The transitive core: inference of welfare from nontransitive preference relations

Hiroki Nishimura*

December 19, 2015

Abstract

This paper studies welfare criteria for decision makers endowed with nontransitive preference relations. If a person has nontransitive preferences, then the classical utilitarian welfare criterion does not identify her welfare order, and the problem of maximizing the welfare becomes unclear. In order to find a reliable method of welfare inference, I propose a series of welfare criteria that apply to nontransitive preference relations. Then, we show that these criteria characterize a unique rule of welfare inference from nontransitive preference relations. This rule, called the transitive core, is applied to a variety of nontransitive preference models, such as semiorders on the commodity space, relative discounting time preferences, justifiable preferences over ambiguous acts, regret preferences on risky prospects, and collective preferences induced by majority voting. These examinations show that the

*Department of Economics, University of California Riverside, 3128 Sproul Hall, Riverside, CA 92521 (Email: hiroki.nishimura@ucr.edu). I am deeply indebted to Efe A. Ok for his continuous support and guidance. I am also grateful to Debraj Ray and David Pearce for their help and encouragement. Lastly, I thank Simone Cerreia-Vioglio, Faruk Gul, Yoram Halevy, Hao Li, Fabio Maccheroni, Michael Peters, Michael Richter, Gil Riella, Ariel Rubinstein, Yuval Salant, Ennio Stacchetti, seminar participants at Bocconi University, Kellogg School of Management at Northwestern University, New York University, Princeton University, University of California Riverside, University of California San Diego, University of California Santa Barbara, University of Southern California, Washington University in St. Louis, Yokohama National University, Bounded Rationality in Choice Workshop, Canadian Economic Association Annual Meeting, Canadian Economic Theory Conference, and anonymous referees for helpful comments and discussions. The financial support from C.V. Starr Center at New York University is gratefully acknowledged. The responsibility of any error is my own.
proposed method provides nontrivial and sensible inference of welfare in respective contexts.

*JEL Classification:* D11, D60.
*Keywords:* Nontransitive preferences, Welfare, Bounded rationality.

## 1 Introduction

Many behavioral models postulate boundedly rational or heuristic choice procedures rather than the standard utility maximization. For example, satisficing, limited cognition, shortlist methods, and framing effects are behavioral choice models that attract significant attention in the literature. (See, for instance, [7], [11], [24], [26], [33], [35], [36], [38] for recent development.) However, when a decision maker follows a behavioral choice procedure, especially when there is no utility function that underlies her behavior, it is unclear how we can infer the decision maker’s welfare. This ambiguity poses difficulty in normative analysis and policy evaluations, for desirability of outcomes is no longer well-defined.

The issue of welfare evaluations under behavioral models has led to the development of behavioral welfare economics. In this field of study, we seek methods of eliciting decision maker’s welfare from observable data. For example, Bernheim and Rangel [7, 8] consider behavioral decision makers affected by ancillary conditions and propose a method of welfare inference for these agents. Following this work, Kőszegi and Rabin [20], Rubinstein and Salant [34], Chambers and Hayashi [12] further study welfare of behavioral decision makers who exhibit choice mistakes, sensitivity to framing effects, and random decisions, respectively.

In this paper, I shed light on one of the classical models of behavioral decision making, namely, cyclic preference relations. To highlight a problem of welfare evaluations under cyclic preference relations, consider a policy maker who tries to make a policy choice in order to maximize the decision maker’s welfare. Suppose that the policy maker observes the decision maker’s preferences over available policies. Then, to maximize the welfare, the policy maker needs to infer from this observation how the policies affect the decision maker’s welfare.

If an observed preference relation is transitive, we may follow the utilitarian welfare criterion. According to the criterion, we understand that one alternative improves the decision maker’s welfare over another if and only if the former is preferred to the latter. Therefore, the objective of the policy maker in this case is simply a preference (utility) maximization problem. However, if an observed
preference relation is cyclic, the utilitarian criterion induces a cyclic welfare order the maximization of which does not have a solution in general. Furthermore, the money-pump argument implies an unrealistic claim that the policy maker may indefinitely improves the decision maker’s welfare. The utilitarian criterion assumes existence of “utility” after all, and it is not applicable for decision makers with cyclic preference relations.

Experimental literature provides evidence of cyclical choices. In turn, many models of nontransitive preferences are developed to explain such observations. For instance, Kahneman and Tversky [18] present laboratory choice data over uncertain prospects that exhibit pairwise choice cycles. They argue that the data is well explained by a certain nonexpected utility representation that entails nontransitive evaluations of prospects. In the same context, Loomes and Sugden [22] develop a model of nontransitive preferences that accounts for experience of regret. Their model accommodates choice anomalies such as certainty effect, common ratio effect, and common consequence effect (also known as Allais paradox). For intertemporal choice problems, Read [29] and Roelofsma and Read [30] provide experimental data that show that subjects make consistent cyclical choices over intertemporal outcomes. Following this observation, Ok and Masatlioglu [28], Read [29], and Rubinstein [32] study alternative discounting models that induce nontransitive intertemporal preferences. On the consumer utility theory, Armstrong [2, 3, 4, 5] discusses that a decision maker will exhibit nontransitive indifference when alternatives are too similar to be discerned. This intuitive argument has led to the introduction of semiorders (Luce [23]) and interval orders (Fishburn [14]).

This paper proposes a method of inferring the decision maker’s welfare from observation of a complete but nontransitive preference relation. How is it possible to infer the welfare ranking of alternatives when an observed preference relation is cyclic? For example, suppose that a decision maker prefers an alternative $x$ to another alternative $y$. In addition, let us assume that her preference of $x$ over $y$ is not involved in any preference cycles. Then, while the observed preference relation might be cyclic for other pairs of alternatives, it seems reasonable to say that the preference of $x$ over $y$ is consistent and that this part at least reveals the decision maker’s welfare. This argument forms a welfare criterion that applies to nontransitive preference relations. The objective of this paper is to seek a reliable method of welfare inference by studying criteria obtained in such a way.

Formally, I introduce a concept called a welfare evaluation rule. A welfare evaluation rule (WER) is a function that maps every complete binary relation to a preorder. Throughout the paper, I assume that we observe a complete preference
relation (possibly restricted to some domain due to limited availability of choice data) and infer a transitive welfare order. In this sense, a welfare evaluation rule represents a rule of welfare inference. Then, we can formulate welfare criteria like one in the previous paragraph as properties of WERs. This paper studies normative properties of WERs and their implications. In fact, we will show that the proposed properties characterize a unique welfare evaluation rule \( c(\cdot) \) such that

\[
x c(\geq) y \quad \text{if and only if} \quad \begin{cases} z \geq x \implies z \geq y \\ y \geq z \implies x \geq z \end{cases} \quad \text{for every } z \in X
\]

for any complete preference relation \( \geq \) on a nonempty set \( X \). The relation \( c(\geq) \) is transitive even if an observed preference relation \( \geq \) is not. I will refer to the order \( c(\geq) \) as the transitive core of a preference relation \( \geq \).

Ex ante we make no assumption on the decision maker’s preference relation other than being complete. The absence of the assumptions serves as both strength and weakness of the paper. On one hand, it allows wide applications of the transitive core. Since the paper assumes no particular model, we can apply its results to any theory of complete preference relations. By taking advantage of this feature, we will examine implications of the transitive core in a variety of contexts in Section 4. This includes preference relations over commodity bundles, intertemporal outcomes, ambiguous acts, risky prospects, and policy alternatives. These applications are possible because of the model-free approach of the paper. On the other hand, the lack of assumptions limits the room to discuss welfare criteria. Because we do not know why an observed preference relation is nontransitive, properties of welfare inference can only refer to the structure of the observed preference relation. As one consequence of this approach, the transitive core may fail to infer complete welfare orders. We can interpret that the transitive core offers welfare comparison for pairs of alternatives that are consistent enough without assuming a particular model. It leaves pairs of alternatives incomparable if no welfare comparison is sufficiently convincing just by looking at a preference relation.

**Observation of preference relations.** Throughout the paper, I assume observability of a complete but not necessarily transitive preference relation. Typically, such an observability assumption is tied to some theory of revealed preferences from observed choice data. However, this is not straightforward for the present analysis because it is unclear how the decision maker should make a choice under a cyclic preference relation. In order to clarify the underlying assumption about observability of the preference relation, I consider the following three cases.
The first (trivial) case is where our observation consists of pairwise choices. A preference of an alternative $x$ over another alternative $y$ is revealed by observation of choice $x$ from the set $\{x, y\}$. While this is trivial, experimental choice data are often collected from this domain. For example, Kahneman and Tversky [18] and Roelofsuma and Read [30] both reveal a preference relation from their pairwise choice observation. The second case is where a person makes a “rationalizable” choice from arbitrary choice set under a cyclic preference relation. Given a preference relation $\succsim$, we say that an alternative $x$ is indirectly $\succsim$-preferred to another alternative $y$ in a choice set $S$ if there exists a finite sequence $(z_l)_{l=1}^k$ in $S$ such that $x = z_1 \succsim z_2 \succsim \cdots \succsim z_k = y$. Facing a choice set $S$, the decision maker chooses alternatives that are indirectly preferred to every alternative in the choice set. This model is called the top-cycle choice, and it accommodates pairwise choice cycles while it explains choice from three or more alternatives. Ehlers and Sprumont [13] provide testable conditions for existence of a preference relation that rationalizes the choice in this manner. The last case is where a decision maker is endowed with a nontransitive preference relation whereas she also follows a behavioral choice procedure. For example, Manzini and Mariotti [24] study a choice model called the rational shortlist method. The model explains choice behavior in two steps. In the first step, a shortlist preference relation is used to eliminate dominated alternatives from a choice set. In the second step, a tie-breaking preference relation is used to make a choice from those that survive the first step. Neither relation is assumed to be transitive.\footnote{In fact, if both relations are transitive, then the model reduces to the standard rational choice under a complete and transitive preference relation.} The welfare order of alternatives is hence unclear even when we know that the choice can be represented by such a model. The authors of the paper provide an axiomatic characterization of the model and a method of revealing the preference relations from choice observation. In each of the three cases above, there is a way of revealing a nontransitive preference relation from observed choice. The paper assumes observability of this revealed preference relation.

Note that observation of pairwise choice cycles does not always imply that it is made by a cyclic preference relation. For example, Masatlioglu et al. [26] study choice under limited attention. Their model postulates that a decision maker has a complete and transitive preference relation but suffers from limited attention. Facing a choice set $S$, the decision maker recognizes only a subset $\Gamma(S)$ of all available alternatives in $S$. Then, her choice from the set $S$ consists of the best alternatives in $\Gamma(S)$ under the preference relation. The model accommodates pair-
wise choice cycles, but the underlying preference relation is transitive. For choice models such as this, the present paper offers no more than the utilitarian welfare criterion for welfare analysis. Same applies to the case of random choice.

More on related literature. Aside from the literature referred above, there are some works in close connection with the present paper. In social choice theory, Fishburn [15] and Miller [27] introduce the covering order for tournaments (that is, complete asymmetric binary relations). The covering order is motivated as a method of inferring social welfare from a social preference relation under majority voting. In individual choice theory, Luce [23] studies how to infer an individual’s utility function when she fails to distinguish similar alternatives. In fact, we can show that these methods and the transitive core obtain the same welfare order in respective contexts. The covering order of a tournament coincides with the transitive core of the same tournament. The transitive core infers a utility function if an observed preference relation is a semiorder. However, we can also show that these methods do not obtain the same result if they are applied in different contexts. The covering order of a semiorder does not reveal the utility function. Luce’s method applied to a tournament is not even transitive in general. Therefore, the transitive core can be viewed as a genuine generalization of welfare inference methods that reduces to the known rules in specialized contexts. In addition, the transitive core applies to the contexts where there is no obvious way of welfare inference, such as theory of intertemporal choice or choice under risk and uncertainty.

