Self-imposed Constraints and Rationalizable Choice

Sean Ingham*

September 19, 2011

1 Introduction

Orthodox rational choice theory assumes that individuals choose what they most prefer. A necessary condition for this assumption is that the individual’s choices exhibit a minimal degree of consistency across different menus of options. In such cases, the choice behavior can be taken to “reveal” a preference relation. One way in which this consistency might fail to obtain is if the agent accepts ethical principles or social norms that constrain her choices from different menus in such a way that whatever she chooses from these constrained menus, her choices are bound to be inconsistent. There is also the possibility that the self-imposed constraint does not by itself guarantee this inconsistency, but does so in combination with the (constrained) optimization of a preference relation. This paper explores the conditions on the self-imposed constraint function and the preference relation which determine whether or not the choice behavior reveals a preference relation.

An important early inquiry into questions of this kind came with Sen (1993), which criticized the exclusive reliance on “internal consistency” conditions in standard accounts as the sole basis for determining whether choice behavior is “rational.” Sen argued that certain forms of choice behavior would appear inconsistent with simple, unconstrained preference maximization, and hence “irrational” on the standard view, if one neglected to consider

---

*Ph.D. candidate, Department of Government, Harvard University. Email: ingham@fas.harvard.edu. Website: [http://scholar.harvard.edu/ingham](http://scholar.harvard.edu/ingham).
the role that norms might play in determining choice. The following example, which Sen used to motivate his criticism of the orthodox view of rational choice, has influenced much of the subsequent literature on norm-constrained choice behavior.

“Suppose the person faces a choice at a dinner table between having the last remaining apple in the fruit basket (y) and having nothing instead (x), forgoing the nice-looking apple. She decides to behave decently and picks nothing (x), rather than the one apple (y). If, instead, the basket had contained two apples, and she had encountered the choice between having nothing (x), having one nice apple (y) and having another nice one (z), she could reasonably enough choose one (y), without violating any rule of good behavior. The presence of another apple (z) makes one of the two apples decently choosable, but this combination of choices would violate the standard consistency conditions... even though there is nothing particularly “inconsistent” in this pair of choices (given her values and scruples)” (Sen 1993, p. 501).

Sen (1993) has inspired further research on the way in which the norm-constrained behavior might be reconciled with the standard theory of rational choice or some revision thereof. Baigent & Gaertner (1996) focus on Sen’s example of a person who never chooses the uniquely “largest” (or “best” according to some quality ordering), and Gaertner & Xu (1999a) provide an axiomatic characterization of choices guided by the norm of choosing the median according to some quality ordering (see also Gaertner & Xu 1999a, Gaertner & Xu 1999b). Xu (2007) considers such norms and identifies necessary and sufficient conditions for each norm under which choice behavior can be represented as maximizing a preference relation subject to the norm.

In response to the general allegation that the standard account of rational choice neglects the importance of norms, a defender of that account might respond that the influence of self-imposed norms on behavior is reflected in the standard account by influencing the preference relation, which the agent is assumed to maximize (without constraint). In other words, a defender of the standard account might respond that behavior which is viewed as maximizing a preference relation subject to a self-imposed constraint can be viewed as maximizing some other “revealed” preference relation without constraints. As the examples of the aforementioned norms might suggest, this response is not in general available (see Sen 1997). Baigent (2007) addresses
the question of when it is available, i.e., of when norm-constrained behavior is rationalizable by a revealed preference relation, and this question also motivates the present paper, which generalizes several of Baigent’s (2007) results. An important result of Baigent (2007) is that if the self-imposed constraint is represented as a choice function rationalizable by a weak order and the agent’s choice behavior results from maximizing a weak order subject to this self-imposed constraint, then her choice behavior can be rationalized by a weak order (Baigent 2007, Theorem 1). Theorem 2 of this paper generalizes his result. Baigent also shows that this condition, while sufficient, is not necessary for rationalizability of the choice behavior. Proposition 2 extends this result, showing that rationalizability of choice behavior is consistent with the self-imposed norm violating not only transitivity but also weaker consistency conditions and with the preferences that determine choices from norm-constrained menus being at the same time cyclic.

Section 2 states definitions used in the results presented in section 3. Section 4 concludes with a discussion of the significance of these results and the recent research on norm-constrained rationalizability for applications of the theory of rational choice in social science.

2 Definitions

A choice function is a function $C : X \to X$ such that $\forall S \in X, C(S) \subseteq S$, where $X$ is the set of nonempty subsets of the set $X$ of alternatives. Let $\mathcal{C}$ denote the set of all such self-maps on $X$.

