

# The transitive core: inference of welfare from nontransitive preference relations

Hiroki Nishimura\*

January 17, 2014

## Abstract

This paper studies welfare criteria under an environment in which a decision maker is endowed with a nontransitive preference relation. In such an environment, the classical utilitarian welfare criterion may not identify the welfare order, and the problem of maximizing the decision maker's welfare becomes ambiguous. In order to find a sensible welfare criterion that applies to nontransitive preference relations, I propose a series of desirable properties of welfare criteria and uniquely identify a consistent rule that infers welfare orders from nontransitive preference relations. This rule, called the transitive core, is applied to a variety of nontransitive preference models, such as semiorders on a commodity space, relative discounting time preferences, justifiable preferences over ambiguous acts, regret preferences on risky prospects, and collective preference relations induced by the majority criterion. These examinations show that the proposed method provides successful inference of welfare in respective contexts.

---

\*Department of Economics, New York University, 19th West 4th Street, New York, NY 10012 (Email: hiroki@nyu.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, Yoram Halevy, Hao Li, Fabio Maccheroni, Michael Peters, Michael Richter, Gil Riella, Ariel Rubinstein, Yuval Salant, Ennio Stacchetti and seminar participants at Bocconi University, Kellogg School of Management at Northwestern University, New York University, Princeton University, University of California in Riverside, Washington University in St. Louis, and Yokohama National University for helpful comments and discussions. The financial support from C.V. Starr Center at New York University is gratefully acknowledged. The responsibility for any error is my own.

*JEL Classification:* D11, D60.

*Keywords:* Nontransitive Preferences, Behavioral Choice, Welfare.

## 1 Introduction

In the recent literature, much interest is directed to behavioral economics, in which a decision maker either suffers from bounded rationality or follows a heuristic choice procedure so that the standard utility maximization problem may not explain the decision maker's behavior. For example, satisficing, limited cognition, shortlist methods, and framing effects are behavioral choice models that attract significant attention recently.<sup>1</sup> However, when a decision maker follows a behavioral choice procedure, especially when there is no utility function that underlies her behavior, there is ambiguity in how to understand the decision maker's welfare, and this ambiguity poses a serious question in normative analysis such as policy evaluations. This issue leads to the development of *behavioral welfare economics*, where efforts are made to elicit the welfare of the decision maker from observable data. Bernheim and Rangel [6, 7], for instance, target behavioral decision makers affected by framing effects and propose certain methods of welfare inference from observable choice behavior. Following this work, Kőszegi and Rabin [17], Rubinstein and Salant [32], Chambers and Hayashi [11] further investigate methods of welfare inference for behavioral decision makers.

In this paper, I shed light on one of the most classical source of behavioral decision making, namely, cyclic preference relations. To be precise, this paper studies welfare criteria under an environment in which a decision maker is endowed with a nontransitive preference relation. For an illustration of a problem, consider a situation where a third party (such as a policy maker or a researcher) who observes a preference relation of the decision maker tries to make a choice from available options for maximizing the decision maker's welfare. Then, clearly, the third party needs to know, in order to solve this problem, how the decision maker's welfare is measured.

If an observed preference relation is transitive, we may follow the utilitarian welfare criterion and understand that one alternative improves the decision maker's welfare over another if and only if the former is preferred to the latter. Therefore, in this case, the objective of the third party can be simply rephrased by

---

<sup>1</sup>See, for instance, [6], [10], [21], [23], [31], [34], [33], [36] for recent development in these models.

a usual utility maximization problem. If, however, an observed preference relation is cyclic, we may not rely on the above criterion to evaluate the decision maker's welfare, for, otherwise, the welfare could be improved for free by cyclically replacing alternatives given to the decision maker. (This is a dual of what is often called the money-pump argument.) Therefore, the objective function for the third party is ambiguous in this case.

Much evidence suggests that the decision maker's preference relation is often not transitive in practice. Armstrong [2, 3, 4] and Luce [20] argue that, in contrast to the classical theory of values, a decision maker suffers from nontransitive indifferences due to the imperfect ability of discrimination, which leads to the introduction of semiorders and interval orders (Fishburn [12]). Kahneman and Tversky [16] present extensive experimental results showing that subjects violate the expected utility hypothesis in a consistent manner that entails cyclic preferences over prospects. To accommodate such anomalies, Loomes and Sugden [19] develop a nontransitive evaluation of prospects that accounts for the experience of regret. In intertemporal choice contexts, Roelofsma and Read [28], Read [27], Rubinstein [30], Ok and Masatlioglu [26] study alternative discounting models that inherently induce cyclic choice patterns.

This paper proposes, for a complete preference relation of the decision maker, a method of obtaining a part of the preference relation where we can safely say that the decision maker's welfare is revealed even if the original preference relation is cyclic. How is it possible for the third party to infer the welfare values of alternatives when an observed preference relation is cyclic? Suppose, for example, that a decision maker expresses a preference relation in which an alternative  $x$  is ranked higher than another alternative  $y$ . In addition, let us assume that the preference of  $x$  over  $y$  is not involved in *any* preference cycles. While the preference relation might be cyclic for other pairs of alternatives and, thus, we may not identify the observed preference relation with the welfare order as a whole, it seems fairly reasonable to say that the preference of  $x$  over  $y$  is consistent and that this part at least reflects the decision maker's well-being. Such an argument forms a criterion that can be applied even for nontransitive preference relations. This paper investigates the extent to which this approach may be generalized in order to obtain a reliable rule of inferring the welfare order.

**A welfare evaluation rule and the transitive core.** A welfare evaluation rule (WER) is a function that maps a preference relation to a welfare order. Throughout the paper, I assume that a complete (but not necessarily transitive) binary relation

over feasible alternatives is observed as a preference relation and that a reflexive and transitive binary relation is inferred as a welfare order. With this construction, we may formulate criteria like one in the previous paragraph as the properties of WERs that we can use to find a sensible rule from the collection of all welfare evaluation rules. This paper discusses desirable properties imposed on WERs and investigates their implications. As a matter of fact, the properties proposed as the natural requirements of the welfare evaluation rules are shown to identify a unique rule such that

$$x \text{ core}(\succsim)y \quad \text{if and only if} \quad \begin{cases} z \succsim x \text{ implies } z \succsim y \\ y \succsim z \text{ implies } x \succsim z \end{cases} \quad \text{for every } z \in S$$

for an arbitrary complete binary relation  $\succsim$  on a nonempty set  $S$ . I will refer to the order  $\text{core}(\succsim)$  as *the transitive core* of a preference relation  $\succsim$ .

**Model-free approach and cautious inference.** This paper is based on two pillars of ideas: a model-free approach and a cautious inference of welfare. The model-free approach means that an observed preference relation is ex ante not assumed to have any structure other than being a complete binary relation. This feature ex post warrants wide applicability of the transitive core. In Section 4, we examine performance of the transitive core by applying it to a variety of nontransitive preference models, such as those on commodities, dated outcomes, risky prospects, and political candidates. The model-free approach allows us to apply the transitive core to all of these models regardless of the difference in primitives. The cautious inference of welfare, on the other hand, means that the paper attempts to find a part of the observed preference relation where we can safely understand that the decision maker's welfare is reflected. In other words, if a welfare order between two alternatives is ambiguous, we allow ourselves to reserve the comparison by saying that their welfare values are *incomparable* on a basis of observed information. Therefore, a welfare evaluation rule maps a preference relation to a possibly incomplete welfare order on the set of feasible alternatives. In this regard, this paper adopts an approach similar to that of Bernheim and Rangel [6, 7] and contrasts with that of Rubinstein and Salant [32]. (See Section 5 for further discussion on the relation between the transitive core and the work of Bernheim and Rangel. In particular, it is shown that, when nontransitivity of a preference relation is indeed a source of behavioral decision making, the transitive core offers a shaper criterion than the method studied by Bernheim and Rangel. While both Bernheim and Rangel and this paper take cautious ap-

proaches, the transitive core achieves stronger implications by restricting its focus to nontransitive preferences.)

**More on related literature.** Aside from the literature referred above, there are some works that are well worth mentioned in connection with this paper. For example, in the social choice theory, Fishburn [13] and Miller [25] introduce *the covering order* for a tournament (i.e. a complete asymmetric binary relation), which is motivated as an inference method of the social welfare. Alternatively, for an individual with imperfect perception, Luce [20] propose his method of inferring a “true” utility measure from observation of a distorted preference relation called a semiorder.<sup>2</sup> It turns out that, in the respective contexts, these methods obtain the same welfare order as one found by the transitive core: the covering order of a tournament is the transitive core of the same tournament, and Luce’s order for a semiorder (introduced in Section 4) is the transitive core of the same relation. However, what is more notable is that these methods do *not* obtain the same result as the transitive core when applied in the different contexts: the covering order of a semiorder, for example, does not reveal the true utility measure, and the Luce’s order applied to a tournament is not even transitive in general (and, of course, they exhibits quite different features from the transitive core when applied to nontransitive time preferences or preferences over risky outcomes.) Therefore, the transitive core may be viewed as a genuine generalization of welfare inference methods that reduces to the known rules in appropriate contexts while it also applies to realms that has not been studied yet such as intertemporal choices or choices under risk and uncertainty. Further discussion on the related literature is given in Section 6.