The paper is structured as follows. The next section introduces notation and terminology we use throughout the paper. In Section 3, we formally define welfare evaluation rules and discuss their normative properties. Then, the transitive core is characterized as a unique welfare evaluation rule that satisfies all the proposed properties. Section 4 provides case studies of the transitive core. This includes its applications on semiorders, time preferences, preferences over ambiguous acts, preferences over risky prospects, and collective preferences. In Section 5, we will compare implications of the transitive core with those of the unambiguous welfare improvement order by Bernheim and Rangel [7, 8]. The related literature is further discussed in Section 6. Section 7 closes the paper with concluding remarks. All proofs and supplementary results are given in the appendix.
2 Preliminaries

Preference relations. For any set \( X \), a binary relation on \( X \) is a subset of \( X \times X \) with a generic notation \( \succeq \). As usual, I write \( x \succeq y \) to mean that \((x, y) \in \succeq \). The notations \( > \) and \( \sim \) represent the strict part and the symmetric part of \( \succeq \), respectively.

A binary relation \( \succeq \) on \( X \) is reflexive if \( x \succeq x \) for all \( x \in X \), complete if either \( x \succeq y \) or \( y \succeq x \) for any \( x, y \in X \), transitive if \( x \succeq y \) and \( y \succeq z \) imply \( x \succeq z \) for any \( x, y, z \in X \), and antisymmetric if \( x \succeq y \) and \( y \succeq x \) imply \( x = y \) for any \( x, y \in X \). A reflexive and transitive binary relation on \( X \) is called a preorder on \( X \). A preorder on \( X \) is called a weak order if it is complete, a partial order if it is antisymmetric, and a linear order if it is complete and antisymmetric. The diagonal order on \( X \) is the trivial partial order \( \Delta_X = \{(x, x) : x \in X\} \). In this paper, I refer to a binary relation as a preference relation when it represents the decision maker’s tastes over alternatives. Then, the statement \( x \succeq y \) is read as “\( x \) is preferred at least as much as \( y \).” An ordered pair \( (x, y) \) is called a preference if \( x \succeq y \). The terms indifference, strict preference, and indecision are analogously defined for ordered pairs.

Graphical notation of preference relations. Throughout the paper, preference relations on finite sets are depicted by graphs as in Figures 1 and 2. In these figures, vertices represent alternatives, and directed arrows are depicted from strictly preferred alternatives to less preferred alternatives. Undirected lines between pairs of alternatives represent indifference of the corresponding pairs.

Preference cycles. For any preference relation \( \succeq \) on a set \( X \), a cycle of \( \succeq \) is a finite sequence \((z_l)_{l=1}^k\) of distinct points in \( X \) such that \( z_1 \succeq z_2 \succeq \cdots \succeq z_k \succeq z_1 \) with at least one strict preference. A preference \( (x, y) \) is involved in a cycle \((z_l)_{l=1}^k\) of the relation \( \succeq \) if \( z_l = x \) and \( z_{l+1} = y \) for some \( l < k \) or if \( z_k = x \) and \( z_1 = y \). A preference relation \( \succeq \) is called cyclic if it has at least one cycle and acyclic otherwise. I identify two cycles of a preference relation if they are identical upon rotation. For example, a preference relation in Figure 1 has three cycles \((y, z, w)\), \((z, w, y)\), and \((w, y, z)\), but they are viewed as the same cycle.

Permutation, inverse, restriction. Let \( \succeq \) be any preference relation on a set \( X \), and \( \pi \) a permutation on \( X \). With abuse of notation, we denote by \( \pi(\succeq) \) the binary relation \( \{(\pi(x), \pi(y)) : x \succeq y\} \). The inverse of a preference relation \( \succeq \) is the binary relation \( \text{inv}(\succeq) = \{(y, x) : x \succeq y\} \). Given a subset \( S \) of \( X \), the restriction of \( \succeq \) on \( S \) is the binary relation \( \succeq_S = \{(x, y) \in S \times S : x \succeq y\} \cup \Delta_X \). Note that the restriction
$\succeq_S$ makes no comparison for distinct alternatives in $X \setminus S$, but it is reflexive on $X$. When $S$ is a finite set, say $S = \{x, y, z\}$, we write $\succeq_{xyz}$ instead of $\succeq_{\{x, y, z\}}$ for brevity.

3 The transitive core

This paper studies a method of inferring a welfare order over alternatives from observation of a preference relation. If the observed preference relation is transitive, the utilitarian welfare criterion suggests that replacing a less preferred alternative with a preferred alternative improves the decision maker’s welfare. However, this criterion brings a cyclic welfare order when the observed preference relation is cyclic. For example, suppose that the decision maker exhibits a preference relation on $\{x, y, z\}$ such that $x \succ y \succ z \succ x$. If she is initially endowed with $z$, then the criterion tells us that she would be better off if the alternative is replaced by $y$ and then by $x$. However, since the decision maker also prefers $z$ to $x$, she would end up with the initial alternative $z$ with her welfare being strictly improved. This is unrealistic. The utilitarian welfare criterion assumes the existence of a transitive preference relation, and it is not applicable when the observed preference relation is cyclic.

A preference relation is still informative in eliciting the decision maker’s welfare even when it is cyclic. For instance, consider a preference relation in Figure 1. We can easily see that the preferences of an alternative $x$ over the other alternatives are not involved in any preference cycles. Given the dominating preference of the alternative $x$, it seems reasonable to conclude that at least this part of the preference relation reveals the decision maker’s welfare. Note that this “cycle-free” criterion for welfare inference is applied to a cyclic preference relation.

Also, a method of welfare inference should be coherent across preference relations. For example, consider two cyclic preference relations $\succeq_1$ and $\succeq_2$ in Figure 2. These preference relations share the identical structure except for labeling of alternatives. In order to obtain a coherent rule of welfare inference, we may require that welfare comparisons between pairs of alternatives (say, $x$ and $y$) made for $\succeq_1$ coincide with those of corresponding pairs ($y$ and $z$, respectively) for $\succeq_2$. Such a requirement guides us to finding a sensible method of welfare inference.

In this section, I propose a method of inferring welfare orders from decision makers’ preference relations. I call this mapping from preference relations to welfare orders a welfare evaluation rule (WER). Then, we discuss properties of WERs including those argued above. The main result of the section shows that a set of properties introduced as normative requirements on WERs characterizes a
unique welfare evaluation rule. This unique rule will be called the *transitive core*.

**Incomplete inference.** One property that we could impose on welfare evaluation rules is inference of complete welfare orders. However, we will not adopt this property in this paper. This is because we assume no observability assumption beyond that of a complete preference relation. As the rules must infer welfare orders only by the structure of observed preference relations, there are cases where any welfare comparisons for some alternatives are just unreliable. For example, recall the preference relation in Figure 1. This relation is cyclic over three alternatives \( y, z, \) and \( w \). But what makes welfare inference further complicated is symmetry of the preference relation. Since there is no welfare ranking of these alternatives that makes better sense than the others, discreet methods of welfare inference should reserve their welfare comparisons. If we assume a particular model of nontran-
sitive preferences, it may be possible to propose a refined method of welfare in-
ference. Inference of incomplete welfare orders leaves room for further criteria
based on specific models of nontransitive preference relations.

3.1 Welfare evaluation rules

Let $X$ be the set of conceivably all alternatives of interest with $|X| > 3$. I assume
that there is a decision maker who has an unknown complete preference relation
on $X$, and we reveal this preference relation from available choice data. We allow
cases where our observation of the preference relation is limited to a subset of $X$.
(For example, this is the case when an available data set consists of choice obser-
vation from present alternatives whereas we are interested in the decision maker’s
welfare on intertemporal outcomes.) If $\succsim$ is the decision maker’s unknown pref-
erence relation on $X$, and a set $S \subseteq X$ is this restricted domain, then we observe a
preference relation $\succsim_S$ such that

$$x \succsim_S y \quad \text{if and only if} \quad x \succsim y \text{ and } x, y \in S$$

for any distinct $x, y \in X$. I study a method of welfare inference as a concept before
observation of a preference relation. This leads us to formulating it as a function
from the set of possibly observed preference relations to the set of welfare orders.
Define

$$\mathcal{P} := \{\succsim_S : \succsim \text{ is a complete binary relation on } X \text{ and } \emptyset \neq S \subseteq X\}.$$ 

This is the set of preference relations that we may observe. Then, a welfare eval-
uation rule (WER) is defined as a map $\sigma$ on $\mathcal{P}$ such that $\sigma(\succsim)$ is a preorder on $X$
for each $\succsim \in \mathcal{P}$. A WER represents a rule of welfare inference from preference re-
lations. As a minimal requirement, it must infer a reflexive and transitive welfare
order for every observed preference relation. These requirements are, however,
not sufficient for the concept to be meaningful. The following examples present
two extremes of welfare evaluation rules.

**Example** (Universal incomparability). The universally incomparable WER is a
map $\sigma_0$ that assigns the diagonal order $\Delta_X$ for every preference relation in $\mathcal{P}$. This
rule only says that every alternative has the identical welfare value as itself, and
it makes no nontrivial welfare judgment. While the rule does not make incorrect
welfare comparison, it is most certainly useless.
The universally indifferent WER is a map \( \sigma_1 \) that assigns the trivial weak order \( X \times X \) for every preference relation in \( \mathcal{P} \). This rule assumes that any pair of alternatives has the same value for the decision maker regardless of her preferences. Even when an observed preference relation is transitive, the rule ignores this information and simply grants the identical value to all alternatives. The universally indifferent WER is likely unacceptable.

As these examples suggest, the notion of welfare evaluation rules itself does not provide sensible methods of welfare inference. In order to find a reliable rule, we will examine some properties of WERs below. Note that properties of WERs describe how we should infer the decision maker’s welfare. In this sense, they are not positive but rather normative properties. For example, the classical utilitarian welfare criterion can be viewed as a property of WERs. Let us say that a welfare evaluation rule \( \sigma \) satisfies the utilitarian welfare criterion if \( \sigma(\preceq_S) = \preceq_S \) for every complete and transitive preference relation \( \preceq \) on \( X \) and every nonempty \( S \subseteq X \).

### 3.2 Properties of welfare evaluation rules

The utilitarian welfare criterion is not applicable if we observe a cyclic preference relation. Therefore, we need properties beyond this criterion to discuss welfare inference for nontransitive preference relations. For this, I propose six axioms of welfare evaluation rules below. These axioms apply for any complete preference relation \( \succeq \) on \( X \), any subset \( S \) of \( X \), and any permutation \( \pi \) on \( X \).

#### Axiom 1 (Prudence).
For any \( x, y \in X \), \( x \sigma(\preceq_S) y \) implies \( x \preceq_S y \).

#### Axiom 2 (Cycle-free).
If no cycle of \( \succeq \) involves its preference \( (x, y) \), then \( x \sigma(\succeq) y \).

#### Axiom 3 (Neutrality).
\( \sigma \circ \pi(\succeq) = \pi \circ \sigma(\succeq) \).

#### Axiom 4 (Inverse).
\( \sigma \circ \text{inv}(\succeq) = \text{inv} \circ \sigma(\succeq) \).

#### Axiom 5 (Reduction).
If \( x \sigma(\preceq_T) y \) and \( x, y \in S \subseteq T \), then \( x \sigma(\preceq_S) y \).