A preference relation $R$ on the set $X$ is a binary relation on $X$, $R \subseteq X \times X$, where the notation $xRy$ is taken to mean that $(x, y) \in R$. In this paper, all preference relations under consideration will be assumed to be complete, meaning that $\forall x, y \in X, xRy$ or $yRx$. I denote the set of all complete binary relations on $X$ by $\mathcal{B}$. Throughout, $P$ denotes the asymmetric part of $R$ (strict preference) — that is, $xPy$ if and only if $xRy$ and not $yRx$. $I$ denotes its symmetric part (indifference) — that is, $xIy$ if and only if $xRy$ and $yRx$. Several additional properties that a binary relation may possess include:

**Definition 1.** A binary relation $R \in \mathcal{B}$ satisfies

1. **transitivity** if and only if $\forall x, y, z \in X$, if $xRy$ and $yRz$, then $xRz$,
2. **quasi-transitivity** if and only if $\forall x, y, z \in X$, if $xPy$ and $yPz$, then $xPz$,
3. **acyclicity** if and only if $\forall \{x_1, x_2, x_3, ..., x_{n-1}, x_n\} \subseteq X$, if $x_1Px_2$ and $x_2Px_3$... and $x_{n-1}Px_n$, then $x_1Rx_n$.
The standard theory of rational choice assumes that choice behavior can be rationalized by a preference relation in the sense specified by definition 2.

**Definition 2.** The maximal set of \( S \subseteq X \) with respect to \( R \in B \) is defined as 
\[
M(R,S) \equiv \{ x \in S : \forall y \in S, xRy \},
\]
and a binary relation \( R \in B \) rationalizes a choice function \( C \in C \) if and only if \( \forall S \in \mathcal{X}, C(S) = M(R,S) \).

An important fact that simplifies the task of determining the rationalizability of a choice function \( C \) is that a choice function is rationalizable by a binary relation \( R \) if and only if \( R \) is the base relation associated with \( C \):

**Definition 3.** The base relation \( R_C \) associated with the function \( C \in C \) is the relation \( R_C \in B \) such that \( \forall x,y \in X, xR_C y \Leftrightarrow x \in C(\{x,y\}) \).

Whether a choice function can be rationalized depends on its contraction- and expansion-consistency properties. The most important properties are the following.

**Definition 4.** A function \( C \in C \) satisfies condition  
(1) \( \alpha \) if and only if \( \forall S,T \in \mathcal{X} \) such that \( S \subseteq T, S \cap C(T) \subseteq C(S) \),  
(2) \( \beta \) if and only if \( \forall S,T \in \mathcal{X} \) such that \( S \subseteq T \), if \( C(S) \cap C(T) \neq \emptyset \), then \( C(S) \subseteq C(T) \),  
(3) \( \gamma \) if and only if \( \forall S,T \in \mathcal{X}, C(S) \cap C(T) \subseteq C(S \cup T) \),  
(4) path-independence if and only if \( \forall S,T \in \mathcal{X}, C(S \cup T) = C(C(S) \cup C(T)) \).

As is well-known, a choice function is rationalizable if and only if it satisfies \( \alpha \) and \( \gamma \); it is rationalizable by a quasi-transitive binary relation if and only if it satisfies \( \gamma \) and path-independence; and it is rationalizable by a transitive binary relation if and only if it satisfies \( \alpha \) and \( \beta \) (Arrow 1959, Sen 1971).

The questions which this paper seeks to answer concern the rationalizability of choice behavior that is somehow informed not only by a preference relation but also by a self-imposed norm. Following Sen (1997), the choice behavior of someone who always chooses subject to respecting some self-imposed norm or constraint is represented by a choice function \( C_1 \in C \) which is defined in terms of another choice function (e.g., the self-imposed constraint) \( C_2 \in C \) and a binary relation \( \tilde{R} \in B \) such that, \( \forall S \in \mathcal{X}, C_1(S) = M(\tilde{R}, C_2(S)) \). Sen (1997) establishes that for any such \( R_1 \), there is some constraint function \( C_2 \) such that the resulting choice function \( C_1 \) cannot be rationalized by a binary relation (Sen 1997, theorem 6.2). The results in 3 explore further the conditions on \( \tilde{R} \) and \( C_2 \) that determine the rationalizability of \( C_1 \).
To ensure that the choice function $C_1$ is well-defined, the set $M(\tilde{R},C_2(S))$ must be nonempty for each $S \in \mathcal{X}$. On account of the role of the constraint function $C_2$, acyclicity of $\tilde{R}$ is not necessary. Instead, it is merely necessary that $\tilde{R}$ is “acyclic on the range” of $C_2$, as specified in definition 5.

**Definition 5.** A binary relation $\tilde{R} \in \mathcal{B}$ is

1. transitive on the range of $C \in \mathcal{C}$ if and only if $\forall S \in \mathcal{X}$ and $\forall x,y,z \in C(S)$, if $x\tilde{R}y$ and $y\tilde{R}z$, then $x\tilde{R}z$,
2. quasi-transitive on the range of $C \in \mathcal{C}$ if and only if $\forall S \in \mathcal{X}$ and $\forall x,y,z \in C(S)$, if $x\tilde{P}y$ and $y\tilde{P}z$, then $x\tilde{P}z$,
3. acyclic on the range of $C \in \mathcal{C}$ if and only if $\forall S \in \mathcal{X}$ and $\forall \{x_1,x_2,x_3,\ldots,x_{n-1},x_n\} \subseteq C(S)$, if $x_1\tilde{P}x_2$ and $x_2\tilde{P}x_3\ldots$ and $x_{n-1}\tilde{P}x_n$, then $x_1\tilde{R}x_n$.