The paper is structured as follows. The next section introduces notation and terminology used throughout the paper. In Section 3, we formally define welfare evaluation rules and investigate their desirable properties. The transitive core is then obtained as a unique welfare evaluation rule that satisfies all the proposed properties, and its computational characterization is provided. Section 4 gives case studies of the transitive core on semiorders, time preferences, preferences under risk, and collective preferences. Section 5 proceeds by making a direct

---

<sup>2</sup>To be precise, Luce defines that, for a complete binary relation  $\succsim$ ,  $x$  *dominates*  $y$  if either (a)  $x \succ y$ , (b)  $x \sim y$  but  $x \sim z \succ y$  for some  $z$ , or (c)  $x \sim y$  but  $x \succ z \sim y$  for some  $z$ , and say that  $x$  and  $y$  are *indifferent* if neither dominates the other. While the definition yields an order identical to the transitive core when applied to semiorders introduced by him (and in Section 4), it is also easy to see that this is not the case in general. (Consider, for example, a preference relation  $\succsim$  such that  $x \succ y \succ z \succ x$ .)

comparison between the transitive core and the unambiguous welfare improvement order proposed by Bernheim and Rangel [6, 7]. The related literature is further discussed in Section 6. Section 7 closes the paper with a few remarks. All the proofs and supplementary results are given in the appendix.

## 2 Preliminaries

**Preference relations.** For an arbitrary set  $S$ , a *binary relation* on  $S$  is a subset of  $S \times S$  with a generic notation  $\succcurlyeq$ . As usual, I write  $x \succcurlyeq y$  to mean that  $(x, y) \in \succcurlyeq$ . Also, the notations  $\succ$  and  $\sim$  are used to denote the strict part and the symmetric part of  $\succcurlyeq$ , respectively. A binary relation  $\succcurlyeq$  on  $S$  is *reflexive* if  $x \succcurlyeq x$  for all  $x \in S$ , *complete* if either  $x \succcurlyeq y$  or  $y \succcurlyeq x$  for any  $x, y \in S$ , *transitive* if  $x \succcurlyeq y$  and  $y \succcurlyeq z$  imply  $x \succcurlyeq z$  for any  $x, y, z \in S$ , and *antisymmetric* if  $x \succcurlyeq y$  and  $y \succcurlyeq x$  imply  $x = y$  for any  $x, y \in S$ . A reflexive and transitive binary relation on  $S$  is called a *preorder* on  $S$ . In turn, a preorder on  $S$  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  $S$  is the trivial partial order on  $S$  defined by  $\Delta_S = \{(x, x) : x \in S\}$ . A *preference relation* is simply a synonym for a binary relation, except that it is used, in particular, to represent the decision maker's taste so that the condition  $x \succcurlyeq y$  is read as “ $x$  is preferred at least as much as  $y$ .” An ordered pair  $(x, y)$  such that  $x \succcurlyeq y$  is called a *preference* of  $\succcurlyeq$ . (Hence, a preference relation is a set of preferences.) 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 Figure 1. In the figure, each vertex represents an alternative, and a directed arrow is drawn from a strictly preferred alternative to a strictly less preferred alternative. An undirected line between a pair of alternatives is used to represent the indifference of the corresponding pair in the preference relation.

**Preference cycles.** For an arbitrary preference relation  $\succcurlyeq$  on a set  $S$ , a *cycle* of  $\succcurlyeq$  is a finite sequence  $(z_l)_{l=1}^k$  of distinct points in  $S$  such that  $z_1 \succcurlyeq z_2 \succcurlyeq \cdots \succcurlyeq z_k \succcurlyeq z_1$  with at least one preference holding strictly. We say that a preference  $(x, y)$  of  $\succcurlyeq$  is *involved* in a cycle  $(z_l)_{l=1}^k$  of  $\succcurlyeq$  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  $\succcurlyeq$  is said to be *cyclic* if there is at least one cycle of  $\succcurlyeq$  and to be *acyclic* otherwise. Two cycles of a preference relation that are identical

upon rotation are identified. For example, a preference relation  $\succsim_1$ , depicted by Figure 2, has three sequences of alternatives,  $(x, y, z)$ ,  $(y, z, x)$ , and  $(z, x, y)$ , which are cycles of  $\succsim_1$ , but they are viewed as the same cycle. In this sense, we say that the preference relation  $\succsim_1$  has a unique cycle.

**Permutation, inverse, restriction.** Let  $\succsim$  be an arbitrary preference relation on a set  $S$ . Given a permutation  $\pi$  on  $S$ , we define a binary relation obtained from  $\succsim$  by relabeling alternatives according to  $\pi$  and write this relation by  $\pi(\succsim)$  with an abuse of notation. Formally, it is defined as  $\pi(\succsim) = \{(\pi(x), \pi(y)) : x \succsim y\}$ . An *inverse* of a preference relation  $\succsim$  is a binary relation  $\text{inv}(\succsim)$  such that  $x \text{inv}(\succsim) y$  if and only if  $y \succsim x$  for all  $x, y \in S$ . Given a subset  $T$  of  $S$ , the *restriction* of  $\succsim$  on  $T$  is a binary relation  $\succsim_T$  on  $T$  such that  $x \succsim_T y$  if and only if  $x \succsim y$  for all  $x, y \in T$ . When  $T$  is a finite set, say  $T = \{x, y, z\}$ , we will write  $\succsim_{xyz}$  instead of  $\succsim_{\{x, y, z\}}$  for brevity. For any  $x \in S$ , the upper and lower contour sets of the preference relation  $\succsim$  at the point  $x$  are denoted by  $U(x, \succsim) = \{z : z \succsim x\}$  and  $L(x, \succsim) = \{z : x \succsim z\}$ , respectively.

### 3 The transitive core

This paper studies a method of inferring a welfare order over alternatives from observation of a decision maker's preference relation. If the observed preference relation is transitive, the utilitarian welfare criterion infers that replacing a less preferred object with a preferred object improves the decision maker's welfare. However, this criterion, as a rule of identifying the welfare order with the observed preference relation, gives a cyclic welfare order whenever the observed preference relation is cyclic. Suppose, for example, 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 were replaced by  $y$  and then by  $x$ . However, since the decision maker prefers  $z$  to  $x$  at the same time, she would end up with the initial alternative  $z$  with her welfare being strictly improved, which is a quite unrealistic conclusion. This argument suggests that, while a preference relation can be cyclic as a notion that describes a decision maker's behavior, the welfare order should be transitive as a measure that evaluates values of alternatives.

A preference relation is, nevertheless, still informative in eliciting the welfare of a decision maker. For instance, suppose that we observe a preference relation  $\succsim_1$  given in Figure 1 from a decision maker. Note that, whereas the preference

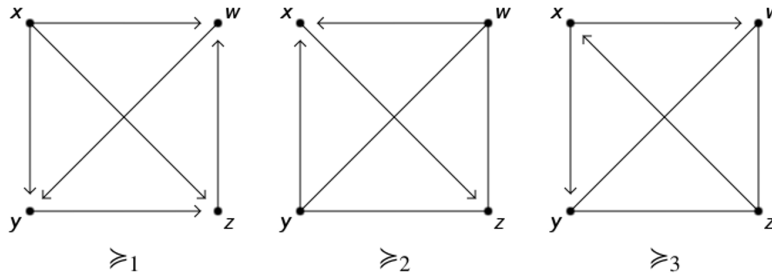


Figure 1: Nontransitive preference relations on four alternatives

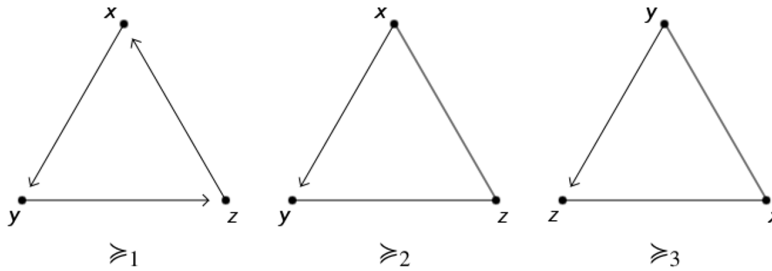


Figure 2: Nontransitive preference relations on three alternatives

relation is cyclic as a whole, the preferences of an alternative  $x$  over the other alternatives are fairly consistent in the sense that they are not involved in *any* preference cycles of  $\succsim_1$ .<sup>3</sup> Having this convincing evidence of the decision maker's preference, therefore, it seems reasonable to conclude that this part of a preference relation at least reflects the welfare values of alternatives. Importantly, the criterion for the welfare inference is applied regardless of the fact that the preference relation is cyclic. Therefore, although observed preferences must be carefully inspected, eliciting welfare values of alternatives from a nontransitive preference relation is, at least in part, still possible.