#### Axiom 6 (Extension).
If \( x \sigma(\preceq_{xyz}) y \) for all \( z \in X \), then \( x \sigma(\succeq) y \).

The first axiom, prudence, says that a welfare evaluation rule should make a welfare comparison only when the decision maker at least exhibits the same preference. In other words, if the decision maker strictly prefers \( x \) over \( y \), then a WER should not rank \( y \) higher than \( x \). Note that the WER can infer strict welfare ranking of alternatives between which the decision maker is indifferent. In this sense, the
axiom does not exclude a possibility of tie-breaking by welfare evaluation rules. Also, the axiom requires that WERs make no welfare comparison for alternatives on which we do not observe the decision maker’s preferences. When we observe a preference relation on a set $S$, WERs should be able to make nontrivial welfare comparisons only for alternatives in $S$.

The second axiom claims that observation of a preference $(x, y)$ should reveal the decision maker’s welfare if this preference is not involved in any preference cycles. For example, recall a preference relation in Figure 1. While this preference relation is cyclic, preferences of an alternative $x$ over the other alternatives are not involved by any preference cycles. Given the consistent structure of the relation, therefore, it would be reasonable to say that the decision maker is better off with the alternative $x$ over the others. The cycle-free axiom requires that WERs reject observed preferences for welfare comparison only if they cause preference cycles.

Note that, if an observed preference relation is transitive, a welfare evaluation rule under the cycle-free axiom infers that all observed preferences reflect the decision maker’s welfare. On the other hand, the prudence axiom ensures that WERs never infer a welfare order that contains unobserved preferences. Put together, if a WER $\sigma$ satisfies both axioms, then we have $\sigma(\preceq_S) = \preceq_S$ for any complete and transitive preference relation $\preceq$ on $X$ and any nonempty set $S \subseteq X$. Hence, the two axioms imply the utilitarian welfare criterion.

The neutrality axiom requires that welfare inference by WERs be independent of labeling of alternatives. Suppose that two complete preference relations $\preceq$ and $\preceq'$ on $X$ are identical upon relabeling of alternatives. (Formally, for some permutation $\pi$ on $X$, $x \preceq y$ iff $\pi(x) \preceq' \pi(y)$ for all $x, y \in X$. Figure 2 illustrates an example of such preference relations.) If welfare inference by a WER $\sigma$ is independent of labeling, then the welfare orders $\sigma(\preceq)$ and $\sigma(\preceq')$ inferred from these preference relations should have the identical structure upon the same relabeling. The axiom claims that $x \sigma(\preceq) y$ iff $\pi(x) \sigma(\preceq') \pi(y)$ for all $x, y \in X$.

Our forth axiom states that, if there are two decision makers whose observed preference relations are inverse to each other, then WERs should infer that their welfare orders are also inverse to each other. Given a preference relation $\preceq$ on $X$, its inverse relation is a binary relation $\text{inv}(\preceq)$ on $X$ such that $x \preceq y$ if and only if $y \text{inv}(\preceq) x$ for all $x, y \in X$. The axiom requires that $\sigma(\text{inv}(\preceq))$ be inverse of $\sigma(\preceq)$, that is, $\sigma(\text{inv}(\preceq)) = \text{inv}(\sigma(\preceq))$.

The reduction axiom imposes coherence of welfare inference when observation of a preference relation is limited to subsets of $X$. For the description of the axiom, we shall consider a special case where $T = X$. (The same interpretation applies for the general case.) Suppose that the decision maker has an unknown
preference relation $\succeq$ on $X$, and the observed choice data reveals her preference relation restricted on a set $S \subseteq X$. Note that any preference cycles of the revealed preference relation $\succeq_S$ are, of course, preference cycles of the decision maker’s preference relation $\succsim$. In this sense, the preference relation $\succeq$ is always at least as much cyclic as the observed relation $\succeq_S$. (For instance, it is possible to observe a transitive relation $\succeq_S$ when the decision maker’s preference relation $\succeq$ is cyclic.)

The axiom claims that, if a WER $\sigma$ would make a welfare comparison for a pair of alternatives in $S$ when we were to observe the full relation $\succeq$, then the rule should make the same welfare comparison when we observe the “less cyclic” relation $\succeq_S$. Note that there is no way to know the full relation $\succeq$ if the available data set only reveals her preference relation $\succeq_S$ on $S$. For example, if the observed relation $\succeq_S$ is transitive, then the decision maker appears rational as far as we can see.

Our last axiom also imposes coherence of welfare evaluation rules under limited observation of a preference relation. The axiom claims that, if a WER infers that an alternative $x$ is a welfare improvement over another alternative $y$ whenever we observe $\succeq_{xyc}$ for some $z \in X$, then this rule should also make the same welfare inference for the pair $(x, y)$ when we observe the “less cyclic” relation $\succeq_S$. To illustrate how this axiom works, let $X = \{1, 2, \ldots, N\}$ for some $N > 3$, and consider a preference relation $\succeq$ on $X$ such that

$$n \succsim m \text{ iff } [n \geq m \text{ and } (n, m) \neq (N, 1)] \text{ or } (n, m) = (1, N) \quad (1)$$

for any $n, m \in X$. The preference relation $\succsim$ is identical to the Euclidean order $\geq$ except that $1 > N$. (Of course, $\succeq$ is a transitive order.) Notice that, however, every preference of this relation is involved in some preference cycles. For example, a preference $(3, 2)$ is involved in a cycle $(3, 2, 1, N)$. So, the cycle-free axiom offers no implication for welfare inference. Given this structure, do we have to conclude that nothing about welfare can be learned from the preference relation $\succsim$?

Our intuition may suggest that the preference $(1, N)$ is the main cause of preference cycles. While the other preferences are also involved in preference cycles, they seem to contribute only indirectly for the cycles. In fact, outside observers can realize that the decision maker’s preference relation $\succeq$ is cyclic if, and only if, they observe her preference $(1, N)$. This means that, for example, an observed preference relation appears transitive whenever it is restricted on $\{x, y, z\}$ for any $x, y, z \in X$ with $N - 1 \geq x, y \geq 2$. Therefore, the presence of a preference $(1, N)$ is crucial to make the preference relation cyclic. We can show that, if a WER $\sigma$ admits the extension axiom and the utilitarian welfare criterion, then the rule must infer that $x$ is a welfare improvement over $y$ for any $x$ and $y$ with $N - 1 \geq x \geq y \geq 2$. 

13
The extension axiom allows WERs to make nontrivial welfare inference by identifying “less problematic” preferences.

In the appendix, I prove that the proposed axioms are mutually independent by showing welfare evaluation rules that satisfy all except one axiom. For example, the universally incomparable WER \( \sigma_0 \) satisfies all but the cycle-free axiom, and the universally indifferent WER \( \sigma_1 \) satisfies all but the prudence axiom. Note that even these trivial welfare evaluation rules satisfy almost all the axioms.

### 3.3 The transitive core

In this section, I introduce a certain WER called the transitive core. Once we observe a preference relation, the transitive core checks consistency of its preference with respect to third alternatives. For example, suppose that we observe a preference relation \( \succeq \) on \( X \) such that \( x \succeq y \) for some alternatives \( x \) and \( y \). The transitive core views this preference \((x, y)\) as inconsistent with respect to a third alternative \( z \in X \) if either \( y \succeq z \succ x \) or \( y \succ z \succeq x \) and as consistent if neither holds. If a given preference is consistent in this way with respect to every third alternative, the transitive core infers that the preference reflects the decision maker’s welfare.

Formally, the transitive core is a welfare evaluation rule \( c(\cdot) \) such that

\[
x \ c(\succeq_S) y \quad \text{if and only if} \quad \begin{cases} z \succeq_S x & \text{implies} & z \succeq_S y \quad \text{for every} \quad z \in X. \\ y \succeq_S z & \text{implies} & x \succeq_S z \end{cases}
\]

for any complete preference relation \( \succeq \) on \( X \), any subset \( S \) of \( X \), and any \( x, y \in X \). It is straightforward to show that the transitive core maps every preference relation in \( \mathcal{P} \) to a reflexive and transitive binary relation on \( X \). It is hence well defined as a welfare evaluation rule.

An important observation regarding the transitive core is that it satisfies all the axioms in the last section. Therefore, if we adopt the axioms as desired properties of WERs, then the transitive core is a possible candidate for a method of welfare inference. Indeed, the rule offers reasonable welfare comparisons in many cases. The next example demonstrates one of such cases. In Section 4, we further study implications of the transitive core on various models of nontransitive preferences.

**Example.** Recall the preference relation \( \succeq \) on \( X = \{1, 2, \ldots, N\} \) defined by (1). As we observed earlier, if \( N - 1 \geq x \geq y \geq 2 \), then the preference relation is transitive when it is restricted on three alternatives \( \{x, y, z\} \) for any \( z \in X \). This means that no \( z \) satisfies \( y \succeq z > x \) nor \( y > z \succeq x \), so the preference \((x, y)\) is consistent with respect to any third alternative. The transitive core thus infers that the preference
(x, y) reflects the decision maker’s welfare. In contrast, the preference (1, N) is inconsistent with respect to some (in fact, all) third alternatives. So, the rule rejects this preference from welfare comparison. After all, the transitive core induces a welfare ranking c(≿) such that

\[ x \ c(≿) \ y \quad \text{if and only if} \quad N - 1 \geq x \geq y \geq 2 \]

for any \( x, y \) in \( X \). Therefore, while all observed preferences make some cycles, the transitive core only reserve welfare comparison for “dubious” preferences. Notice that this example demonstrates a difference of welfare inference by the transitive core from simple removal of preference cycles.

What is further notable about the transitive core is that it is the only WER that satisfies all Axioms 1 through 6. Though there are many welfare evaluation rules that satisfy all but one axiom, a rule that satisfies all the axioms at the same time turns out to be unique. In this sense, the axioms characterize the transitive core.

**Theorem 1.** A WER satisfies Axioms 1-6 if and only if it is the transitive core.

The theorem implies that, if a WER satisfies all axioms, then we can uniquely identify the welfare order it infers for each observed preference relation. But it is worthwhile to note that the axioms pin down the inferred welfare order in tandem. For example, even determining welfare orders for transitive preference relations requires both Axiom 1 and Axiom 2. (Recall that the utilitarian welfare criterion is an implication of the two axioms.) Therefore, it would be of interest to examine how the axioms coordinate to characterize the welfare order for nontransitive preference relations. While the formal proof of the theorem is left to the appendix, the next paragraph illustrates derivation of the unique welfare order for a preference relation on three alternatives.

**Example.** Let \( \sigma \) be a WER that satisfies Axioms 1 through 6, and suppose that we observe a preference relation \( \succsim_1 \) in Figure 3. Theorem 1 then implies that \( \sigma(\succsim_1) = c(\succsim_1) \), which we will verify in this example. First, we can observe that preferences \((x, z)\) and \((z, y)\) are not involved in any preference cycle of the relation \( \succsim_1 \). So, Axiom 2 implies that \( x \sigma(\succsim_1) z \) and \( z \sigma(\succsim_1) y \), which in turn implies \( x \sigma(\succsim_1) y \) by transitivity. Now, consider \( \succsim_2 \) and \( \succsim_3 \) in Figure 3. The relation \( \succsim_2 \) is the inverse of the preference relation \( \succsim_1 \). The relation \( \succsim_3 \) has the identical structure as \( \succsim_2 \) upon relabeling of alternatives. Therefore, it follows from Axiom 3 and Axiom 4 that

\[
\begin{align*}
\text{Inverse} & \quad z \sigma(\succsim_1) x & \iff & \quad x \sigma(\succsim_2) z \\
\text{Neutrality} & \quad y \sigma(\succsim_3) z.
\end{align*}
\]
However, observe that $\succsim_1$ and $\succsim_3$ are the same preference relation. So, if $z \sigma(\succsim_1)x$, then we must have $y \sigma(\succsim_1)z$ and thus $y \sigma(\succsim_1)x$ by transitivity. Since the inference $y \sigma(\succsim_1)x$ would contradict with Axiom 1, we cannot have $z \sigma(\succsim_1)x$. The same argument also proves that $y \sigma(\succsim_1)z$ do not hold. This concludes the characterization of $\sigma(\succsim_1)$ as such that $\sigma(\succsim_1) = \{(x, y), (x, z), (z, y)\}$. It is straightforward via (2) to check that the transitive core $c(\succsim_1)$ obtains the same order.