This form of acyclicity is assumed in what follows whenever reference is made to a choice function $C_1$ given by $C_1(S) = M(\tilde{R},C_2(S))$. For a fuller treatment of the conditions on the choice function $C_1$ and constraint function $C_2$ under which $C_1$ permits a characterization of the kind assumed here, see Bossert & Suzumura (2008).

### 3 Results

The self-imposed norms used both in Sen’s (1997, theorem 6.2) and in Sen’s (1993) informal discussion of not choosing the uniquely largest violate condition $\alpha$, which is known to be a necessary condition on choice functions for their rationalizability. This fact might lead one to think that the constrained choice function $C_1$ inherits its contraction- and expansion-consistency properties from the constraint function $C_2$. This thought is on the right track, but not wholly correct. For example, the constraint function $C_2$ may satisfy the condition $\gamma$ and the condition of path-independence — which guarantee that $C_2$ is rationalizable by a quasi-transitive relation — but the base relation $R_{C_1}$ may be cyclic and thus fail to rationalize $C_1$, even if $\tilde{R}$ is transitive.

**Proposition 1.** For some $C_1,C_2 \in \mathcal{C}$ and transitive $\tilde{R} \in \mathcal{B}$ such that $C_2$ satisfies condition $\gamma$ and is path-independent and $\forall S \in \mathcal{X}$, $C_1(S) = M(\tilde{R},C_2(S))$, the base relation $R_{C_1}$ is cyclic.

**Proof.** Let $X = \{x,y,z\}$ and suppose that $C_2(\{x,y,z\}) = \{x,z\}$, $C_2(\{x,y\}) = \{x,y\}$, $C_2(\{y,z\}) = \{z\}$ and $C_2(\{x,z\}) = \{x,z\}$, and
further that $y \bar{P} x, x \bar{P} z,$ & $y \bar{P} z$. Then $C_2$ is path-independent and satisfies condition $\gamma$ and $\bar{R}$ is transitive. But it follows from these suppositions that $C_1(\{x, y\}) = \{y\}$, $C_1(\{y, z\}) = \{z\}$, and $C_1(\{x, z\}) = \{x\}$. Thus, the base relation $R_{C_1}$ is cyclic.

Proposition 1 indicates how the rationality of $C_1$ can be less than that of its parts: the base relation associated with $C_1$ is cyclic even though for any set $S \in \mathcal{X}$, the choice set $C_1(S)$ is just the elements of $S$ chosen according to a transitive binary relation from a set of elements chosen from $S$ according to a quasi-transitive relation. Hence, norm-constrained choice functions do not inherit the properties of their ingredients in any simple or direct way.

Nonetheless, it is true that increasing the consistency of the constrained function $C_2$ eventually entails the rationalizability of the constrained choice function $C_1$. Baigent (2007) establishes the following result.

**Theorem 1.** For any $C_1, C_2 \in \mathcal{C}$, if $C_1$ and $C_2$ each satisfy conditions $\alpha$ and $\beta$, then there exists some transitive $R \in \mathcal{B}$ such that $\forall S \in \mathcal{X}$, $C_1(C_2(S)) = M(R, S)$.

**Proof.** See Baigent (2007, theorem 1)

Translated into the slightly different framework used in this paper and in Sen (1997), Baigent’s (2007, theorem 1) states that a choice function $C_1$ satisfies $\alpha$ and $\beta$ — and is therefore rationalizable by a transitive binary relation — if it is given by $C_1(S) = M(\bar{R}, C_2(S))$ for transitive $\bar{R} \in \mathcal{B}$ and $C_2 \in \mathcal{C}$ satisfying $\alpha$ and $\beta$. It is possible to generalize Baigent’s result, however, to cases where $\bar{R}$ fails to be transitive, or even fails to be acyclic, provided that it is transitive on the range of $C_2$. (That is, one can generalize his result to cases where only the constraint function $C_2$ satisfies conditions $\alpha$ and $\beta$, provided that the choice behavior is informed by a preference relation that is transitive on the range of $C_2$.) Furthermore, in cases where the choice behavior is informed by a preference relation that is intransitive on the range of $C_2$, but quasi-transitive (acyclic) on the range of $C_2$, then the choice behavior is rationalizable by a quasi-transitive (acyclic) binary relation, if the constraint function satisfies $\alpha$ and $\beta$.