In addition, a method of inferring welfare orders should be coherent across various preference relations. For example, Figure 2 depicts two cyclic preference relations,  $\succsim_2$  and  $\succsim_3$ , that share the identical structure except for specific labels of alternatives. Having this association, it seems plausible to require that the inferred

<sup>3</sup>In fact,  $(y, z, w)$  is a unique cycle of  $\succsim_1$ .



welfare order between two alternatives (say,  $x$  and  $y$ ) for  $\succsim_2$  coincide with that of the corresponding two ( $y$  and  $z$ , respectively) for  $\succsim_3$ . Such a requirement provides coherence to a rule of welfare inference.

**Cautious inference.** As noted in the introduction, this paper does not necessarily attempt to infer a *complete* welfare order from every preference relation of the decision maker. Instead, it aims to clarify when it is safe to say that one alternative improves the decision maker's welfare over another. A preference relation  $\succsim_1$  shown in Figure 2 gives here a good example that illustrates this approach. First, this preference relation is cyclic, and we may not understand that every preference of  $\succsim_1$  reflects the welfare values of alternatives. However, inferring a welfare order from this preference relation is further complicated by the fact that it has a precisely symmetric structure in preferences. In particular, preserving one preference over another as a part of the welfare order necessarily goes beyond a criterion based on the observed preference relation. The cautious solution in this case, especially when we do not *ex ante* assume any particular model that induces  $\succsim_1$ , is to reserve every welfare comparison between alternatives.

It should be remarked that inferring indifference as a welfare order is different from inferring incomparability. In general, if we assume a particular model that induces observation of a cyclic preference relation, it is often possible to propose a refined method of welfare inference by relying on the additional structures. As the safe inference under a model-free approach, therefore, the method studied in this paper is expected to give a part of the welfare order that is agreed by the refined methods of welfare inference in a variety of contexts. Inferring incomparability leaves room for further criteria based on a specific model of nontransitive preference relations, whereas inference of indifference is a certain statement that two alternatives have the same values in welfare.

The present section proceeds by formulating a method of inferring welfare orders as a certain map called a *welfare evaluation rule* (WER), and we seek desirable properties of WERs, including those argued above. The main finding of the section shows that a set of axioms proposed as the natural requirements on the WER characterizes a unique welfare evaluation rule, which we call the *transitive core*.

### 3.1 Welfare evaluation rules

Let  $X$  be the set of conceivably all alternatives of interest with  $|X| > 3$ . I consider an environment in which nature or an experiment may restrict feasibility of alternatives to a subset of  $X$ , and once a feasibility constraint is set, a preference relation is observed on this restricted domain. I assume that a complete (but not necessarily transitive) binary relation is observed as a preference relation of the decision maker when a set of feasible alternatives is resolved. A method of welfare inference is formulated as a function that maps a preference relation to a welfare order, prepared by the third party *before* the resolution of a feasible constraint and an observation of a preference relation. Let  $\mathcal{P}$  be the class of all preference relations that are potentially observed prior to the resolution of a feasibility constraint:

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

For any  $\succsim \in \mathcal{P}$ , write  $D_{\succsim} = \{x \in X : x \succsim x\}$  for the domain of the preference relation  $\succsim$ . A *welfare evaluation rule* (WER) is a map  $\sigma$  on  $\mathcal{P}$  such that  $\sigma(\succsim)$  is a preorder on  $D_{\succsim}$  for each  $\succsim \in \mathcal{P}$ . A WER is interpreted as a rule employed by the third party to infer the welfare order from an observed preference relation of the decision maker. Transitivity and reflexivity are imposed as minimal requirements on the inferred welfare orders. Of course, these requirements are less than adequate for a welfare evaluation rule to be understood as a sensible method of welfare inference. The following examples present two rules that are either of little use or hardly acceptable.

**Example** (Universal incomparability). The universally incomparable WER is a map  $\sigma_0$  that assigns the diagonal order  $\Delta_{D_{\succsim}}$  for every preference relation  $\succsim$  in  $\mathcal{P}$ . This rule infers that every alternative has the same value as itself, but it makes no other welfare judgment. So, the universally incomparable WER is correct wherever it gives a valid comparison but, most likely, useless.

**Example** (Universal indifference). The universally indifferent WER is a map  $\sigma_1$  that assigns the trivial weak order  $D_{\succsim} \times D_{\succsim}$  for every preference relation  $\succsim$  in  $\mathcal{P}$ . This rule alleges that any two alternatives have the same value regardless of a preference relation expressed by a decision maker. The universally indifferent WER is likely unacceptable for the intended purpose.

As these examples imply, the class of all the welfare evaluation rules is quite large. In order to refine the class to one that provides insight, below I examine

desirable properties of welfare evaluation rules. Note that the classical welfare criterion can also be viewed as a property of the WER: We say that a welfare evaluation rule  $\sigma$  confirms *the utilitarian welfare criterion* if  $\sigma(\succsim) = \succsim$  for every complete and transitive preference relation  $\succsim \in \mathcal{P}$ .

**Remark.** Note that the main problem of this paper, i.e. understanding the welfare for a behavioral decision maker, is a normative exercise. A welfare evaluation rule represents a method of welfare inference adopted by the third party such as a policy maker, not by a decision maker, and desirability of properties imposed on WERs should be evaluated from the normative perspective. Accordingly, observability of preference relations in  $\mathcal{P}$  is not necessary in evaluating these properties.<sup>4</sup> (This is a similar point in evaluating the axiomatization of Nash bargaining solution or the Pareto criterion in the social choice theory.)

### 3.2 Desirable properties of welfare evaluation rules

The utilitarian welfare criterion is, however, silent when an observed preference relation is cyclic. Therefore, any nontrivial choice of WER must be made by the criteria beyond this property. In this section, I propose six axioms on welfare evaluation rules and discuss their justification from the normative perspective. The following axioms apply to any subsets  $S, T$  of  $X$  with  $T \subseteq S$ , any complete preference relation  $\succsim$  on  $S$ , and any permutation  $\pi$  on  $S$ .<sup>5</sup>

**Axiom 1** (prudence).  $x \sigma(\succsim) y$  only if  $x \succsim y$ .

**Axiom 2** (consistency). If  $(x, y)$  is involved in no cycle of  $\succsim$  and  $x \succsim y$ ,  $x \sigma(\succsim) y$ .

**Axiom 3** (anonymity).  $\sigma \circ \pi(\succsim) = \pi \circ \sigma(\succsim)$ .

**Axiom 4** (symmetry).  $\sigma \circ \text{inv}(\succsim) = \text{inv} \circ \sigma(\succsim)$ .

**Axiom 5** (reduction). If  $x, y \in T$  and  $x \sigma(\succsim) y$ , then  $x \sigma(\succsim_T) y$ .

**Axiom 6** (extension). If  $x \sigma(\succsim_{xyz}) y$  for all  $z \in S$ , then  $x \sigma(\succsim) y$ .

---

<sup>4</sup>Of course, in order for the third party to *use* this method of welfare inference, it must observe a preference relation from an individual decision maker whose welfare the third party is interested in. An only point stressed here is that the model by no means assumes observability of many preference relations in  $\mathcal{P}$  from a single decision maker.

<sup>5</sup>It is shown in the appendix that the proposed axioms are mutually independent.

The first axiom, *prudence*, requires a welfare evaluation rule to elicit the precedence of one alternative over another in welfare values only when the decision maker at least expresses a preference of the former over the latter. Put differently, it asks the WER not to add a preference to the welfare order that the decision maker did not even manifest. Notice that, under the prudence axiom, a WER may infer the strict welfare order for a pair of alternatives that are evaluated as indifferent by the original preference relation.<sup>6</sup> However, when  $x$  is strictly preferred to  $y$ , the property prevents the welfare evaluation rule from inferring the strict precedence of  $y$  over  $x$  in welfare values.