Before proceeding, I remark that the assumption of completeness for the decision maker’s preference relation is crucial for Theorem 1. In general, if we extend the transitive core for incomplete preference relations by the same rule (2), then it does not even satisfy the cycle-free criterion. (For example, the transitive core maps some acyclic relations to the diagonal order.) In addition, if we extend the domain of welfare evaluation rules by allowing incomplete preference relations, then there is no welfare evaluation rule that satisfies all the axioms in the last section. Finding a reliable rule of welfare inference for incomplete preference relations and identifying a set of normative axioms that characterizes such a rule are both open questions.

4 Applications

In the last section, we show that the transitive core is the unique welfare evaluation rule that satisfies Axioms 1 through 6. While this characterization gives a certain normative justification for the rule, it does not necessarily imply that the rule is a useful concept. In this section, we will examine the performance of the transitive core by applying it to models of nontransitive preference relations of economic interest.
4.1 Semiorders: imperfect ability of discrimination

Nontransitive indifference due to imperfect ability of discrimination has long been studied in the literature. Armstrong [2, 3, 4, 5] poses a question on the assumption of transitive indifference and first introduces a utility model of imperfect discrimination. Luce [23] brings a notion of semiorders into economics and provides its axiomatic foundation. Subsequently, many generalizations of semiorders, such as interval orders by Fishburn [14], are developed in search of descriptive models of nontransitive indifference. In this section, we take $X$ as a connected metric space and consider the following representation of semiorders introduced by Luce.

**Definition.** A semiorder is a binary relation $\succeq$ on $X$ for which there is a pair $(u, \epsilon)$ of a continuous function $u: X \rightarrow \mathbb{R}$ and a nonnegative number $\epsilon \geq 0$ such that

$$x \succeq y \text{ if and only if } u(x) \geq u(y) - \epsilon$$

holds for all $x, y \in X$.

Obviously, if $\epsilon = 0$, the representation reduces to the standard utility representation. When $\epsilon > 0$, on the other hand, a semiorder shows cyclic preferences. Note that the representation implies that $x \sim y$ iff $|u(x) - u(y)| \leq \epsilon$. So, a semiorder shows an indifference between alternatives $x$ and $y$ even when their utility values are different. Luce refer to the coefficient $\epsilon$ as the just noticeable difference. Figure 4 illustrates the regions of preferred, indifferent, and less preferred alternatives to a given $x \in X$ when $X$ is a real line.

While a semiorder is cyclic in general, the “right” welfare order for the decision maker seems to be obvious in this context. Nontransitivity of the preference

![Figure 4: A semiorder representation](image-url)
relation is induced by imperfect perception. If it were possible to eliminate the perception error $\epsilon$, the decision maker would consistently evaluate alternatives by the utility function $u$. A ranking induced by the utility function is transitive, and it is a natural candidate for the decision maker’s welfare order. It turns out that the transitive core infers this utility order for each semiorder.

**Proposition 2.** Let $\succeq$ be a semiorder on $X$ with a representation $(u, \epsilon)$. Then,

$$x \succeq (\succeq) y \text{ if and only if } u(x) \geq u(y)$$

for any $x, y \in X$, provided that $\sup |u(x) - u(y)| > 2\epsilon$.\(^2\)

Note that the utility function is not observable. The proposition implies that the transitive core infers an unobservable welfare order associated with the utility function from an observed preference relation. Also, the inferred welfare order turns out to be complete in this case.

### 4.2 Time preferences

Let $Z$ be a nonempty open interval in $\mathbb{R}_+$, and let $X = Z \times [0, \infty)$. In this section, a generic member $(x, t)$ of $X$ is interpreted as a dated outcome where the decision maker receives a prize of $x$ dollars at time $t$. Correspondingly, I refer to a complete preference relation on $X$ as a time preference. Whereas there are many models of time preferences, the class of absolute discounting time preferences attracts a particular interest in the literature. These preference relations are represented as

$$(x, t) \succeq (y, s) \text{ if and only if } \delta(t)u(x) \geq \delta(s)u(y)$$

for each $(x, t), (y, s)$ in $X$ under some discounting function $\delta : [0, \infty) \to [0, 1]$ and utility function $u : Z \to \mathbb{R}$. The exponential discounting and hyperbolic discounting models are special cases of this class. Note that any absolute discounting time preference is transitive, for it has a utility representation $(x, t) \mapsto \delta(t)u(x)$.

Read [29] finds that a large fraction of subjects in an experiment appear inconsistent with the model of absolute discounting and are better explained by that of

---

\(^2\)The added condition is a necessary and sufficient condition for uniqueness the utility order. For example, suppose that $\sup |u(x) - u(y)| < \epsilon$. Then, since the semiorder is indifferent for every pair of alternatives, any function $u' : X \to \mathbb{R}$ with $\sup |u'(x) - u'(y)| < \epsilon$ would represent the same semiorder. If the given condition is satisfied, an ordinal ranking of the underlying utility function is uniquely determined.
subadditive discounting. The subadditive discounting model postulates that a discount factor for a delay is larger than the product of discount factors for subdelays that partition the original delay. To be more specific, let us introduce the following general representation of time preferences studied by Ok and Masatlioglu [28].

**Definition.** A time preference $\succeq$ is a relative discounting time preference if there exist continuous functions $u : \mathbb{Z} \to \mathbb{R}^+$ and $\eta : \mathbb{R}_+^2 \to \mathbb{R}^+$ that satisfy

$$(x, t) \succeq (y, s) \text{ if and only if } u(x) \geq \eta(s, t)u(y)$$

for each $(x, t), (y, s)$ in $X$, where $u$ is an increasing homeomorphism, $\eta(\cdot, t)$ is decreasing with $\eta(\infty, t) = 0$, and $\eta(t, s) = \eta(s, t)^{-1}$ for any $t, s \geq 0$.

The function $\eta$ is called a relative discount function, and it measures a relative discount factor for delays between any two points in time. So, the decision maker discounts a utility value of a prize at time $s$ by the factor $\eta(s, t)$ in order to compare it with that of another prize at time $t$. If $\eta(s, t) = \delta(s)/\delta(t)$, the relative discounting model reduces to the absolute discounting model. Given the representation, we can write the subadditive discounting model as a property on $\eta$:

$$\eta(r, t) \geq \eta(r, s)\eta(s, t) \quad \text{for every } r \geq s \geq t.$$  \hspace{1cm} (3)

Note that absolute discounting time preferences satisfy (3) with equality. (Indeed, it is a characterization of the absolute discounting model.) Read’s finding suggests that allowing (3) as inequality provides significant improvement in data fitting.

For another alternative to the absolute discounting model, Rubinstein [31] proposes a similarity-based time preference. This model postulates that the decision maker follows up to three steps of heuristic procedures in order to compare a pair of dated outcomes. In the first step, the decision maker looks for dominance of dated outcomes. So, if $x > y$ and $t < s$, then the dated outcome $(x, t)$ is preferred to $(y, s)$. Provided that there is no dominance between them, she next looks for a similarity of dated outcomes either in delivery dates or in prizes. For example, if the delivery dates are similar while the prizes are not, the decision maker chooses one that offers the larger prize. Lastly, if neither of the first two steps resolves her choice, another criterion is applied to compare $(x, t)$ and $(y, s)$. The author argues that this model explains observed data from an experiment better than the hyperbolic discounting model and is more intuitive as a description of the decision maker’s reasoning process. Ok and Masatlioglu [28] proves that a variety of similarity-based time preferences are, in fact, relative discounting time preference.
We can easily see that some relative discounting time preferences are not transitive. For example, consider a preference relation $\succsim$ represented under a relative discounting function $\eta$ such that

$$
\eta(s, t) = \begin{cases} 
\delta^{s-t} & \text{if } T \geq s - t \geq 0 \\
\delta^T & \text{if } s - t > T
\end{cases}
$$

for any $s$ and $t$, where $T > 0$ is some constant. The decision maker exponentially discounts a delay if it is shorter than $T$. However, she perceives any delays longer than $T$ as same and apply a constant discount factor for them. (This is an example of subadditive discounting. Note that $\eta$ satisfies (3).) Then, whenever $x$, $y$, and $z$ are prizes such that $(x, 0) \sim (y, T) \sim (z, 2T)$, we have $(z, 2T) \succ (x, 0)$.

This example of a nontransitive time preference is not a special case. Indeed, a relative discounting time preference turns out to be transitive if and only if it is an absolute discounting time preference ([28, Corollary 1]). Therefore, any empirical implications of the relative discounting model beyond that of the absolute discounting model are attributed to nontransitivity of time preferences. For better explanations of observed data offered by the subadditive discounting model or the similarity-based time preferences, allowing nontransitivity of preference relations is essential.

While the relative discounting model generalizes the class of time preferences to the extent that it contains cyclic relations, we can show that their welfare orders inferred by the transitive core have a representation under absolute discounting.

**Theorem 3.** Let $\succsim$ be a relative discounting time preference with a representation $(u, \eta)$. Then, there is a set $\mathcal{D}$ of continuous functions $\delta : \mathbb{R}_+ \rightarrow \mathbb{R}_{++}$ such that

$$(x, t) \preceq (y, s) \text{ if and only if } \delta(t)u(x) \geq \delta(s)u(y) \text{ for all } \delta \in \mathcal{D}$$

whenever $(x, t)$ and $(y, s)$ are dated outcomes in $X$.

The theorem shows that the transitive core of a relative discounting time preference is represented by multi absolute discounting functions. In the proof of the theorem, we show that the collection $\mathcal{D}$ consists of functions of the form $\eta(\cdot, r)$ for arbitrary fixed points $r \geq 0$ in time. Hence, the transitive core suggests that a dated outcome $(x, t)$ is a welfare improvement over $(y, s)$ if and only if the former has the higher discounted utility value than the latter regardless of a time at which they are evaluated.
4.3 Justifiable preferences

Let $\Omega$ be a finite nonempty set of states of the world, and $Y = \Delta(\mathbb{R})$ be the set of all lotteries over real prizes. An arbitrary function $f : \Omega \to Y$ mapping each state to a lottery is referred to as an Anscombe and Aumann [1] act. I denote the collection of all acts by $X$. An act is a description of state-contingent prize schedules, where objective likelihood of each state is not known.

Consider a group of agents each of whom has a subjective belief about likelihood of the states. To study justifiable collective decision making in this context, Lehrer and Teper [21] propose a model where the group prefers one act over another if and only if at least one agent in the group has the higher expected utility for the former act than the latter. We say that a preference relation $\succeq$ on $X$ is justifiable if there exist a continuous affine monotone utility function $u : Y \to \mathbb{R}$ and a nonempty closed convex set $P$ of probability distributions over $\Omega$ such that

$$f \succeq g \quad \text{if and only if} \quad \exists p \in P \text{ s.t. } \sum_{\omega \in \Omega} p(\omega)u(f(\omega)) \geq \sum_{\omega \in \Omega} p(\omega)u(g(\omega))$$

for any two acts $f$ and $g$.

While the agents share the same utility function $u$ over lotteries, they can disagree on evaluations of acts due to different subjective beliefs about the states of the world. In particular, it is possible that, for three acts $f, g, h \in X$, some agents prefer $f$ over $g$, and $g$ over $h$, whereas the others prefer $h$ over $f$, and $f$ over $g$. The justifiable preference of this group is such that $f \sim h \sim g$ and $f \succ g$. Therefore, a justifiable preference relation is not transitive in general ([21, p.763]).