---

Baigent mistakenly claims that conditions $\alpha$ and $\gamma$ are necessary and sufficient conditions on a choice function for it to be rationalizable by a transitive binary relation, but this mistake does not vitiate his proof: he shows that if the choice functions $C_1$ and $C_2$ are each rationalizable by a transitive preference relation, then the composite choice function $C_1 \circ C_2$ is also rationalizable by a transitive preference relation.
Theorem 2. For any \( C_1, C_2 \in \mathcal{C} \) and binary relation \( \tilde{R} \in \mathcal{B} \) such that \( \forall S \in \mathcal{X}, C_1(S) = M(\tilde{R}, C_2(S)) \), if \( C_2 \) satisfies conditions \( \alpha \) and \( \beta \), then there exists an acyclic preference relation \( R \in \mathcal{B} \) such that

\[
\forall S \in \mathcal{X}, C_1(S) = M(R, S),
\]

and if \( \tilde{R} \) is quasi-transitive on the range of \( C_2 \), then \( R \) is quasi-transitive and if \( \tilde{R} \) is transitive on the range of \( C_2 \), then \( R \) is transitive.

Proof. Assume that \( C_2 \) satisfies \( \alpha \) and \( \beta \) and \( \tilde{R} \) is acyclic on the range of \( C_2 \), but that that \( C_1 \) is not rationalizable. Then it follows that

\[
\exists S \in \mathcal{X} : C_1(S) = M(\tilde{R}, C_2(S)) \neq M(R_{C_1}, S).
\]

This implies that

\[
\exists x \in S : \forall y \in S, xR_{C_1}y \land [x \notin C_2(S) \lor \exists z \in C_2(S) : z\tilde{P}x],
\]

or

\[
\exists y \in C_2(S) : \forall z \in C_2(S), y\tilde{R}z \land \exists w \in S : wP_{C_1}y.
\]

Suppose first that (1) is true. Since \( C_2 \) satisfies \( \alpha \) and \( \beta \), it is rationalized by a transitive relation, which is its base relation. Hence, (1) can be rewritten as,

\[
\exists x \in S : \forall y \in S, xR_{C_1}y \land [x \notin M(R_{C_2}, S) \lor \exists z \in M(R_{C_2}, S) : z\tilde{P}x]
\]

For any pair \( \{v, w\} \subset \mathcal{X}, v \in C_1(\{v, w\}) \) implies \( v \in C_2(\{v, w\}) \), by definition of the constrained choice function \( C_1 \). Hence, \( xR_{C_1}y \) implies \( xR_{C_2}y \) and (1) therefore implies that \( \forall y \in S, xR_{C_2}y \). Thus, \( \exists z \in M(R_{C_2}, S) : z\tilde{P}x \) and \( zI_{C_2}x \). Consequently, \( M(\tilde{R}, C_2(\{x, z\})) = M(\tilde{R}, \{z, x\}) = \{z\} \), but this contradicts the supposition that \( xR_{C_1}y, \forall y \in S \). Hence, (1) is false.

Suppose, then, that (2) is true.

\[
\exists y \in C_2(S) : \forall z \in C_2(S), y\tilde{R}z \land \exists w \in S : wP_{C_1}y.
\]

Since \( C_2 \) satisfies \( \alpha \) and \( \beta \), (2) can be rewritten as,

\[
\exists y \in M(R_{C_2}, S) : \forall z \in M(R_{C_2}, S), y\tilde{R}z \land \exists w \in S : wP_{C_1}y.
\]
wP_{C_1,y} implies that either \( wR_{C_2}y \) and \( \hat{w}P_2y \) or \( wP_{C_2,y} \). If \( wP_{C_2,y} \), then \( y \notin M(R_{C_2}, S) \). Hence, \( wR_{C_2}y \) and \( \hat{w}P_2y \). Hence, \( w \notin M(R_{C_2}, S) \), because \( \forall z \in M(R_{C_2}, S), y \hat{R}z \). Thus, \( \exists v \in S \) such that \( vP_{C_2,w} \). Since \( wR_{C_2}y \) and \( R_{C_2} \) is transitive, it follows that \( vP_{C_2,w} \), but this contradicts the supposition that \( y \in M(R_{C_2}, S) \). Thus, (2) is false and so

\[
\forall S \in X, \ C_1(S) = M(R_{C_1}, S).
\]

To demonstrate that \( R_{C_1} \) is transitive if \( \hat{R} \) is transitive on the range of \( C_2 \), suppose that \( \hat{R} \) is transitive on the range of \( C_2 \) and \( R_{C_1} \) is not transitive. Then, \( \exists x, y, z \in X \) such that \( xR_{C_2}y \) and \( yR_{C_1,z} \), but \( zP_{C_1,x} \). By definition of \( C_1 \), this implies that \( xR_{C_2}y, yR_{C_2,z} \), and \( zR_{C_2}x \). By transitivity of \( R_{C_2} \), it then follows that \( xI_{C_2}z, yI_{C_2}z, \) and \( xI_{C_2}y \). Since \( C_2 \) is rationalized by \( R_{C_2} \), \( C_2(\{x,y,z\}) = \{x,y,z\} \). For it to be true that \( xR_{C_1}y \) and \( yR_{C_1,z} \), but \( zP_{C_1,x} \), as supposed, it must by definition of the constrained choice function \( C_1 \), be true that \( x\hat{R}y, y\hat{R}z, \) and \( z\hat{P}x \). Since \( C_2(\{x,y,z\}) = \{x,y,z\} \), this conclusion contradicts the assumption that \( \hat{R} \) is transitive on the range of \( C_2 \).