A welfare evaluation rule is said to be *consistent* if it rejects a preference between alternatives as the welfare order only on the basis of preference cycles. For example, consider a preference relation  $\succsim_1$ , depicted in Figure 1. This preference relation, although cyclic as a whole, exhibits strong consistency in preferences of an alternative  $x$ . In fact, a unique cycle of  $\succsim_1$ ,  $(y, z, w)$ , does not negate the preferences of  $x$  to other alternatives, in the sense that the cycle does not involve the preferences of  $x$ . This consistency makes it fairly persuasive for us to believe that the choice of  $x$  brings the decision maker higher welfare than any other alternatives do. The consistency axiom is a requirement on the WER to preserve a preference as a part of the welfare order the credibility of which is not impaired by any preference cycles.

Note that, since a transitive preference relation admits no preference cycle within, a consistent welfare evaluation rule must preserve all the preferences expressed by the relation. With the prudence axiom, which guarantees no addition of unobserved preferences to the welfare order, the welfare evaluation rule maps every transitive preference relation to itself. Therefore, it follows that the prudence and consistency axioms imply the utilitarian welfare criterion.

**Fact 1.** A prudent and consistent WER confirms the utilitarian welfare criterion.<sup>7</sup>

The anonymity axiom can be equivalently rephrased as follows: If there is a way,  $\pi$ , to relabel alternatives so that two preference relations,  $\succsim$  and  $\succsim'$ , are related in such a manner that  $x \succsim y$  iff  $\pi(x) \succsim' \pi(y)$  for any  $x$  and  $y$ , then the

<sup>6</sup>An example where this might be the case of interest is the preference relation  $\succsim_2$  in Figure 2. In this preference relation, although the decision maker shows an indifference between  $x$  and  $z$ , there is an indication that implies the superiority of  $x$  over  $z$ . Indeed, she evaluates the value of  $z$  as equal to that of  $y$ , while giving  $x$  a higher value than  $y$ . In order to resolve a preference cycle shown by the decision maker, therefore, it may be reasonable to do so by breaking the indifference of  $x$  and  $z$  in favor of  $x$ .

<sup>7</sup>Recall that the utilitarian welfare criterion means that  $\sigma(\succsim) = \succsim$  for every transitive  $\succsim \in \mathcal{P}$ .

welfare orders inferred from these preference relations should relate in the same manner; that is,  $x \sigma(\succcurlyeq) y$  iff  $\pi(x) \sigma(\succcurlyeq') \pi(y)$  for every  $x$  and  $y$ . To illustrate, a pair of preference relations  $\succcurlyeq_2$  and  $\succcurlyeq_3$ , depicted in Figure 2, is an example of this case. (Observe that two preference relations share the same structure upon relabeling of alternatives.) The axiom then requires that a welfare evaluation rule accept, say, the preference of  $x$  over  $y$  as a part of the welfare order for  $\succcurlyeq_2$  if and only if it accepts the preference of  $y$  over  $z$  for  $\succcurlyeq_3$ .

Analogously, the symmetry axiom is rephrased by stating that, if  $\succcurlyeq$  and  $\succcurlyeq'$  are the inverse of each other—that is,  $x \succcurlyeq y$  iff  $y \succcurlyeq' x$  for any  $x$  and  $y$ —then the inferred welfare orders for these preference relations are related in the same way. Two preference relations  $\succcurlyeq_2$  and  $\succcurlyeq_3$ , given in Figure 1, are an example of this case. (Notice that every directed arrow in  $\succcurlyeq_2$  is flipped by  $\succcurlyeq_3$ .) The symmetry axiom then requires the WER to preserve, say, the preference of  $x$  over  $z$  as a part of the welfare order for  $\succcurlyeq_2$  if and only if it does the same to the preference of  $z$  over  $x$  for  $\succcurlyeq_3$ .

The last two axioms, *reduction* and *extension*, are coherence requirements on the welfare evaluation rules for the restrictions of a preference relation. (Recall that the notations  $\succcurlyeq_T$  and  $\succcurlyeq_{xyz}$  represent the restrictions of a preference relation  $\succcurlyeq$  on the smaller domains.) To interpret the reduction axiom, notice that every cycle of the restricted preference relation  $\succcurlyeq_T$  is a cycle of the original preference relation  $\succcurlyeq$ , which means that  $\succcurlyeq$  always has at least as many (and possibly more) preference cycles as  $\succcurlyeq_T$ . The axiom states that, if a preference is acceptable to form a part of the welfare order with respect to the preference relation  $\succcurlyeq$ , then the same should be the case for the preference relation  $\succcurlyeq_T$ .

Lastly, the extension axiom claims that, if a preference is credible as a part of the welfare order for every triangle restriction of the original preference relation, then so is the case for the original preference relation. With regard to this axiom, one might wonder why the axiom requires only the triangle restrictions but not the square or larger restrictions for evidence of a preference  $(x, y)$  being supportable. One fact is that, for an arbitrary complete preference relation, whenever there is a preference cycle of length four or more, there exists a cycle of length three within it. Put differently, if a preference relation is free from triangle cycles, it is *necessarily* transitive. In this sense, triangle cycles are fundamental cycles that make a preference relation cyclic. The axiom claims that, if a preference of one alternative over another is credible as a part of the welfare order relative to every fundamental cycle of a preference relation, then it is safe to say that this preference is credible for the original preference relation.

Note that, while a number of axioms are proposed in this section, each axiom

is, to some extent, a plausible assumption on the welfare evaluation rules. In fact, there are some indications suggesting that the proposed axioms might be too weak to find an insightful rule from the class of all welfare evaluation rules. The following facts show that two extreme welfare evaluation rules, previously introduced as examples of little use or unjustifiable method, admit all the axioms except for one axiom each.

**Fact 2.** The universally incomparable WER  $\sigma_0$  satisfies Axioms 1, 3, 4, 5, and 6.

**Fact 3.** The universally indifferent WER  $\sigma_1$  satisfies Axioms 2, 3, 4, 5, and 6.

### 3.3 The transitive core

Contrary to the weakness of the axioms implied by the two facts above, it turns out that the proposed axioms are just enough to identify a *unique* welfare evaluation rule.

**Theorem 1.** *A welfare evaluation rule that admits Axioms 1-6 is unique.*

So, while many welfare evaluation rules, including those that are unjustifiable, confirm each of the proposed axioms, there exists at most one rule that admits all axioms at the same time. Since a welfare evaluation rule that satisfies all the axioms is unique, we may give a name to refer this particular rule for convenience. This name given to the unique rule is *the transitive core*. Therefore, the transitive core is innately justified through a set of desirable properties of welfare evaluation rules.

**Definition.** The transitive core is a welfare evaluation rule that satisfies Axioms 1-6.

For the transitive core, I reserve the notation  $\text{core}(\cdot)$  to distinguish it from a generic welfare evaluation rule. Accordingly, for every  $\succsim \in \mathcal{P}$ , the transitive core of a preference relation is  $\text{core}(\succsim)$  so that  $x \text{core}(\succsim) y$  means that “an alternative  $x$  is evaluated by the transitive core to have a higher value in welfare than another alternative  $y$ .” The strict part of the transitive core is denoted by  $\text{core}^*(\succsim)$ .

**Remark.** Though the formal proof of Theorem 1 is given in the appendix, it is straightforward to see how the axioms determine the welfare order for a preference

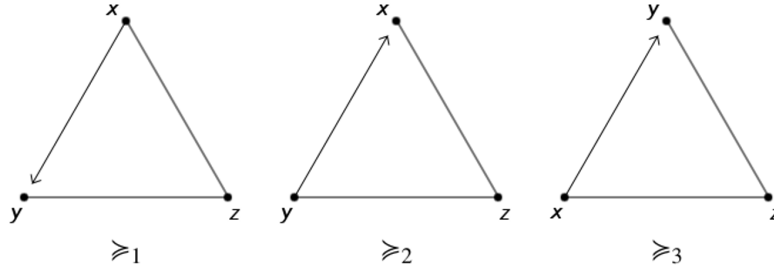


Figure 3: Illustration of Theorem 1

relation  $\succcurlyeq_1$  in Figure 3.<sup>8</sup> Notice that preferences  $(x, z)$  and  $(z, y)$  are not involved in any cycle of  $\succcurlyeq_1$ , making them credible as a part of the welfare order  $\sigma(\succcurlyeq_1)$  by the consistency axiom. It is then implied that  $x \sigma(\succcurlyeq_1) y$  as the welfare order is transitive. As the last step, observe the following implications of the anonymity and symmetry axioms on three preference relations depicted in Figure 3:

$$z \sigma(\succcurlyeq_1) x \quad \overset{\text{symmetry}}{\iff} \quad x \sigma(\succcurlyeq_2) z \quad \overset{\text{anonymity}}{\iff} \quad y \sigma(\succcurlyeq_3) z.$$

(Note that  $\succcurlyeq_1$  and  $\succcurlyeq_2$  are the inverse of each other and that  $\succcurlyeq_3$  is obtained from  $\succcurlyeq_2$  by swapping the labels  $x$  and  $y$ .) But  $\succcurlyeq_1$  and  $\succcurlyeq_3$  are *identical*. Therefore,  $\sigma(\succcurlyeq_1)$  must either preserve both  $(z, x)$  and  $(y, z)$  or reject both  $(z, x)$  and  $(y, z)$ . The welfare order  $\sigma(\succcurlyeq_1)$  cannot preserve both of these preference relations, for otherwise  $\sigma(\succcurlyeq_1)$  would coincide with  $\succcurlyeq_1$ , which is not transitive.<sup>9</sup> The observation concludes the identification of  $\sigma(\succcurlyeq_1)$  as one that evaluates  $x$  strictly higher than  $y$ , and  $y$  strictly higher than  $z$ .