Lehrer and Teper contrast justifiable preferences with Knightian preferences by Bewley [9]. Under Knightian preferences, the group prefers one act over another if and only if all agents in the group have the higher expected utility for the former act than the latter. Formally, a preference relation $\succeq$ on $X$ is called Knightian if there exist a continuous affine monotone utility function $u : Y \to \mathbb{R}$ and a nonempty closed convex set $P$ of probability distributions over $\Omega$ such that

$$f \succeq g \quad \text{if and only if} \quad \sum_{\omega \in \Omega} p(\omega)u(f(\omega)) \geq \sum_{\omega \in \Omega} p(\omega)u(g(\omega)) \quad \forall p \in P$$

Throughout the section, a Borel probability measure on $\mathbb{R}$ is called a lottery. I endow the set $Y$ of lotteries with the weak topology and the set $X$ with the product topology. A function $u : Y \to \mathbb{R}$ is said to be monotone if $u(\mu) > u(\nu)$ whenever $\mu$ first-order stochastic dominates $\nu$.

The model of justifiable preferences studied in this section differs from the original work by Lehrer and Teper in two points. First, I take the set $\mathbb{R}$ of real prizes for the outcome space of lotteries. Second, the utility function is assumed to be monotone. For a result of this section, we can weaken these assumptions so long as the utility function $u$ remains locally nonsatiable.
for any two acts \( f \) and \( g \). It is straightforward to show that a Knightian preference relation is not necessarily complete, but it is always transitive. Lehrer and Teper remark that a justifiable preference relation is a completion of a Knightian preference relation: if \( \succeq_J \) and \( \succeq_K \) are respectively justifiable and Knightian preferences associated with the same pair \((u, P)\) of a utility function and a set of subjective beliefs, then \( f \succeq_K g \) implies \( f \succeq_J g \).

A justifiable preference relation may have preference cycles, but it offers complete comparisons of acts and never contradicts with a Knightian preference relation. A Knightian preference relation provides an incomplete but transitive ranking over acts unanimously supported by the agents. The transitive core associates these two models in the straight manner.

**Proposition 4.** The transitive core of a justifiable preference relation is a Knightian preference relation under the same pair \((u, P)\) of a utility function and a set of beliefs.

If we observe a justifiable preference relation from a group of agents, the welfare order of this group may not be clear due to presence of preference cycles. The transitive core suggests that an act will improve the group’s welfare over another act when all agents in the group unanimously prefers the former act to the latter. Also, notice that we do not have to know a representing pair \((u, P)\) of an observed preference relation to apply the transitive core. As long as we observe the group’s preference relation, the transitive core infers the unanimity ranking by the agents.

### 4.4 Regret theory

Let \( \{1, \ldots, n\} \) be a finite set of states of the world with \( n \geq 3 \). Suppose that there is a nature that resolves a state according to a probability distribution \( p \) such that \( p_i > 0 \) for every \( i \in \{1, \ldots, n\} \) and \( \sum_{i=1}^{n} p_i = 1 \). A prospect is a real valued function on the set of the states, and it is interpreted as a state-contingent prize schedule delivered to a decision maker. Let \( X = \mathbb{R}^n \) be the set of all prospects, and we shall consider preference relations on \( X \) in this section.

The main body of economic analysis under uncertainty relies on the expected utility theory developed by von Neumann and Morgenstern [39]. It postulates that prospects are compared by their expected utility values \( \mathbb{E}(u \circ x) \) for some real function \( u : \mathbb{R} \rightarrow \mathbb{R} \). This theory has been acknowledged as the model of rational decision making under uncertainty and is justified from the normative perspective.

However, experimental studies find disparities between observed behavior and the prediction of the expected utility theory. The celebrated work by Kahneman
and Tversky [18], for example, provides extensive evidence that shows that subjects violate the expected utility hypothesis in consistent manners. Some of these violations are known as the certainty effect, the common consequences effect (or Allais paradox), and the isolation effect.

To accommodate the observed violations of the expected utility theory, Loomes and Sugden [22] propose an alternative theory of decision making that reflects the experience of regret. The regret theory takes into account that the decision maker may regret or rejoice for the chosen prospect upon realization of a state. Specifically, when a state is resolved, the decision maker may regret (rejoice) if an outcome of the chosen prospect happens to be worse (better) than that of the alternative prospect. The theory postulates that the psychological factor as such affects ex ante tastes over prospects by introspection. In this section, we consider the following representation of a regret preference by Loomes and Sugden.

**Definition.** A regret preference is a preference relation on $X$ for which there exist two continuous functions $u : \mathbb{R} \to \mathbb{R}$ and $Q : \mathbb{R} \to \mathbb{R}$ that satisfy

$$x \succeq y \quad \text{if and only if} \quad \sum_{i=1}^{n} p_i Q(u(x_i) - u(y_i)) \geq 0$$

for every $x, y \in X$, where $u$ is increasing homeomorphism with $u(0) = 0$, and $Q$ is convex, strictly increasing and satisfies $Q(-a) = -Q(a)$ for all $a \geq 0$.

The function $Q$ measures the decision maker’s regret/rejoice due to difference in realized prizes of two prospects. The model of regret preferences reduces to an expected utility representation when the function $Q$ is linear. Loomes and Sugden show that regret preferences robustly explain the observed choice anomalies under an assumption of strict convexity of the function $Q$.

We can show that a regret preference is not transitive whenever an associated function $Q$ is strictly convex. For example, let $n = 3$ and $p_1 = p_2 = p_3 = 1/3$, and consider three prospects $x, y, z$ that give utility values as in Table 1. Then,

$$\sum_{i=1}^{3} p_i Q(u(x_i) - u(y_i)) = \frac{1}{3} Q(2) - \frac{1}{3} Q(1) - \frac{1}{3} Q(1) > 0$$

<table>
<thead>
<tr>
<th>State</th>
<th>$u(x_i)$</th>
<th>$u(y_i)$</th>
<th>$u(z_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1: Utility for $i = 1, 2, 3$
and thus $x \succ y$. The decision maker prefers $x$ to $y$ since the large rejoice from the choice of $x$ over $y$ at state 1 is more than enough to compensate the small regrets at state 2 and 3. The same argument applies to pairs $(y, z)$ and $(z, x)$ by symmetry, resulting in a preference cycle $x \succ y \succ z \succ x$.

Indeed, this observation regarding nontransitivity of regret preferences holds in general. Bikhchandani and Segal [10] show that a regret preference is transitive if and only if it admits an expected utility representation. Therefore, any descriptive power of the regret theory beyond that of the classical expected utility theory is inseparable from the presence of preference cycles.

When a decision maker is endowed with a regret preference, her welfare order is hence unclear due to nontransitivity of the preference relation. Then, we might find the transitive core useful for inferring the decision maker’s welfare. In favor of such an expectation for the transitive core, the rule consistently infers that a prospect is a welfare improvement over another prospect whenever the former state-wise dominates the latter.

**Proposition 5.** Let $\succsim$ be a regret preference on $X$. If $x_i \succeq y_i$ for every state $i$, then $x \succsim y$. If, in addition, $x_i > y_i$ for some state $i$, then $y \succsim x$ does not hold.

This proposition shows that the transitive core infers a welfare order for a pair of prospects when one prospect dominates the other prospect. In general, we can show that the transitive core also offers welfare comparison for two prospects that do not dominate each other. However, a characterization of the transitive core for regret preferences is not known and left as an open question.

### 4.5 Majority voting

Let $n$ be an arbitrary natural number representing the number of voters in a society, and $X$ be a set of policies to be chosen. In this section, we shall consider a society as a representative decision maker and examine its preference relation induced by the majority voting rule. Suppose that an individual voter $i$, for each $i \in \{1, \ldots, n\}$, evaluates policies according to a linear order $\succsim_i$ on $X$ and has an equal share of votes. It is well known that a social preference induced by the majority criterion fails transitivity in general. Arrow [6] gives the celebrated impossibility theorem that proves that there is no voting system satisfying certain desirable criteria at the same time. Given the inevitable ambiguity in policy evaluations, the social choice theory has been extensively studied in the literature. The purpose of this section is to examine implications of the transitive core on social preferences.
Ranking $\preceq_1 \preceq_2 \preceq_3$

<table>
<thead>
<tr>
<th></th>
<th>$x$</th>
<th>$w$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x$</td>
<td>$w$</td>
<td>$z$</td>
</tr>
<tr>
<td>2</td>
<td>$y$</td>
<td>$x$</td>
<td>$w$</td>
</tr>
<tr>
<td>3</td>
<td>$z$</td>
<td>$y$</td>
<td>$x$</td>
</tr>
<tr>
<td>4</td>
<td>$w$</td>
<td>$z$</td>
<td>$y$</td>
</tr>
</tbody>
</table>

Table 2: Individual preferences

**Definition.** A majority preference is a preference relation $\preceq$ on $X$ defined by

$$x \preceq y \text{ if and only if } \left| \{ i : x \preceq_i y \} \right| \geq \left| \{ i : y \preceq_i x \} \right|,$$

for every $x$ and $y$ in $X$.

Note that a majority preference can also be interpreted as a preference relation of an *individual* decision maker over alternatives with $n$ many attributes. For every $i \in \{1, \ldots, n\}$, the decision maker has a ranking $\preceq_i$ of the alternatives according to the $i$th attribute. The majority preference arises when she prefers one alternative to another if and only if the former beats the latter by the number of attributes in which the former is ranked higher than the latter. In this case, the transitive core applied to the majority preference infers the individual’s welfare order.

A majority preference is cyclic in general. Table 2 gives an example of voters’ preference relations when $n = 3$ and $X = \{x, y, z, w\}$. A majority preference of this society has a preference cycle $x \succ z \succ w \succ x$, making desirability of each policy unclear. Many voting criteria have been proposed to study an optimal choice for the society. Some of them are listed below.

**Pareto criterion.** We say that a policy $x$ is a *Pareto improvement* over another policy $y$ if every voter in the society prefers $x$ over $y$. The Pareto criterion requires that $y$ be not chosen over $x$ when $x$ is unanimously preferred over $y$.

**Condorcet principle.** A *Condorcet winner* is a policy $x \in X$ that beats every other policy under pairwise majority voting. The principle claims that a Condorcet winner is an optimal choice for the society. The Condorcet winner may not exists, but it is always unique if it exists.

**Smith’s principle [37].** Let $S$ and $T$ be a partition of $X$ such that every policy in $S$ beats every policy in $T$ under pairwise majority voting. This principle suggests that an optimal social choice be made from the set $S$.  

25
Exclusive Condorcet principle. A set \( \{ x \in X : y \succ x \text{ for no } y \in X \} \) consists of all undominated policies under pairwise majority voting. The exclusive Condorcet principle assumes that the optimal social choice can be found in this set of policies if it is nonempty.

Pareto criterion and Condorcet principle are commonly viewed as desirable properties of social choice. Indeed, these criteria are often used as touchstones for evaluating voting systems. Smith’s principle and the exclusive Condorcet principle imply, but not implied by, Condorcet principle. Fishburn [15] argues, in favor of Smith’s principle, that “I find it hard to imagine an argument against Smith’s Condorcet Principle that would not also be an argument against Condorcet’s Principle.” The social welfare order inferred by the transitive core turns out to satisfy all principles except for the exclusive Condorcet principle. In the proposition below, I denote by \( c^*(\succ) \) the strict part of the transitive core, so that \( x c^*(\succ) y \) means \( x c(\succ) y \) and not \( y c(\succ) x \) for any \( x, y \in X \).