To demonstrate that \( R_{C_1} \) is quasi-transitive if \( \hat{R} \) is quasi-transitive on the range of \( C_2 \), suppose that \( \hat{R} \) is quasi-transitive on the range of \( C_2 \) and \( R_{C_1} \) is not quasi-transitive. Then, \( \exists x, y, z \in X \) such that \( xP_{C_1,y} \) and \( yP_{C_1,z} \), but \( zR_{C_1}x \). By the definition of \( C_1 \), this implies that \( xR_{C_2}y, yR_{C_2,z} \), and \( zR_{C_2}x \). By transitivity of \( R_{C_2} \), it implies that \( xI_{C_2}y, yI_{C_2}z, \) and \( xI_{C_2}z \). Since \( C_2 \) is rationalized by \( R_{C_2} \), \( C_2(\{x,y,z\}) = \{x,y,z\} \). For it to be true that \( xP_{C_1,y} \) and \( yP_{C_1,z} \), but \( zR_{C_1}x \), as supposed, it must, by definition of the constrained choice function \( C_1 \), be true that \( x\hat{P}y, y\hat{P}z, \) and \( z\hat{R}x \). Since \( C_2(\{x,y,z\}) = \{x,y,z\} \), this conclusion contradicts the assumption that \( \hat{R} \) is quasitransitive on the range of \( C_2 \).

The intuition behind the result is that when the constraint function \( C_2 \) is rationalizable by a transitive base relation \( C_2 \) — as it is when \( C_2 \) satisfies \( \alpha \) and \( \beta \) — then the role of the relation \( \hat{R} \) in the constrained choice function \( C_1(\cdot) = M(\hat{R}, C_2(\cdot)) \) is to break ties in \( R_{C_2} \). Whenever \( xP_{C_2,y} \), it immediately follows that \( xP_{C_1,y} \), irrespective of how \( \hat{R} \) ranks \( x \) and \( y \). Alternatively, if \( xI_{C_2}y \), then the ranking of \( x \) and \( y \) according to \( R_{C_1} \) coincides with the ranking of the two elements according to \( \hat{R} \). In such cases, any failure in the transitivity of \( R_{C_1} \) must be due to such a failure in \( \hat{R} \), which is why the (quasi)transitivity of the latter on the range of \( C_2 \) implies the (quasi)transitivity of the former. This fact about the relation between \( \hat{R} \) and
While it has been more common in the literature to think of norms as constraining the maximization of preferences, one can also, as Baigent (2007) points out, consider maximization of preferences as constraining the satisfaction of norms. At some fundamental level, this order of things must surely be more accurate, unless one thinks that individuals are willing to incur arbitrarily large costs for the sake of satisfying norms. If one wanted to maintain the current model that assumes a lexical relationship between norms and preferences, then the most plausible view might require adding additional choice functions to the composite function characterizing choice behavior. For example, an individual may choose elements by first eliminating from consideration any that compromise her most fundamental interests and needs, then eliminate from consideration the options that violate whatever norms she has internalized, and then finally choose from the remaining menu of options according to some preference relation. But fundamental needs and interests are most plausibly modeled by a transitive binary relation: if \( x \) compromises one’s most basic interests more than \( y \) and \( y \) compromises them more than \( z \), then \( x \) compromises them more than \( z \). In many cases, considerations of basic needs will underdetermine a choice, meaning that there will be indifference in the binary relation representing these basic interests, and hence norms, and preferences reflecting less fundamental desires, will have a nontrivial role to play in determining choice behavior. The possibility of multiple kinds of considerations — fundamental interests, various norms, less fundamental desires and tastes, etc. — determining choice behavior raises the question of whether theorem 2 can be generalized to cases where choice behavior is modeled by arbitrarily many lexically ordered choice functions.

**Theorem 3.** For any \( n \in \mathbb{N} \) and \( (C_1, C_2, \ldots, C_n), (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_n) \) such that

\[
\forall i = 2, \ldots, n, \forall S \in \mathcal{X}, \quad C_i(S) = M(\tilde{R}_i, C_{i-1}(S)),
\]

where each \( \tilde{R}_i \) is transitive on the range of \( C_{i-1} \), there exists a transitive \( R \in \mathcal{B} \) such that \( \forall S \in \mathcal{X}, \ C_n(S) = M(R, S) \), if for some \( i \leq n \), \( C_i \) satisfies conditions \( \alpha \) and \( \beta \).