Two issues immediately arise with respect to the definition. The first is the existence of the rule. As Theorem 1 only guarantees the uniqueness of a welfare evaluation rule satisfying the axioms, it is not clear if the definition is made on an existing object. The second is the accessibility of the rule: It is not clear how we can compute the transitive core for a given preference relation because the rule

<sup>8</sup>In sharp contrast, proving that the axioms uniquely identify the welfare order for a preference relation  $\succcurlyeq$  on  $\{x, y, z\}$  such that  $x > y > z \sim x$  is much less straightforward. It should be also remarked that Theorem 1 fails if the assumption  $|X| > 3$  made in Section 3.1 is weakened. See Appendix 7 for details.

<sup>9</sup>Also, note that unobserved preferences in  $\succcurlyeq_1$  may not be added to  $\sigma(\succcurlyeq_1)$  by the prudence axiom.

is defined by properties instead of a particular operation. The concept will be of little use in applications if the process of computation is too hard.

As a matter of fact, these problems can be dealt with at a time. The next proposition introduces a simple criterion to evaluate, for any alternatives  $x$  and  $y$ , whether the value of  $x$  in welfare is higher than that of  $y$ . On the one hand, this criterion defines a welfare evaluation rule for which we can readily verify that all the axioms are met. On the other hand, the criterion, thus by definition, obtains the transitive core, providing a method to compute the rule.

**Proposition 2.** *The transitive core exists. Moreover, it is obtained by*

$$x \text{ core}(\succsim) y \quad \text{if and only if} \quad \begin{cases} z \succsim x \text{ implies } z \succsim y \\ y \succsim z \text{ implies } x \succsim z \end{cases} \text{ for every } z \in X.$$

for any preference relation  $\succsim \in \mathcal{P}$  and any  $x, y \in D_{\succsim}$ .

In words, a welfare value of  $x$  is evaluated higher by the transitive core than that of  $y$  if and only if any third alternative  $z$  preferred over  $x$  is preferred over  $y$ , and any alternative  $z$  less preferred to  $y$  is less preferred to  $x$ . It is straightforward to see that the transitive core may be characterized by the upper and lower contour sets of preference relations. The next fact holds for any preference relation  $\succsim$  in  $\mathcal{P}$  and any  $x, y \in D_{\succsim}$ .

**Fact 4.**  $x \text{ core}(\succsim) y$  if and only if  $U(x, \succsim) \subseteq U(y, \succsim)$  and  $L(y, \succsim) \subseteq L(x, \succsim)$ .

Notably, the welfare criterion obtained by Proposition 2 allows applications of the transitive core to various models of nontransitive preference relations. The next section studies the performance of the transitive core applied to semiorders, relative discounting time preferences, justifiable preferences, regret preferences, and social preferences. Although the axiomatic foundation of the transitive core does not (at least directly) guarantee its performance in the applications, we will indeed verify that the rule successfully infers an intuitive welfare order in respective contexts.

**Remark** (The transitive core is more than preferences free from cycles). While the consistency axiom is fairly reasonable, it is only a partial criterion that forms the transitive core, and preferences can be viewed as a part of welfare order even if they are involved in preference cycles. Indeed, the implication of the axiom is often not strong enough to make any welfare inference in a situation where our intuition may suggest certain judgment. Let, for example,  $X = \{1, 2, 3, \dots, n\}$



for an arbitrary  $n \geq 4$ , and let  $\succsim$  be a preference relation on  $X$  such that  $x \succ y$  iff  $x > y$  whenever  $(x, y) \neq (n, 1)$ , and  $1 \succ n$ . Note that the preference relation is transitive when restricted on a subset of  $X$  that misses either 1 or  $n$ , just as the descending order on natural numbers. It lacks full transitivity only due to a “small” mistake  $n \succ 1$ , and the statement that the preference of  $n$  over 1 is more ambiguous to form a welfare order compared to other less problematic parts (such as the preference of 3 over 2) seems intuitive, if not formal. However, the implication of the consistency axiom here is empty, for every component of the preference relation is involved in at least one preference cycle. On the other hand, the transitive core does infer a nonempty welfare order,  $3 \text{ core}(\succsim) 2$  being supported as a part of the welfare order, for example.<sup>10</sup> While the transitive core is developed with aim of making agreeable welfare judgment in many contexts, it is not the *safest* method of welfare inference. The rule makes one step ahead towards finer welfare inference by means of the axiomatic criteria proposed in the last section.

## 4 Applications

### 4.1 Semiorders: imperfect ability of discrimination

Nontransitive indifference due to imperfect ability of discrimination has long been confirmed in the literature. Armstrong [2, 3, 4] poses a question on the assumption of transitive indifference and first introduces a utility model of imperfect discrimination. Luce [20] brings a notion of semiorders into economics and provides its axiomatic foundation.<sup>11</sup> Subsequently, many generalizations of semiorders, such as interval orders by Fishburn [12], are developed in search of further descriptive models of nontransitive indifference. In this section, let  $X$  be a connected metric

---

<sup>10</sup>Although the example here might be deemed technical, it can be shown that the same (i.e. the empty welfare inference by the sole application of the consistency axiom) is the case for any nontransitive semiorders, relative discounting time preferences, and regret preferences studied in the next section. In contrast, the transitive core provides a nonempty and acceptable criterion for evaluating the decision maker’s welfare even for these models.

<sup>11</sup>A well-known example of a preference relation under imperfect discrimination is of “coffee and sugar,” first introduced by Luce. He argues that, even when a subject has a strict preference for, say, a cup of coffee with one cube of sugar to another with five cubes, we can easily expect that a subject will exhibit indifference between any two cups of coffee with just a grain difference of sugar, because she is simply not able to distinguish the difference. Therefore, for a finite sequence of cups of coffee, the decision maker may exhibit indifference to every adjacent pair but strict preference between the first and the last cups.

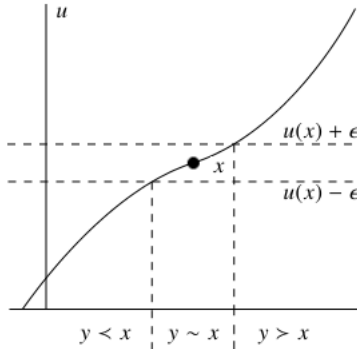


Figure 4: A semiorder representation

space, and let us consider the following representation of semiorders originally studied by Luce.

**Definition.** A semiorder is a binary relation  $\succsim$  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

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

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 exhibits cyclic preferences of alternatives. In particular, the representation can be equivalently rephrased by two conditions: (a)  $x \sim y$  iff  $|u(x) - u(y)| \leq \epsilon$  and (b)  $x \succ y$  iff  $u(x) > u(y) + \epsilon$ . Therefore, a semiorder shows an indifference between alternatives  $x$  and  $y$  even when their utility values are different (provided that the difference is smaller than the threshold  $\epsilon$ ; Luce called  $\epsilon$  *the just noticeable difference* for this reason). Figure 4 illustrates the regions of preferred, indifferent, and less preferred options relative to a given  $x \in X$  when  $X$  is a real line.

Importantly, while a semiorder is nontransitive in general and thus does not fully represent the welfare values of alternatives, the “right” welfare order for the decision maker is indeed obvious in this context. Note that nontransitivity of a semiorder is induced not from cyclical utility values of alternatives, but from imperfect perception of the decision maker toward the values. Therefore, if it were possible to reduce the perception error  $\epsilon$  to zero, the decision maker could correctly evaluate alternatives according to the utility function  $u$ . The transitive

core allows a third party to do this for the decision maker: For any observed preference relation that is a semiorder, taking the transitive core infers the order induced by a utility function  $u$ .

**Proposition 3.** *Let  $\succsim$  be a semiorder on  $X$  with a representation  $(u, \epsilon)$  that satisfies  $\sup |u(x) - u(y)| > 2\epsilon$ .<sup>12</sup> Then,*

$$x \text{ core}(\succsim) y \quad \text{if and only if} \quad u(x) \geq u(y)$$

for any  $x, y \in X$ .