Proposition 6. The transitive core of a majority preference satisfies

(a) Pareto criterion: if \( x \) is a Pareto improvement over \( y \), then \( x c^*(\succ) y \),

(b) Condorcet principle: if \( x \) is a Condorcet winner, \( x c^*(\succ) y \) for all \( y \in X \setminus \{ x \} \),

(c) Smith’s principle: if \( S \) and \( T \) are a partition of \( X \) such that \( x \succ y \) for every \( x \in S \) and \( y \in T \), then \( x c^*(\succ) y \) for any \( x \in S \) and \( y \in T \),

but not the exclusive Condorcet principle.

The next example verifies that the transitive core violates the exclusive Condorcet principle in general. While there are many reasons to support the criterion, the example suggests that the exclusive Condorcet principle might exclude the choice of an attractive policy in some cases.

Example. Consider a society where a policy is chosen from \( \{ x, y, a_1, \ldots, a_k \} \) for \( 2N + 2 \) citizens (with a large number \( N \)). Let us assume that voters’ preferences of this society are distributed as in Table 3. We can observe that everyone except one voter ranks a policy \( x \) second or higher, whereas \( y \) is a controversial policy that splits the society about in half. Under the exclusive Condorcet principle, the social choice is uniquely determined by \( y \) eliminating a possibility of choosing an attractive “second best” policy \( x \). The transitive core reserves the welfare comparison between \( x \) and \( y \) and, thus, leaves room to choose the policy \( x \).
In the context of social choice, many alternative methods of welfare inference are proposed in the literature, including those that offer complete welfare rankings over policies. For example, Rubinstein [31] studies a ranking of policies induced by the point system. We can define the point system as a WER $\sigma$ such that

$$x \sigma(z) y \text{ if and only if } |\{z : x \succeq z\}| \geq |\{z : y \succeq z\}|$$

for every $x$ and $y$ in $X$, assuming that $X$ is finite. Then, $\sigma(z)$ is clearly a complete and transitive order on $X$. Moreover, $\sigma(z)$ is a completion of the transitive core so that $x c(z) y$ implies $x \sigma(z) y$ and that $x c^*(z) y$ implies $x \sigma^*(z) y$, where $c^*(z)$ and $\sigma^*(z)$ denote the strict parts of $c(z)$ and $\sigma(z)$, respectively. Therefore, the point system always agrees with welfare comparisons by the transitive core, whereas it offers a complete welfare ranking at the same time. However, this does not necessarily imply that the transitive core is an inferior method of welfare inference. For example, the point system infers that the controversial policy $y$ is a strict welfare improvement over the second-best policy $x$ in the example of Table 3. We might still find the transitive core useful as it identifies a more reliable part of welfare comparisons by the point system.

### Table 3: Exclusive Condorcet principle

<table>
<thead>
<tr>
<th>ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N)$ $x$ $a_1$ $\ldots$ $a_k$ $y$</td>
</tr>
<tr>
<td>$(N + 1)$ $y$ $x$ $a_1$ $\ldots$ $a_k$</td>
</tr>
<tr>
<td>$(1)$ $a_k$ $\ldots$ $a_1$ $y$ $x$</td>
</tr>
</tbody>
</table>

5 Unambiguous welfare improvement

Reflecting a recent interest in behavioral models of decision making, behavioral welfare economics has attracted attention in the literature. For example, Bernheim and Rangel [7, 8] develop a framework of behavioral welfare analysis that largely extends the classical utilitarian approach. They shed light on a generalized choice situation in which a decision maker might make a different choice from the same choice set contingent on ancillary conditions. Given such extended choice data, they introduce a method of welfare inference that captures an unambiguous part of welfare comparison. In this section, I compare implications of the transitive core with that of their unambiguous welfare improvement order.
Let $X$ be a nonempty set of alternatives, $\mathcal{A}$ the collection of all nonempty subsets of $X$, and $Z$ a nonempty set. We interpret $\mathcal{A}$ as the collection of choice sets and $Z$ as the set of ancillary conditions. We call $C : \mathcal{A} \times Z \rightarrow X$ an extended choice correspondence if $C(S, z) \subseteq S$ for all $S \in \mathcal{A}$ and $z \in Z$. The set $C(S, z)$ consists of alternatives chosen by the decision maker when she faces a choice set $S$ and an ancillary condition $z$. Given observation of an extended choice correspondence $C$, we say that an alternative $x$ in $X$ is an unambiguous welfare improvement over another alternative $y$ if presence of the alternative $x$ blocks choice of the alternative $y$ regardless of choice sets and ancillary conditions, that is, $y \notin C(S, z)$ for every $S \in \mathcal{A}$ and $z \in Z$ with $x \in S$.

The unambiguous welfare improvement order has a clear advantage over the transitive core in its applicability. Unlike the transitive core, the former rule does not require an assumption that the choice is made by a decision maker endowed with a nontransitive preference relation. In fact, it imposes no structural assumptions on observed choice behavior. (In particular, it does not assume that $C(\cdot, z)$ is rationalizable for each $z \in Z$. So, behavioral models under the standard choice framework are special cases of this model where $C(\cdot, z) = C(\cdot, z')$ for any $z, z' \in Z$.) The unambiguous welfare improvement order allows us to infer the welfare of a decision maker as long as her choice behavior is observable.

For comparing the transitive core with the unambiguous welfare improvement order, I restrict our attention to choice under nontransitive preference relations. To be specific, consider an extended choice correspondence $C$ such that there exists a complete but not necessarily transitive preference relation $\succeq$ on $X$ such that

$$C(S, z) = \{x \in S : x \text{ is indirectly } \succeq \text{-preferred to } y \text{ in } S \text{ for all } y \in S\}$$

for any $S \in \mathcal{A}$ and $z \in Z$. Recall that this is the top-cycle choice model discussed in Section 1 where the decision maker makes "rationalizable" choice under a nontransitive preference relation. This provides a common ground to compare the two rules, that is, between the unambiguous welfare improvement order for $C$ and the transitive core of the preference relation $\succeq$.

It turns out that, when applied to the choice correspondence (4), an alternative $x$ is an unambiguous welfare improvement over another alternative $y$ if, and only if, a preference $(x, y)$ is not involved in any preference cycles. Thus, by the cycle-free axiom, the implication of the unambiguous welfare improvement order is always weaker than that of the transitive core. Moreover, the difference of implications might be significant in some cases. As we saw earlier, a preference relation $\succeq$ defined by (1) has no preference that is free from preference cycles,
whereas the transitive core infers a nontrivial welfare order. Not only this simple example but also other preference relations of interest have the same feature. We can show that semiorders, justifiable preferences, and relative discounting time preferences discussed in Section 4 do not have any preference that is free from preference cycles unless the whole relations are transitive. The unambiguous welfare improvement order will be an empty relation if it is applied to choice under any of these preference relations.

However, this should not be viewed as a weakness of the unambiguous welfare improvement order. After all, the rule is applied to a very special case of its scope where the choice is induced by a nontransitive preference relation. Moreover, its implication coincides with the cycle-free axiom which I argue in this paper as a reasonable requirement for methods of welfare inference. The implications of the unambiguous welfare improvement order might be weak for some cases, but they are reliable at least. On the other hand, the transitive core does not offer as much applicability as the unambiguous welfare improvement order does. It restricts the scope of welfare analysis to cases where a decision maker’s choice is induced by a nontransitive preference relation. However, it takes advantage of its assumption. The assumption of choice under nontransitive preference relations lets us discuss normative properties of welfare inference beyond the cycle-free axiom. This leads to the axioms we introduced in Section 3, followed by its characterization as the transitive core. As we studied in Section 4, the rule offers nontrivial inference of the decision maker’s welfare in many contexts of interest.

6 Related literature

This paper relates to three branches of the economic literature. The first branch is descriptive studies that provide supporting evidence for choice under nontransitive preference relations. The second is the literature in behavioral welfare economics which develop welfare inference rules for behavioral decision makers. The third is the social choice theory which provides welfare analysis for a group of individuals who do not share the same interest.

Descriptive analysis. The descriptive studies of nontransitive preference relations provide evidence of choice under nontransitive preferences and develop models that explain such choice data. Several works in this branch were already cited in Section 4 while we examine the performance of the transitive core. The theory of nontransitive preference relations has a long history, but it attracts substantial
interest in the recent literature. Manzini and Mariotti [25] study choice by lexicographic applications of semiorders and give its axiomatic characterization. For the regret theory, Hayashi [17] studies choice of prospects by a decision maker whose motive is driven by anticipated regrets. In particular, he proposes a model which is flexible enough to allow both regret aversion and nontrivial likelihood judgement.

Note that the development in this branch of the literature motivates the purpose of this paper. Seeking a reliable method of welfare inference from nontransitive preference relations is of particular interest because of the presence of supporting evidence of choice under nontransitive preferences. In addition, the descriptive studies often offer methods of revealed preferences for choice under nontransitive preferences. Such methods embody an observability assumption for the decision maker’s preference relation, which I adopt throughout this paper.

**Behavioral welfare economics.** Bernheim and Rangel [7, 8] study a choice environment in which a decision maker’s choice is affected by ancillary conditions and develop a method of welfare inference that extends the utilitarian criterion. (See Section 5 for further discussion.) Likewise, Rubinstein and Salant [34] study an individual decision maker whose behavior is affected by choice frames. They assume a set of preference relations as observable data (each of which accounts for the decision maker’s behavior under some frame) and seek an unobservable welfare order that underlies her behavior. K˝oszegi and Rabin [20] discuss a method to identify choice observations that are due to the decision maker’s mistakes. Using an example of the gambler’s fallacy, they argue that understanding one’s mistakes helps us analyze her welfare. Chambers and Hayashi [12] examine welfare criteria that apply for random choice data. With the formulation of welfare inference rules as mappings from random choice data into weak orders, they study implications of normative properties on such mappings.

These works and this paper go parallel for the same goal. All are motivated to seek reliable methods of welfare inference when a decision maker is observed to be a non-utility maximizer. However, none nests another work as each differs in choice environments, aspects of behavioral choice, and proposed welfare criteria.

**Social choice theory.** In social choice theory, many studies are devoted for analyzing voting systems and collective welfare criteria. The theory provides a foundation for evaluating policies that affect welfare of each member of the society. While this branch of the literature discusses inference of social welfare, we should
note that it typically adopts the rationality assumption for individual members of the society. Nontrivial questions regarding social welfare arise due to aggregation of their preferences rather than evaluation of individual welfare. The purpose of the theory differs from that of behavioral welfare economics at this point, and the two fields are mostly disjoint in their scopes and results.

Having mentioned that, there are some results in the social choice theory that we may use to study welfare of individual decision makers. For this paper, an important example is the covering order by Fishburn [15] and Miller [27]. For any preference relation $\succeq$ on a set $X$, and for every $x, y \in X$, we say that $x$ covers $y$ if $y \succeq z$ implies $x \succeq z$ for all $z \in X$. Then, the uncovered set is defined as the set of alternatives in $X$ that are not strictly covered by any other alternative.\(^5\) The authors show that the uncovered set has desirable properties of social choice and that well known voting systems choose a policy from this set.

By the condition (2), we can easily observe that the rule of the covering order resembles that of the transitive core. Therefore, we might be inclined to argue that the covering order is just a similar method of welfare inference that is potentially useful for analysis of individual welfare. It turns out that the slight difference of the two rules brings quite different implications for analysis of both social welfare and individual welfare. For example, Fishburn [15] shows that the covering order satisfies the exclusive Condorcet principle, whereas the transitive core does not as we proved in Section 4.5. For analysis of individual welfare, the uncovered set of a semiorder may contain alternatives that do not maximize the underlying utility function $u$. In contrast, the transitive core of a semiorder always agrees with the ranking induced by the utility function (Proposition 2). Therefore, in spite of their apparent resemblance, these two methods of welfare inference differ not only in their motivations but also in their implications.