**Proof.** The result is trivial if \( C_n \) satisfies \( \alpha \) and \( \beta \), so consider some \( i < n \), and suppose that \( C_i \) satisfies conditions \( \alpha \) and \( \beta \). Since by assumption
∀S ∈ X, C_{i+1}(S) = M(\tilde{R}_{i+1}, C_i(S)) and \tilde{R}_{i+1} is transitive on the range of C_i, it follows from theorem 1 that C_{i+1} is rationalized by a transitive binary relation. Consequently, C_{i+1} satisfies α and β and so, by similar reasoning, C_{i+2} satisfies conditions α and β. Continuing this line of reasoning leads to the conclusion that C_n is rationalized by a transitive relation R ∈ B.

With regard to the three-tiered decision procedure described above, theorem 3 tells us that if the norms can be represented by a binary relation that, while possibly cyclic, is acyclic on the subsets of alternatives that are maximal according to the binary relation representing the individual’s most fundamental interests, and the preferences representing the final determinants of choice are transitive on the norm-constrained menus, then the choice behavior is itself rationalizable by a transitive binary relation.

Theorem 2 may also be of interest in contexts other than rationalizing the choices of a single individual. Theorem 2 also tells us something about the rationalizability of a choice process that consists in one agent choosing from the alternatives chosen from a set S by a second agent. For example, if, say, an agency chooses a policy from a set of possible policies by first having a team of advisors submit a menu of choices from this larger set to their advisee, and the advisee then chooses what she most prefers from this constrained menu, the entire selection process will be rationalizable by an acyclic preference relation if the choices made by the advising team are rationalizable by a transitive preference relation and the preference relation of the advisee is acyclic on the range of the advising team’s choice function.

Theorem 2 does not run in the other direction, however. Theorem 4 of Baigent (2007) establishes that there are composite choice functions of the form C_1 ∘ C_2 that are rationalizable by transitive binary relations, even though neither C_1 nor C_2 is rationalizable by a transitive binary relation. Proposition 2 is a strengthening of this result: not only may the function fail to be rationalizable by a transitive relation, but it and C_2 may fail to be rationalizable by any binary relations, and yet their composition may still be rationalizable by a transitive binary relation.

**Proposition 2.** There exist C_1, C_2 ∈ C and cyclic \tilde{R} ∈ B such that

∀S ∈ X, \quad C_1(S) = M(\tilde{R}, C_2(S)) = M(R_{C_1}, S),

\footnote{Note that theorem 3 is stronger than what may be derived immediately from Baigent (2007, theorem 1), since theorem 3 requires only that each of the binary relations R_i is transitive on the range of C_{i-1}.}
and \( R_{C_1} \) is transitive, but \( C_2 \) satisfies neither \( \alpha, \gamma, \beta \), nor path-independence.

**Proof.** Let \( X = \{ w, x, y, z \} \) and suppose that \( C_2(\{ w, x, y, z \}) = C_2(\{ x, y, z \}) = \{ x, y \} \), \( C_2(\{ x, y \}) = \{ y \} \), \( C_2(\{ z, y \}) = \{ y \} \), \( C_2(\{ x, z \}) = \{ x, z \} \), \( C_2(\{ w, x \}) = \{ w, x \} \), \( C_2(\{ w, y \}) = \{ w, y \} \), and \( C_2(\{ w, z \}) = \{ w, z \} \). \( C_2 \) violates conditions \( \alpha, \beta, \gamma \), and path-independence. Consistent with these suppositions, we can suppose that \( x\tilde{P}z, z\tilde{P}y, y\tilde{P}x \), and \( a\tilde{P}w, \forall a \in X \setminus \{ w \} \), which means that \( \tilde{R} \) is cyclic. It then follows that \( C_1(\{ w, x, y, z \}) = C_1(\{ x, y, z \}) = \{ y \} \), \( C_1(\{ x, y \}) = \{ y \} \), \( C_1(\{ z, y \}) = \{ y \} \), \( C_1(\{ x, z \}) = \{ x \} \), and \( w \notin C_1(\{ a, w \}), \forall a \in X \setminus \{ w \} \). \( C_1 \) has a transitive base relation that rationalizes these choices: \( yP_{C_1}xP_{C_1}zP_{C_1}w. \)

Proposition\ref{2} illustrates how apparent counterexamples to orthodox rational choice theory might be merely apparent counterexamples. Suppose that person \( j \) has a horrible time making decisions from menus of alternatives in a set \( X \), because \( j \)’s preferences \( R_j \) on \( X \) are cyclic. As an intended remedy, he always rules out of consideration any alternative that, he thinks, neither of his parents, \( f \) and \( m \), would choose, and then chooses according to \( R_j \) from the remaining alternatives. Hence his choices are given by the constrained choice function \( C_j(S) = M(R_j, C_f(S) \cup C_m(S)) \), where \( C_f \), \( C_m \) denote the choice functions characterizing his father’s and mother’s choices, respectively.