It should be remarked that, while the right welfare order may be given by a utility function  $u$ , this function is *not* directly observable. An observed preference relation is a semiorder that is distorted due to the imperfect ability of discrimination. Proposition 3, therefore, implies that, although the welfare order is unobservable at first, the third party may proceed with the welfare analysis (such as a choice of the best alternative for the decision maker) only on the basis of an observed preference relation, for the transitive core infers the utility function  $u$  from the preference relation.

## 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 that the decision maker receives a prize of  $x$  dollars at period  $t$ . Correspondingly, a complete preference relation on  $X$  is called a *time preference*. There are many interesting models of time preferences that are special cases of an *absolute discounting time preference* represented as

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

for every  $(x, t), (y, s)$  in  $X$  under some functions  $\delta$  and  $u$  (where the former is often called a discounting function and the latter is a utility function). *The exponential*

---

<sup>12</sup>The condition on the supremum distance of  $u$  is a necessary and sufficient condition for the uniqueness of an order on  $X$  induced by a utility function  $u$  that represents a semiorder. For instance, consider an extreme opposite case where a semiorder admits a representation  $(u, \epsilon)$  that satisfies  $\sup |u(x) - u(y)| < \epsilon$ . Then, the semiorder is indifferent for every pair of alternatives, and any function  $u' : X \rightarrow \mathbb{R}$  would instead represent the same semiorder as long as  $\sup |u'(x) - u'(y)| < \epsilon$  holds. In contrast, the ordinal ranking of alternatives induced by the utility function is unique if, and only if, the condition above is satisfied.

*discounting* and *the hyperbolic discounting* time preferences are examples of this class. Notice that any absolute discounting time preference, including its special case, is transitive for having a utility representation  $(x, t) \mapsto \delta(t)u(x)$ .

In contrast, Read [27] finds that a significant fraction of subjects in an experiment exhibits *subadditive discounting* that is illustrated by a preference relation such as

$$(x, \text{no delay}) \sim (y, 1 \text{ day delay}) \sim (z, 2 \text{ days delay}) > (x, \text{no delay}),$$

where  $x < y < z$ . The subadditive discounting predicts greater discounting for a delay when the delay is divided into subdelays and each is evaluated independently, as compared to when the decision maker discounts for a mass of delay at a time. (A possible explanation for this discounting behavior is called *the support theory*, which assumes subjects' limited cognitive power for processing information about alternatives.) Importantly, the author shows that the subadditive discounting offers a better description of observed choice behavior compared to the hyperbolic discounting. As indicated in the example above, an observed preference relation may have cycles under this model of discounting.

Likewise, Rubinstein [29] proposes *a similarity-based time preference* as a potential alternative to the hyperbolic discounting model that provides an intuitive procedure of evaluating dated outcomes. In this model, the decision maker follows up to three steps of procedures in order to compare a pair of dated outcomes. First, the decision maker tests for dominance of two dated outcomes: if  $x > y$  and  $t < s$ , then  $(x, t)$  is preferred to  $(y, s)$ . Provided that there is no dominance between them, he next looks for a similarity of the dated outcomes either in delivery dates or in prizes. For example, if their delivery dates are similar while the prizes are not, the decision maker chooses one that brings the larger prize. Lastly, if neither of the first two steps applies, another criterion is used to evaluate  $(x, t)$  and  $(y, s)$ . The author remarks that the proposed procedures may conflict with transitivity, as the partial criterion determined by the first two steps is not likely to be consistent with one determined by the third step.

In this section, we shall study a general representation of time preferences that contains all discounting models discussed above as special cases. The following definition is due to Ok and Masatlioglu [26].

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

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

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

A function  $\eta$  in the definition above that measures a discount factor between any two points in time is called a *relative discounting function*. Clearly, setting  $\eta(s, t) = \delta(s)/\delta(t)$  reduces a relative discounting time preference to an absolute discounting time preference. As a consequence, the classical discounting models, such as exponential discounting or hyperbolic discounting, are viewed as special cases of this representation. In fact, it is known that a relative discounting time preference is transitive if and only if we can write  $\eta(s, t) = \delta(s)/\delta(t)$  for a some function  $\delta$  ([26, Corollary 1]).

**Example** (Subadditive discounting). The subadditive discounting time preference can be viewed as a relative discounting time preference in which  $\eta(r, t) \geq \eta(r, s)\eta(s, t)$  for every  $r > s > t$ . So, for instance, the decision maker who applies a discount factor of .9 for every single period delay may use a discount factor of .85, instead of .81, for some two-period delays, resulting in cyclic preference relations such as one shown above.

**Example** (Similarity-based time preference). Certain similarity-based time preferences are represented as a relative discounting time preference. For instance, suppose that the decision maker views as similar two delivery dates that are equal or differ by a delay smaller than  $\theta > 0$ , whereas no two distinct prizes are treated as similar. Let

$$\eta(s, t) = \begin{cases} 1 & \text{if } s \leq t + \theta, \\ \delta^{s-t-\theta} & \text{otherwise} \end{cases}$$

for all  $s, t$  with  $s \geq t$ .<sup>13</sup> It follows that  $(1 + 2\epsilon, 2\theta) > (1 + \epsilon, \theta) > (1, 0)$  for all  $\epsilon > 0$ , but  $(1, 0) > (1 + 2\epsilon, 2\theta)$  when  $\epsilon$  is sufficiently small.

Although the relative discounting time preference greatly generalizes the model of time preferences, even to the extent that they encompass cyclic relations, it turns out that the welfare order which we can reliably elicit from them is still consistent with the representation by an absolute discounting. This result suggests that deviation from the classical theory of discounting made by the relative discounting time preference (which is necessary to accommodate the experimental findings) is smaller than it appears.

---

<sup>13</sup>The relative discounting time preference can also represent a similarity-based time preference where two distinct prizes are evaluated as similar by the decision maker. See [26, Example 3].

**Theorem 4.** 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

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

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

Therefore, Theorem 4 shows that the transitive core of a relative discounting time preference is represented by *multi* absolute discounting functions. Indeed, an interpretation of the representation is further clarified by opening the collection  $\mathcal{D}$  of discounting functions. In the proof, it is shown that each  $\delta$  in  $\mathcal{D}$  is  $\eta(\cdot, r)$  for an arbitrary fixed point  $r \geq 0$  in time. So, the representation of the transitive core claims that if, and only if, a discounted utility value of a dated outcome  $(x, t)$  is at least as good as that of  $(y, s)$  evaluated at *all* relative points  $r \geq 0$  in time, we can safely infer that the former is a welfare improvement over the latter.

### 4.3 Justifiable preferences

Let  $\Omega$  be a finite nonempty set of states of nature, and  $Y = \Delta(\mathbb{R})$  be the set of all lotteries on the real line  $\mathbb{R}$ . A function  $f : \Omega \rightarrow Y$  that maps each state to a lottery is referred to as an Anscombe and Aumann [1] *act*, and the collection of all the acts is denoted by  $X$ .<sup>14</sup> An act is a description of state-contingent prize schedules, where likelihood of each state is not objectively known.

Now, consider an organization of agents each of whom has a subjective belief over likelihood of states. In the aim of studying the organization's decision making that is justifiable in this context, Lehrer and Teper [18] propose a model in which the organization prefers one act over another if and only if at least one agent in the organization evaluates the former act higher than the latter according to his subjective belief of states. Formally, a preference relation  $\succsim$  on  $X$  is called *justifiable* if there exist a continuous affine monotone utility function  $u : Y \rightarrow \mathbb{R}$ <sup>15</sup> and a nonempty closed convex set  $P$  of probability distributions over  $\Omega$  such that,

<sup>14</sup>Throughout the section, a Borel probability measure on  $\mathbb{R}$  is called a *lottery*. As usual, the weak topology is assumed on  $Y = \Delta(\mathbb{R})$ , and the set  $X$  of acts is endowed with the product topology.

<sup>15</sup>Note that  $u$  is bounded as an implication of  $u(Y) \subseteq \mathbb{R}$ . Also, we can easily show that, for every  $\mu \in Y$ ,  $u(\mu)$  is an expected utility value of a lottery  $\mu$ , i.e.  $\int u d\mu$ , by affinity and continuity of  $u$ . Lastly, by monotonicity of  $u$ , I mean that  $u$  is strictly increasing when restricted on  $\mathbb{R} \subseteq Y$ , where, with abuse of notation, any real value  $r \in \mathbb{R}$  is identified with a sure lottery that gives  $r$  with probability 1.

for every two acts  $f$  and  $g$ ,

$$f \succcurlyeq g \quad \text{if and only if} \quad \exists p \in P \text{ such that } \sum_{\omega \in \Omega} p(\omega)u(f(\omega)) \geq \sum_{\omega \in \Omega} p(\omega)u(g(\omega)),^{16}$$

Note that, while agents in the organization share the same evaluation  $u$  over lotteries, they might still disagree on evaluations of acts due to different subjective beliefs over states. 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  $g$ . In this case, the justifiable preference of the organization is such that  $f \sim h \sim g$  but  $f \succ g$ . Therefore, a justifiable preference relation for an organization is not transitive in general.