7 Concluding remarks

This paper studies methods of welfare inference from nontransitive preference relations. We formulate such methods as a concept we call welfare evaluation rules, that is, functions mapping complete preference relations to transitive welfare orders. We discuss certain normative properties of welfare evaluation rules, and the transitive core is characterized as a unique welfare evaluation rule that satisfies all

\(^5\)Galaabaatar and Karni [16] and Karni [19] study a rule similar to the covering order for the theory of individual preferences. They use the concept to distinguish indifference and indecisiveness for a decision maker endowed with a transitive but incomplete preference relation.
the properties. The transitive core infers that an alternative $x$ is a welfare improvement over another alternative $y$ when the decision maker’s preference $(x, y)$ stands consistent with respect to arbitrary third alternatives.

We examine implications of the transitive core for various models of nontransitive preference relations. It is shown that, when we apply the rule to a semiorder, the transitive core infers an underlying utility function that measures the values of alternatives for the decision maker. For preferences over intertemporal outcomes, the transitive core of relative discounting time preferences admits a representation by multi absolute discounting functions. In other contexts, we study applications of the rule for preferences over ambiguous acts (justifiable preference relations), uncertain prospects (regret preference relations), and policy alternatives (majority preferences of the society). These examinations verify that the transitive core offers nontrivial and reasonable inference of welfare in respective contexts.

I remark that this paper focuses on nontransitive preference relations for the source of behavioral decision making. This feature distinguishes the current paper from works of Bernheim and Rangel [8] and others discussed in Section 6 which assume different reasons of behavioral decision making such as framing effects, choice mistakes, and random preferences. However, these aspects do not by any means exhaust possible causes of behavioral decision making. Finding intuitive and reliable methods of welfare inference for other sources of behavioral decision making is an open problem of interest in behavioral welfare economics.

Appendix A: Proofs

Proof of Theorem 1. (If) We show that the transitive core meets the axioms. Let $\succeq$ be a complete preference relation on $X$, $S$ a subset of $X$, and $\pi$ a permutation on $X$.

Prudence. Take any $x, y \in X$. As $\succeq_S$ is a reflexive binary relation on $X$, we have $x \succeq_S x$. Then, by (2), $x c(\succeq_S) y$ implies $x \succeq_S y$.

Cycle-free. Let $(x, y)$ be a preference of $\succeq$ which is involved in none of its cycles. By contradiction, suppose that $x c(\succeq) y$ does not hold. Then, there is a $z \in X$ that meets either $y > z > x$ or $y > z > x$. In both cases, the cycle $(x, y, z)$ involves the preference $(x, y)$, a contradiction.

Neutrality. Recall that we denote by $\pi(R)$ the binary relation $\{ (\pi(x), \pi(y)) : x R y \}$ for any binary relation $R$ on $X$. So, $x \pi(R) y$ iff $\pi^{-1}(x) R \pi^{-1}(y)$ for any $x$ and $y$ in
Then, a statement
\[
\begin{aligned}
\{ z \pi(\geq) x \text{ implies } z \pi(\geq) y \\
y \pi(\geq) z \text{ implies } x \pi(\geq) z 
\}
\text{ for every } z \in X
\end{aligned}
\tag{5}
\]
is equivalent to a statement
\[
\begin{aligned}
\{ \pi^{-1}(z) \geq \pi^{-1}(x) \text{ implies } \pi^{-1}(z) \geq \pi^{-1}(y) \\
\pi^{-1}(y) \geq \pi^{-1}(z) \text{ implies } \pi^{-1}(x) \geq \pi^{-1}(z) 
\}
\text{ for every } z \in X.
\end{aligned}
\tag{6}
\]
But, since \( \pi^{-1} \) is bijective, the statement (6) is equivalent to
\[
\begin{aligned}
\{ z \geq \pi^{-1}(x) \text{ implies } z \geq \pi^{-1}(y) \\
\pi^{-1}(y) \geq z \text{ implies } \pi^{-1}(x) \geq z 
\}
\text{ for every } z \in X.
\end{aligned}
\tag{7}
\]
Then, \( x \circ \pi(\geq) y \iff (5) \iff (7) \iff \pi^{-1}(x) \circ (\geq) \pi^{-1}(y) \iff x \pi \circ c(\geq) y \), as desired.

**Inverse.** For any \( x \) and \( y \) in \( X \), a statement
\[
\begin{aligned}
\{ z \inv(\geq) x \text{ implies } z \inv(\geq) y \\
y \inv(\geq) z \text{ implies } x \inv(\geq) z 
\}
\text{ for every } z \in X
\end{aligned}
\tag{8}
\]
is equivalent to a statement
\[
\begin{aligned}
\{ x \geq z \text{ implies } y \geq z \\
z \geq y \text{ implies } z \geq x 
\}
\text{ for every } z \in X.
\end{aligned}
\tag{9}
\]
Then, \( x \circ \inv(\geq) y \iff (8) \iff (9) \iff y \circ (\geq) x \iff x \inv \circ c(\geq) y \).

**Reduction.** Suppose that \( x \circ c(\geq) y \) for every \( z \in S \subseteq T \). Let \( z \in X \) be such that \( z \geq_S x \). Then, by definition of the restricted relation \( \geq_S \), we have \( z \in S \) and \( z \geq_T x \). As \( x \circ c(\geq) y \), it follows that \( z \geq_T y \) and \( z \geq_S y \). A similar argument shows that any \( z \in X \) with \( y \geq_S z \) satisfies \( x \geq_S z \). So, \( x \circ c(\geq_S) y \).

**Extension.** Suppose that \( x \circ c(\geq) y \) for every \( z \in S \). Let \( z \in X \) be such that \( z \geq x \). Then, \( z \geq_{xy} x \), and thus the hypothesis implies \( z \geq_{xy} y \) and \( z \geq y \). Similarly, we can show that \( y \geq z \) implies \( x \geq z \) for any \( z \in X \). So, \( x \circ c(\geq) y \).

(Only if) Above we have shown that the transitive core is a WER that satisfies Axioms 1 through 6. For this part, it suffices to prove uniqueness of such a WER. We first note that a WER is characterized by its welfare inference for preference relations restricted on domains of three alternatives if it meets all the axioms.
Claim 1. Let \( \sigma \) and \( \sigma' \) be two welfare evaluation rules that meet Axiom 1 through Axiom 6. If \( \sigma(\succeq_{xyz}) = \sigma'(\succeq_{xyz}) \) holds for any complete preference relation \( \succeq \) on \( X \) and any \( x, y, z \in X \), then \( \sigma = \sigma' \).

Proof of Claim 1. Let \( \sigma \) and \( \sigma' \) be two WERs that meet the hypothesis. Take any complete preference relation \( \succeq \) on \( X \) and any set \( S \subseteq X \). Define a complete binary relation \( \succeq_S := \succeq_S \cup (X \times X \setminus S) \) on \( X \), and observe that (a) \( \succeq_S = \succeq_S \) and (b) \( \succeq_{xyz} \) is transitive if \( x, y \in S \) and \( z \in X \setminus S \). For any \( x, y \in S \), we can follow implications

\[
\begin{align*}
x \sigma(\succeq_S) y \Rightarrow & \quad x \sigma(\succeq_{xyz}) y \text{ for any } z \in S \quad \text{by Axiom 5} \\
\Rightarrow & \quad x \sigma(\succeq_{xyz}) y \text{ for any } z \in S \quad \text{by (a)} \\
\Rightarrow & \quad x \sigma(\succeq_{xyz}) y \text{ for any } z \in X \quad \text{by Axiom 2 and (b)} \\
\Rightarrow & \quad x \sigma(\succeq_S) y \quad \text{by Axiom 6} \\
\Rightarrow & \quad x \sigma(\succeq_S) y \quad \text{by Axiom 5} \\
\Rightarrow & \quad x \sigma(\succeq_S) y \quad \text{by (a)}
\end{align*}
\]

to prove an equivalence \( x \sigma(\succeq_S) y \Leftrightarrow x \sigma(\succeq_{xyz}) y \) for any \( z \in S \). Also, by replacing \( \sigma \) with \( \sigma' \) in the argument above, a similar equivalence holds for \( \sigma' \). Then, by the hypothesis, it follows that \( x \sigma(\succeq_S) y \) if \( x \sigma'(\succeq_S) y \) for every \( x, y \in S \). As \( \sigma \) and \( \sigma' \) both admit Axiom 1, this shows that \( \sigma(\succeq_S) \) and \( \sigma'(\succeq_S) \) are identical. \( \square \)

By Claim 1, all we need to show is that the axioms uniquely determine \( \sigma(\succeq_{xyz}) \) for any complete preference relation \( \succeq \) and any \( x, y, z \in X \). As Axiom 1 and Axiom 2 imply the utilitarian welfare criterion, if \( \succeq_{xyz} \) is transitive, then we immediately have \( \sigma(\succeq_{xyz}) = \succeq_{xyz} \). Also, we proved that \( \sigma(\succeq_{xyz}) = \{(x, y), (x, z), (y, z)\} \) when \( \succeq_{xyz} \) is given as \( \succeq_1 \) in Figure 3 (Section 3.3). If \( \succeq_{xyz} = \{(x, y), (y, z), (z, x)\} \cup \Delta_X \), we can readily show that Axiom 1 and Axiom 3 imply \( \sigma(\succeq_{xyz}) = \Delta_X \). The rest of the proof covers a case where \( \succeq_{xyz} = \{(x, y), (y, z), (z, x), (x, z)\} \cup \Delta_X \). This exhausts all cyclic preference relations \( \succeq_{xyz} \) on three alternatives by symmetry. Figure 5 presents all binary relations used below. For the ease of reference, I denote by \( \succeq_{ij} \) a preference relation of the \( i \)th row and the \( j \)th column in the figure. We consider a case where \( \succeq_{xyz} = \succeq_{11} \), and the proof will show that \( \sigma(\succeq_{11}) = \succeq_{14} \).

First, we show that \( \sigma(\succeq_{11}) \) must be either \( \succeq_{14}, \succeq_{21} \) or \( \succeq_{22} \). To see this, note that a preference \( (x, z) \) of \( \succeq_{11} \) is not involved in any cycles of \( \succeq_{11} \), and hence Axiom 2 implies \( x \sigma(\succeq_{11}) z \). Also, observe that

\[
x \sigma(\succeq_{11}) y \quad \text{Inverse} \quad \Leftrightarrow \quad y \sigma(\succeq_{12}) x \quad \text{Neutrality} \quad \Leftrightarrow \quad y \sigma(\succeq_{13}) z
\]
Figure 5: Proof of Theorem 1

by Axiom 3 and Axiom 4. But \( \succeq_{11} = \succeq_{13} \). So, \( \sigma(\succeq_{11}) \) either contains both \((x, y)\) and \((y, z)\) or contains neither of them. These two observations along with Axiom 1 imply that \( \sigma(\succeq_{11}) \) is either of \( \succeq_{14} \), \( \succeq_{21} \), or \( \succeq_{22} \).

Next, we show that \( \sigma(\succeq_{11}) \) cannot be \( \succeq_{21} \). Assume the contrary. Take any \( w \) in \( X \) distinct from \( x, y, z \), and let us consider a preference relation \( \succeq_{23} \). Then, by the hypothesis, the utilitarian welfare criterion, and Axiom 6, we have \( z \sigma(\succeq_{23}) x \) and \( x \sigma(\succeq_{23}) w \). Since \( \sigma(\succeq_{23}) \) is transitive, these also imply \( z \sigma(\succeq_{23}) w \). We then have \( z \sigma(\succeq_{23}) w \) by Axiom 5, where \( \succeq_{23} \) is the restriction of \( \succeq_{23} \) on \( \{y, z, w\} \). This contradicts our previous observation \( \sigma(\succeq_{23}) = \Delta_X \).