Note that the choice function \( C_P \equiv C_f \cup C_m \) may not satisfy \( \beta \), even if \( C_f \) and \( C_m \) each satisfy \( \alpha \) and \( \beta \). For example, we may have \( C_f(\{ x, y \}) = \{ x, y \} \) and \( C_m(\{ x, y \}) = \{ x \} \), but \( C_P(\{ x, y, z \}) = \{ z \} \) and \( C_m(\{ x, y, z \}) = \{ x \} \), in which case \( C_P(\{ x, y, z \}) = \{ x, z \} \) and \( C_P(\{ x, y \}) = \{ x, y \} \). Despite the strangeness by which person \( j \) makes decisions, however, proposition\ref{2} cautions against inferring that his choices cannot be rationalized by a transitive preference relation (one different from \( R_j \) of course). For some specifications of \( R_j, C_m, \) and \( C_f \), his choices are rationalizable.

A note on the relation between theorem\ref{2} and Arrow’s General Possibility Theorem is worth making here. When the constraint function \( C_2 \) is assumed to satisfy \( \alpha \) and \( \beta \) and \( \tilde{R} \) is assumed to be transitive, then the constrained choice function \( C_1 \), defined so that \( C_1(S) = M(\tilde{R}, C_2(S)) \) for all \( S \in \mathcal{X} \), can be thought of as collective choice rule \( \phi : \mathcal{R}^2 \times \mathcal{X} \to \mathcal{X} \), where \( \mathcal{R} \) is the set of all transitive and complete binary relations on \( X \). For any \( S \in \mathcal{X} \) and \( (\tilde{R}, R_{C_2}) \in \mathcal{R}^2 \), \( \phi((\tilde{R}, R_{C_2}), S) = M(\tilde{R}, M(R_{C_2}, S)) \). The fact that \( \phi \) is guaranteed to be rationalizable by a transitive relation, given the assumptions that \( C_2 \) satisfies \( \alpha \) and \( \beta \) and \( \tilde{R} \) is transitive, can be viewed as a special case
of Arrow’s theorem, since the collective choice rule $\phi$ is dictatorial: for any $(\tilde{R}, R_{C_2}) \in \mathcal{R}^2$ and any $x, y \in X$, if $x P_{C_2} y$, then $y \notin \phi((\tilde{R}, R_{C_2}), S), \forall S \in \mathcal{X}$ such that $x \in S$.

Note, however, that it is only possible to view constrained choice functions as examples of collective choice rules when the constraint function $C_2$ satisfies $\alpha$ and $\gamma$ and is therefore rationalizable by its base relation. If $C_2$ does not satisfy these conditions, it is not rationalizable by its base relation, and the constrained choice function cannot be viewed as the result of aggregating the relation $\tilde{R}$ and the relation $R_{C_2}$ into a “collective” choice function $C_1$. When $C_2$ is not rationalizable, the information contained in it cannot be summarized by its base relation, hence one cannot reinterpreted the constrained choice function $C_1$ as just a collective choice rule that aggregates the information contained in $\tilde{R}$ and $R_{C_2}$.

Furthermore, if $C_2$ does satisfy $\alpha$ and $\gamma$ so that $C_1$ can be viewed as induced by a collective choice rule, if $C_2$ does not satisfy $\alpha$ and $\beta$, then Arrow’s theorem will not underwrite any inferences concerning its rationalizability. As proposition [1] indirectly implies, the dictatorship property does not suffice to guarantee “social rationality” (or rationalizability of $C_1$) when the domain of the preference aggregation rule is the set of all complete and quasi-transitive binary relations on $X$, rather than the set of all complete and transitive binary relations on $X$, as is assumed in Arrow’s theorem.

On account of Arrow’s specification of the domain of possible preference profiles, theorem [2] is also not simply deducible from Arrow’s theorem even when $C_2$ satisfies $\alpha$ and $\beta$, for theorem [2] covers cases where $\tilde{R}$ is not transitive, but rather only acyclic or only quasi-transitive or only transitive on the range of $C_2$.