Lehrer and Teper contrast justifiable preferences with Bewley's [8] Knightian preferences. Under a Knightian preference, the organization prefers one act over another if all agents in the organization evaluate the former higher than the latter. Formally, a preference relation  $\succcurlyeq$  on  $X$  is called *Knightian* if there exist a continuous affine monotone utility function  $u : Y \rightarrow \mathbb{R}$  and a nonempty closed convex set  $P$  of probability distributions over  $\Omega$  such that, for every two acts  $f$  and  $g$ ,

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

We can easily observe that a Knightian preference relation is not necessarily complete, but it is always transitive. Lehrer and Teper note that a justifiable preference is a completion of a Knightian preference relation: If  $\succcurlyeq_J$  and  $\succcurlyeq_K$  are a justifiable preference and a Knightian preference associated with the same pair  $(u, P)$  of a utility function and a set of priors, then  $f \succcurlyeq_K g$  implies  $f \succcurlyeq_J g$ .

Therefore, in one hand, though it may contain cyclical evaluations of acts, a justifiable preference relation offers complete pairwise comparisons of acts without making inconsistency with unanimous opinions of agents. On the other hand, a Knightian preference relation provides a transitive ranking over acts that is supported by every agent's prior, although the ranking falls incomplete whenever at least two agents disagree on the evaluations of acts. The transitive core offers a formal connection between these two preference relations.

**Proposition 5.** *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 priors.*

---

<sup>16</sup>The model of justifiable preference relations studied in this section differs slightly from the original work by Lehrer and Teper in that the real line  $\mathbb{R}$  is used for the outcome space for lotteries and that the utility function is assumed to be monotonic on certain lotteries. An essential for a result of this section is the local nonsatiability of the utility function  $u$ . Unboundedness of the domain and monotonicity can be weakened.

When the organization makes its decision by delegating the responsibility of choice to one of the agents, and thus a justifiable preference relation describes the organization's behavior, the welfare of the organization may not be clear due to the existence of preference cycles. The proposition shows that, even if this is the case, the evaluations of acts on which all the agents unanimously agree can be at least viewed as reflecting the welfare of the organization.

It should be remarked that, while the claim of the proposition is fairly reasonable given the model of justifiable preference relations, the transitive core obtains the result without knowing any structure, i.e. a representing pair  $(u, P)$ , of the preference relation. What induces a Knightian preference relation from a justifiable preference relation (and what induces all the other results in this section) is the single axiomatization which is independently proposed as desirable welfare criteria under the abstract framework.

#### 4.4 Regret theory

Let  $\{1, \dots, n\}$  be a finite set of states of the world with  $n \geq 3$ , and suppose that there is a *nature* that resolves a state according to a probability distribution  $p$  over states such that  $p_i > 0$  for every  $i$  and  $\sum_{i=1}^n p_i = 1$ . A *prospect* is a state-contingent prize schedule delivered to a decision maker, formally defined as a real valued function on the set of states. Let  $X = \mathbb{R}^n$  be the set of all prospects, and we shall consider preference relations observed 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 [37], according to which prospects are compared by their expected utility values  $\mathbb{E}(u \circ x)$  for some real function  $u : \mathbb{R} \rightarrow \mathbb{R}$ . The theory has been acknowledged as the model of rational decision making under uncertainty and is justified from the normative perspective. Experimental and empirical studies, however, consistently find a disparity between observed behavior in reality and the prediction of the expected utility theory. The celebrated work of Kahneman and Tversky [16], for example, presents extensive evidence that shows that subjects violate the expected utility hypothesis.

To accommodate the observed violations of the expected utility theory, Loomes and Sugden [19] proposed 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. More specifically, once a state is resolved, the decision maker may regret (or rejoice) if an outcome of the chosen prospect happens to be worse (resp. better) than



State	$u(x_i)$	$u(y_i)$	$u(z_i)$
1	2	0	1
2	1	2	0
3	0	1	2

Table 1: Utility for  $i = 1, 2, 3$

that of the alternative prospect. The theory claims 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, initially proposed in the work of Loomes and Sugden.

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

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

for every  $x, y \in X$ , where  $u$  is increasing homeomorphism with  $u(0) = 0$ , and  $Q$  is convex, strictly increasing and odd.<sup>17</sup>

Obviously, a regret preference admits an expected utility representation when  $Q$  is linear. On the other hand, Loomes and Sugden proved that a regret preference well explains various choice anomalies (such as *the certainty effect* and *the common ratio effect*) as observed in reality with strict convexity of  $Q$ . Also, note that a regret preference relation cannot be transitive when an associated function  $Q$  is strictly convex. To illustrate the last point, 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, it follows that

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

and thus  $x \succ y$ . Intuitively put, the decision maker puts precedence to the great joy of having chosen  $x$  over  $y$  at state 1 over the small regret at state 2 and 3 and, consequently, prefers the former to the latter. (Note that the evaluations of regret and joy are nonlinear due to the strict convexity of  $Q$ .) By symmetry, the same logic applies to pairs  $(y, z)$  and  $(z, x)$ , resulting a preference cycle  $x \succ y \succ z \succ x$ .

As a matter of fact, this observation on the transitivity of a regret preference is a result that holds in general. Bikhchandani and Segal [9] prove that, even with

<sup>17</sup>A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is said to be *odd* when  $f(-a) = -f(a)$  for all  $a \geq 0$ .

further general formulation, a regret preference is transitive if and only if it is represented by an expected utility function. This observation, similarly to the remark made for the relative discounting time preference, implies that every descriptive power of the regret theory attributes to the cyclic structure of the preference relation.

The welfare order for a decision maker whose preference accounts for regret is therefore ambiguous, except when the preference relation admits the expected utility representation. As a consequence, the transitive core may be of use in order for a partial revelation of the decision maker's welfare. Indeed, it confirms the intuitive criterion that we can agree on without specifying any detail of the representation.

**Proposition 6.** *Let  $\succsim$  be a regret preference on  $X$ . Then, for all  $x, y \in X$ ,  $x \geq y$  implies that  $x \text{ core}(\succsim)y$ , and  $x > y$  implies that  $x \text{ core}^*(\succsim)y$ .<sup>18</sup>*

The proposition shows that the dominance relation over prospects is preserved by the transitive core, suggesting that the rule gives a justifiable welfare order, at least when our intuition suffices to make a welfare judgment. Also, it can be shown through examples that the transitive core offers further criterion in general: Two prospects that does not dominate each other may be ordered by the transitive core, thus providing a means of eliciting the welfare order where our intuition does not reach directly.<sup>19</sup>

## 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 society as a representative decision maker and examine its preference relation induced by the majority criterion. Suppose that an individual voter  $i$ , for each  $i \in \{1, \dots, n\}$ , evaluates alternatives 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 lacks transitivity in general and, therefore, fails to give a consistent aggregation of voters' preferences. Furthermore, Arrow [5] gave the celebrated impossibility theorem that proves that there is no voting system satisfying certain desirable criteria at a time. Having the inevitable ambiguity in evaluation of policies, the

<sup>18</sup>I write  $x \geq y$  to mean that  $x_i \geq y_i$  for every state  $i = 1, \dots, n$ , and  $x > y$  that  $x \geq y$  and  $x \neq y$ .

<sup>19</sup>On the other hand, the characterization of the transitive core in terms of primitives of a regret preference, that is,  $(u, Q)$  has not been found yet and is an open question.

Ranking	$\succsim_1$	$\succsim_2$	$\succsim_3$
1	$x$	$w$	$z$
2	$y$	$x$	$w$
3	$z$	$y$	$x$
4	$w$	$z$	$y$

Table 2: Individual preferences

social welfare theory has long been studied in the literature. The purpose of this section is to examine implications of the transitive core in search of a consistent part of the social preference relation.

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

$$x \succsim y \quad \text{if and only if} \quad |\{i : x \succsim_i y\}| \geq |\{i : y \succsim_i x\}|.$$

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

**Remark.** A majority preference may be interpreted instead as a preference relation of an *individual* decision maker over commodities, where each commodity is characterized by  $n$ -many attributes (such as price or quality). For every  $i \in \{1, \dots, n\}$ , the order  $\succsim_i$  then defines a ranking of the alternatives according to the  $i$ th attribute. The majority preference arises from the decision maker who, facing any two commodities, chooses one that beats the other by the number of attributes in which the former is ranked higher than the latter with respect to the orders  $\{\succsim_i\}_{i=1}^n$ .