We show that \( \sigma(\succeq_{11}) \) is not \( \succeq_{22} \) either. Assume the contrary. Take any \( w \) in \( X \) distinct from \( x, y, z \). Let us first prove that \( \sigma(\succeq_{32}) = \{(z, x), (z, w), (w, x)\} \). For this, consider \( \succeq_{24} \). By the hypothesis, the utilitarian welfare criterion, and Axiom 6, we have \( x \sigma(\succeq_{24}) y \) and \( y \sigma(\succeq_{24}) z \). These then imply that \( x \sigma(\succeq_{24}) w \) and \( w \sigma(\succeq_{24}) z \) by Axiom 3. Restricting \( \succeq_{24} \) on \( \{x, z, w\} \), Axiom 5 therefore implies that \( x \sigma(\succeq_{31}) w \) and \( w \sigma(\succeq_{31}) z \). As \( \succeq_{32} \) is inverse to \( \succeq_{31} \), it follows that \( w \sigma(\succeq_{32}) x \) and \( z \sigma(\succeq_{32}) w \) by Axiom 4. By transitivity of \( \sigma(\succeq_{32}) \) and Axiom 1, this completes to show that
Now, consider \( \preceq_{33} \). Then, we have \( w \sigma(\preceq_{33}) x \) and \( x \sigma(\preceq_{33}) y \) by Axiom 6 and thus \( w \sigma(\preceq_{33}) y \) by transitivity. Axiom 5 hence implies \( w \sigma(\succeq) y \) where \( \succeq \) is the restriction of \( \preceq_{33} \) on \( \{y, z, w\} \). This contradicts our previous observation \( \sigma(\succeq) = \Delta_X \). A conclusion: \( \sigma(\preceq_{11}) = \preceq_{14} \).

We have shown that the axioms uniquely determine \( \sigma(\succeq_{33}) \) for any complete preference relation \( \succeq \) on \( X \) and any \( x, y, z \in X \). By Claim 1, this proves uniqueness of a welfare evaluation rule that satisfies Axiom 1 through Axiom 6. The proof of Theorem 1 is now complete. \( \square \)

**Proof of Proposition 2.** Let \( X \) be a connected metric space and \( \succeq \) be a semiorder on \( X \) with a representation \( (u, \epsilon) \), where \( \sup |u(x) - u(y)| > 2\epsilon \). Take any \( x, y \in X \) with \( u(x) \succeq u(y) \). If \( z \in X \) is such that \( z \succeq x \), then \( u(z) \succeq u(x) - \epsilon \geq u(y) - \epsilon \), and thus \( z \succeq y \). If \( z \in X \) is such that \( y \succeq z \), then \( u(x) \succeq u(y) \succeq u(z) - \epsilon \) and thus \( x \succeq z \). So, \( x \c(\succeq) y \). For the converse, we will prove the contrapositive. Take any \( x, y \in X \) with \( u(y) > u(x) \), and we shall show that \( x \c(\succeq) y \) does not hold. Note that, by the hypothesis, there exists a \( z \in X \) that meets either \( u(z) > u(x) + \epsilon \) or \( u(y) - \epsilon > u(z) \).

Assume the existence of \( z \) with the former inequality. (The proof is similar if the latter holds.) Then, we can let \( u(y) + \epsilon > u(z) > u(x) + \epsilon \) without loss of generality, for \( u(X) \) is an interval by continuity of \( u \). So, \( y \succeq z > x \), negating \( x \c(\succeq) y \). \( \square \)

**Proof of Theorem 3.** Let \( \succeq \) be a relative discounting time preference with an associated representation \( (u, \eta) \). Define \( D := \{\eta(\cdot, r) : r \in [0, \infty)\} \). Then, every \( \delta \in D \) is a continuous function from \( \mathbb{R}_+ \) to \( \mathbb{R}_{++} \). Suppose that \( (x, t), (y, s) \in X \) are dated outcomes such that \( (x, t) \c(\succeq) (y, s) \). Take any \( \delta \in D \), and let \( r \in [0, \infty) \) be such that \( \delta(\cdot) = \eta(\cdot, r) \). As \( u \) is a homeomorphism from \( Z \) to \( \mathbb{R}_{++} \), there exists a \( z \in Z \) such that \( u(z) = \eta(s, r)u(y) \). Then, we have \( (y, s) \sim (z, r) \) by the representation and \( (x, t) \succeq (z, r) \) by the hypothesis. So, \( \delta(t)u(x) = \eta(t, r)u(x) \geq u(z) = \eta(s, r)u(y) = \delta(s)u(y) \), as desired. For the converse, let \( (x, t), (y, s) \in X \) be dated outcomes such that \( \delta(t)u(x) \geq \delta(s)u(y) \) for any \( \delta \in D \). Take any \( (z, r) \in X \) with \( (y, s) \succeq (z, r) \). We have \( \eta(s, r)u(y) \geq u(z) \) by the representation and \( \eta(t, r)u(x) \geq \eta(s, r)u(y) \) by the hypothesis. Then, \( \eta(t, r)u(x) \geq u(z) \) and thus \( (x, t) \succeq (z, r) \). We can similarly show that any \( (z, r) \) with \( (z, r) \succeq (x, t) \) satisfies \( (z, r) \succeq (y, s) \). So, \( (x, t) \c(\succeq) (y, s) \). \( \square \)

**Proof of Proposition 4.** Let \( \succeq_J \) and \( \succeq_K \) be a justifiable preference relation and a Knightian preference relation on \( X \), respectively, with a representing pair \( (u, P) \) of a utility function and a set of priors. Below I will write \( \varphi(f, p) = \sum_{s \in S} p(s)u(f(s)) \) for any \( f \in X \) and \( p \in P \). We wish to show that \( c(\succeq_J) = c(\succeq_K) \). First, take any \( f, g \in X \) with \( f \succeq_K g \), that is, \( \varphi(f, p) \geq \varphi(g, p) \) for any \( p \in P \). If \( h \in X \) is such that \( h \succeq_J f \), then there is a \( p \in P \) with \( \varphi(h, p) \geq \varphi(f, p) \geq \varphi(g, p) \), and thus \( h \succeq_J g \). If \( h \in X \) is
such that \( g \succeq_J h \), then there is a \( p \in P \) with \( \varphi(f, p) \geq \varphi(g, p) \geq \varphi(h, p) \), and hence \( f \succeq_J g \). So, \( f \succeq_J h \). To show the converse, take any \( f, g \in X \) with \( f \succeq_J h \) and suppose that \( f \succeq_K g \) does not hold by contradiction. Then, there is a \( p^* \in P \) such that \( \varphi(g, p^*) > \varphi(f, p^*) \). For any positive real number \( \epsilon > 0 \), let \( f^\epsilon \) be an act that gives a prize larger than that of \( f \) by \( \epsilon \) at any realization of state \( s \in S \) and any resolution of lottery \( f(s) \). By continuity of \( u \), we can pick an \( \epsilon > 0 \) small enough so that \( \varphi(g, p^*) > \varphi(f, p^\epsilon) \). On the other hand, for such an \( \epsilon \), \( \varphi(f^\epsilon, p) > \varphi(f, p) \) for all \( p \in P \) as \( f^\epsilon(s) \) first-order stochastic dominates \( f(s) \) for all states \( s \in S \). So, \( g \succeq_J f^\epsilon \succeq_J f \), contradicting \( f \succeq_J h \). \( \square \)

Proof of Proposition 5. Let \( \succeq \) be a regret preference on \( X \) with an associated representation \( (u, Q) \). Let \( x, y \in X \) be two prospects such that \( x_i \succeq y_i \) for all \( i \). If \( z \in X \) is such that \( y \succeq z \), then by monotonicity of \( u \) and \( Q \),
\[
\sum_{i=1}^n p_i Q(u(x_i)) - u(z_i)) \geq 0
\]
and hence \( z \succeq y \). If \( z \succeq x \), then \( \sum_{i=1}^n p_i Q(u(z_i) - u(x_i)) \geq 0 \) and hence \( z \succeq y \). So, \( x \succeq y \). If, in addition, \( x_i > y_i \) for some \( i \), then \( \sum_{i=1}^n p_i Q(u(x_i) - u(y_i)) > 0 \), for \( u \) and \( Q \) are strictly increasing and \( Q(0) = 0 \). Hence, we have \( x > y \), implying that \( y \succeq z \). \( \square \)

Proof of Proposition 6. Let \( X \) be an arbitrary nonempty set, and \( \succeq \) be a majority preference induced by a set of linear orders \( \succeq_i \) on \( X \) for \( i = 1, \ldots, n \). To verify Pareto criterion, let \( x, y \in X \) be two policies such that \( x > y \) for all \( i \). Observe that, for an arbitrary policy \( z \in X \), \( \{ i : y \succeq_i z \} \subseteq \{ i : x \succeq_i z \} \) and \( \{ i : z \succeq_i y \} \subseteq \{ i : z \succeq_i x \} \) as each \( \succeq_i \) is transitive. So, if \( y \succeq z \), then
\[
\|i : x \succeq_i y\| \geq \|i : y \succeq_i z\| \geq \|i : z \succeq_i x\| \geq \|i : z \succeq_i y\|
\]
and thus \( x \succeq z \). We can similarly show that \( z \succeq x \) implies \( z \succeq y \). So, \( x \succeq y \). Also, the hypothesis implies that \( x > y \), and thus \( y \succeq z \). \( \square \)

Appendix B: Supplementary results

Independence of the axioms We show that the axioms introduced in Section 3.2 are mutually independent. As remarked above, the universally incomparable WER
satisfies all but Axiom 2, and the universally indifferent WER \( \sigma_1 \) satisfies all but Axiom 1. Also, the covering order discussed in Section 6 is a WER that satisfies all but Axiom 4. For Axiom 3, fix any two alternatives \( x^* \) and \( y^* \) in \( X \), and define a WER \( \sigma \) by the following rule. Set \( \sigma(\succsim) = c(\succsim_{x^*y^*}) \) for any complete preference relation \( \succsim \) on \( X \) and \( x, y, z \in X \) unless \( x > y > z > x \) and \( \{x^*, y^*\} \subseteq \{x, y, z\} \), in which case we set \( \sigma(\succsim) = \succsim \cap \{(x^*, y^*), (y^*, x^*)\} \). In turn, for any complete preference relation \( \succsim \) on \( X \) and any \( S \subseteq X \) with \(|S| > 3\), we define \( \sigma(\succsim_S) \) by \( x \sigma(\succsim_S) y \) if and only if \( x \sigma(\succsim_{x^*y^*}) y \) for all \( z \in X \). Then, \( \sigma \) meets all the axioms except Axiom 3.

For Axiom 5, define a WER \( \sigma \) by \( \sigma(\succsim_S) = \{x^*, y^*\} \) for any complete preference relation \( \succsim \) on \( X \) and any \( S \subseteq X \) unless \( \succsim_S = \succsim_{x^*y^*} \) in Figure 5 for some \( x, y, z, w \in X \). In the excluded case, let \( \sigma(\succsim_S) = \{(x, z), (z, x), (y, w), (w, y)\} \). Then, \( \sigma \) satisfies all but Axiom 5. Lastly, for any binary relation \( \succeq \) on \( X \), define a binary relation \( \succeq^\circ \) by \( x \succeq^\circ y \) if \( x \succeq y \) and no cycle of \( \succeq \) involves \( (x, y) \). In turn, define a welfare evaluation rule \( \sigma \) by mapping \( \succsim_S \) to the transitive closure of \( \succsim^\circ_S \) for any complete preference relation \( \succsim \) on \( X \) and any \( S \subseteq X \). Then, \( \sigma \) satisfies all but Axiom 6.

References