### 4 Conclusion

In contrast with much of the recent literature on norm-constrained rationalizability, Sen (1993) and Sen (1997) were not focused on how specific norms — such as ‘the choose the median’ and ‘never choose the uniquely largest’ norms — might give rise to choice behavior that is not rationalizable by a preference relation, but rather with the question of whether the incorporation of norms in general creates difficulties for orthodox rational choice theory. Indeed, the language of Sen (1997) indicates that Sen was tempted to state rather strongly the obstacles to modeling norm-constrained behav-
ior as unconstrained preference-maximization: on the basis of his theorem 6.2, Sen suggests that if a person’s choices are given by a constrained choice function, where the constraint function $C_2$ is interpreted as a self-imposed ethical or social constraint, then the only way to model this person’s choices by rationalizable choice function is to impute to them “menu-dependent” preference relations. Formally, a menu-dependent preference relation $R^S$ on $S \subseteq X$ is a complete and transitive binary relation $R^S \subseteq S \times S$. The choice behavior of an agent who acts according to menu-dependent preference relations can be described by a choice function $C : \mathcal{X} \to \mathcal{X}$ such that $\forall S \in \mathcal{X}, C(S) = M(R^S, S)$, for some $R^S \subseteq S \times S$. The difference between such an agent and the agent that ordinarily populates rational choice models is that a menu-dependent preference-maximizer might choose $C(S) = M(R^S, S)$ and $C(T) = M(R^T, T)$, where $M(R^T, S) \neq M(R^S, S)$. Sen writes: “It is, thus, clear [from theorem 6.2] that while the approach of "as if" preferences can take on the role of “mimicking” the use of self-imposed choice constraints, the indexation $S$ in $R^S$ is necessary for this to work (a menu-independent “as if” preference $R$ would not do” (Sen 1997, p. 770, emphasis added). There is, he suggests, a “major technical gulf” between modeling ethical commitments in the form of self-imposed constraints, on the one hand, and modeling them by incorporating them into a preference relation which the agent then maximizes, on the other, so long as one assumes menu-independent preferences, as is traditionally done (Sen 1997, p.773).

Theorem 2 identifies conditions sufficient to bridge this technical gulf, while 2 indicates why the preferences and norms determining choice behavior need satisfy neither these conditions nor other consistency conditions typically imposed on preferences and choice functions in order for choice behavior to be rationalizable. The latter result counsels against assuming that choice behavior will not be assimilable to the orthodox model of rational choice simply because it is driven in part by a norm which cannot be modeled by a transitive preference relation, while the former result identifies a class of norms — those representable by choice functions satisfying $\alpha$ and $\beta$ — for which behavior resulting from maximization of preferences subject to these norms is guaranteed to be assimilable to the orthodox model of rational choice. Both of these results generalize previous results found in Baigent (2007).

Arguably more important, however, than these technical questions to an appraisal of how easily the orthodox model of rational choice can accommodate norms is the question of how one ought to represent the set $X$ of
possible choices. Consider again the example used to motivate so much of the research into norm-constrained choice behavior: a host has two guests, $A$ and $B$, over for dinner. If he were to offer $A$ first choice of three pieces of cake, $A$ would choose the second-largest, but if he were offer the choice first to $B$ and $B$ chose the smallest piece available, then $A$ would choose the largest. Clearly, if one identifies the set of possible choices with the different cake pieces, one cannot construct a preference relation that would be consistent with each of these two hypothetical choices. The explanation for the agent’s choice behavior must then be something like the one that Sen gives and that others writing on norm-constrained choices have affirmed, namely that the agent imposes a politeness constraint on her choices, a constraint representable by a choice function violating condition $\alpha$.

Alternatively, one might represent the set of $A$’s possible choices, not as the set of cake pieces, but rather as elements of a two-dimensional space, where one dimension measures the size of cake pieces on offer and the other dimension measures the esteem in which $B$ and $A$’s host hold $A$ and how offended they are by his behavior. If $A$ is committed to following the rule apart from the derivative consequences of doing so, the second dimension might instead just measure whether or not he follows the rule, and if he is absolutely committed to the rule, then he will have a lexicographic, but transitive preference relation on this two-dimensional space. There is nothing hokey about the latter approach, and it corresponds to the normal practice in applications of rational choice theory. Modelers instinctively describe the set of options facing an agent in a way that includes the properties of states of affairs that are thought to be relevant to the agent’s choice. For example, if an economist believes that a worker cares only about his wage and the amount of leisure time (or, more precisely, can be so treated for the sake of the inquiry), then the economist will model the worker’s set of choices as different combinations of wages and leisure time. But if instead it is thought that the arduousness of tasks is also relevant to the agent, then the set of choices must be described so as to reflect this fact through the inclusion of an additional dimension to the space of options. If, contrary to this requirement, each of the worker’s feasible options is represented as a wage-leisure vector, then the worker might choose a wage-leisure combination $(w, l)$ when $(w', l')$ is available, and later choose $(w', l')$ when $(w, l)$ is available, violating the assumption of preference-maximization. But once it is recognized that the worker also cares about this third aspect of the states of affairs which he can influence, then his options will be better represented as choices of points.
$(w, l, e)$ and $(w', l', e')$ in some three-dimensional space. This representation may of course render his choices consistent with preference-maximization. Likewise, once a modeler realizes that an agent cares about whether or not her choices conform with some norm, then there is no reason not to describe the agent’s choices as a multidimensional space, one of whose dimensions measures compliance with the norm.

References