A majority preference is cyclic in general. Table 2 gives an example of individual preference relations when  $n = 3$  and  $X = \{x, y, z, w\}$ . A majority preference induced for this society ranks  $x \succ z \succ w \succ x$ , making the order of desirability across alternatives ambiguous. Many voting criteria have been proposed to obtain guidance for the social choice in case the welfare value of policies are not immediately clear.

**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  $y$ ; that is,  $x \succ_i y$  for all  $i \in \{1, \dots, n\}$ . 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 by a majority vote; that is,  $x \succ y$  for all  $y \in X \setminus \{x\}$ . The Condorcet principle claims that a Condorcet winner should be a unique choice of the society if one exists. A Condorcet winner may not exist but is always unique.

**Smith's principle.** Let  $S$  and  $T$  be a partition of  $X$  such that every member of  $S$  beats all members of  $T$  by a majority vote; that is,  $x \succ y$  for each  $x \in S$  and  $y \in T$ . Under the Smith's principle, the social choice should be made from  $S$ .<sup>20</sup>

**Exclusive Condorcet principle.** A set  $\{x \in X : y \succ x \text{ for no } y \in X\}$  gathers all the undominated policies by the majority rule. The exclusive Condorcet principle demands that the social choice be made from this set of policies if the set is nonempty.

These four criteria are listed roughly in order of desirability for the social choice. In fact, the Pareto criterion and the Condorcet principle are often used as touchstones for evaluating particular voting systems. Also, the Smith's principle and the exclusive Condorcet principle are stronger than the Condorcet principle in the sense that each of the former implies, but is not implied by, the latter. Fishburn [13] argued in favor of Smith's principle, stating 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."

**Proposition 7.** *The transitive core of a majority preference satisfies*

- (a) *Pareto criterion: if  $x$  is a Pareto improvement over  $y$ ,  $x \text{ core}^*(\succ) y$ ,*
- (b) *Condorcet principle: if  $x$  is a Condorcet winner,  $x \text{ core}^*(\succ) y$  for all  $y \in X \setminus \{x\}$ ,*
- (c) *Smith's principle: if the hypothesis holds,  $x \text{ core}^*(\succ) y$  for all  $x \in S$  and  $y \in T$ ,*

*but not the exclusive Condorcet principle.*

Therefore, the proposition confirms consistency of the transitive core with the first three criteria by showing that an alternative is evaluated to have strictly higher

---

<sup>20</sup>The Smith's principle is originally introduced in [35].

ranking					
(N)	x	a <sub>1</sub>	...	a <sub>k</sub>	y
(N + 1)	y	x	a <sub>1</sub>	...	a <sub>k</sub>
(1)	a <sub>k</sub>	...	a <sub>1</sub>	y	x

Table 3: Exclusive Condorcet principle

value in welfare than another when the appropriate hypothesis is met.<sup>21</sup> As for the exclusive Condorcet principle, while there are many good reasons to support the criterion, there is also a situation in which this principle might hinder the choice of an attractive policy.

**Example.** Consider a society where a policy needs to be chosen from  $\{x, y, a_1, \dots, a_k\}$  for  $2N + 2$  citizens (with  $N$  being a large number). Suppose that individual preferences of this society are distributed as in Table 3. It will be noticed that everyone except one citizen ranks a policy  $x$  second or higher, while  $y$  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 leaves the welfare values for  $x$  and  $y$  *incomparable* and, thus, maintains an option to choose  $x$  as well as  $y$ .

As remarked in the introduction, the transitive core is a cautious method of welfare inference that allows an inferred welfare order to be incomplete. For the present context, however, many alternative methods are proposed in the literature, especially in an attempt to provide finer criteria for the social choice. For example, Rubinstein [29] studies a ranking of policies by *the point system*, defined as

$$x \sigma(\succsim) y \quad \text{if and only if} \quad |L(x, \succsim)| \geq |L(y, \succsim)|$$

for every  $x$  and  $y$  in  $X$ , assuming that  $\succsim$  is antisymmetric and  $X$  is finite.<sup>22</sup> Then,  $\sigma(\succsim)$  is clearly a weak order on  $X$ . Moreover,  $\sigma(\succsim)$  is a completion of the transitive core so that  $x \text{ core}(\succsim) y$  implies  $x \sigma(\succsim) y$  and that  $x \text{ core}^*(\succsim) y$  implies  $x \sigma^*(\succsim) y$ , where  $\text{core}^*(\succsim)$  and  $\sigma^*(\succsim)$  denote the strict parts of  $\text{core}(\succsim)$  and  $\sigma(\succsim)$ , respectively.

<sup>21</sup>On the other hand, the characterization of the transitive core applied to a majority preference relation has not been found yet. As a matter of fact, McGarvey [24] shows that any complete binary relation on  $X$  is obtained as a majority preference under some society. Therefore, no structure on a majority preference can be assumed ex ante, making it even more difficult to obtain the general characterization.

<sup>22</sup>Recall that  $L(x, \succsim) = \{y : x \succsim y\}$  denotes the lower contour set of a preference relation  $\succsim$  at  $x$ .

Therefore, a part of a preference relation that the transitive core judges safe to be seen as the welfare order is, indeed, preserved by the finer criterion of welfare inference. This provides a good reason to call the transitive core as a cautious rule of welfare inference.<sup>23</sup>

## 5 Unambiguous welfare improvement

Reflecting an interest in behavioral models of decision makers, behavioral welfare economics has attracted great attention in the literature. Bernheim and Rangel [6, 7] focus on developing a framework of behavioral welfare analysis that extends the classical approach. Notably, they shed light on a generalized choice situation where a decision maker might make a different choice from the same choice set contingent on ancillary conditions. Given the extended choice data, they introduce a welfare improvement order that captures an unambiguous part of comparisons and apply it to carry out welfare analysis. Although Bernheim and Rangel view the framing effects, and not necessarily nontransitive preference relations, as the main source of behavioral decision making, we can nevertheless apply their method to choice behavior induced by a nontransitive preference relation. For this, we induce a *top-cycle choice* from a preference relation. Let  $\succsim$  be a preference relation on  $X$ , and  $S$  be a nonempty subset of  $X$ . For any  $x, y \in S$ , an alternative  $x$  is said to be *indirectly preferred* in  $S$  to another alternative  $y$  if there is a finite sequence  $(z_i)$  in  $S$  such that  $x = z_1 \succsim z_2 \succsim \cdots \succsim z_k = y$ . The *top-cycle choice* from  $S$  by the preference relation  $\succsim$  is then defined as the set of all alternatives in  $S$  that are indirectly preferred to every other alternative in  $S$ .<sup>24</sup> The top-cycle choice induces a choice correspondence on  $X$  (i.e. a map that assigns a nonempty subset of  $S$  to every nonempty subset  $S$  of  $X$ ) from a preference relation on  $X$ , which we may interpret as a behavioral choice procedure by the decision

<sup>23</sup>On the other hand, there is a case where the point system might be viewed as too bold in making welfare evaluations. For example, take a preference relation  $\succsim$  on  $\{1, \dots, n\}$  considered in the last remark in Section 3. Having the preference relation identical to the descending order except for a “small” mistake  $1 \succ n$ , it might be plausible to understand that the right welfare order  $>$  is one that  $n > \cdots > 1$ . The point system disagrees with this order  $>$ , whereas the transitive core, as a method of cautious welfare inference, accepts both orders as its completion. Also, for a majority preference induced from the society given in Table 3, the point system rejects an attractive second best alternative  $x$  as an inferior alternative to  $y$ , whereas the transitive core leaves them incomparable.

<sup>24</sup>Succinctly, the top-cycle choice from  $S$  is the maximizers of the transitive closure of  $\succsim$  relative to  $S$ .

maker who is endowed with a (possibly cyclic) preference relation  $\succsim$ . Note that the top-cycle choice chooses  $x$  from a binary set  $\{x, y\}$  if and only if  $x \succsim y$ . So, whenever  $\succsim$  is cyclic, the top-cycle choice indeed entails cyclic choice patterns. To extend the choice-based welfare analysis beyond the standard models, Bernheim and Rangel [6, 7] propose *the unambiguous improvement order*. Formally, given a choice correspondence  $C$  on  $X$ , they say that an alternative  $x$  in  $X$  is *an unambiguous welfare improvement* over another alternative  $y$  in  $X$  if  $y \notin C(S)$  whenever  $x \in S$ . Now, we can easily verify that, if a choice correspondence is a top-cycle choice induced from a preference relation  $\succsim$ , the unambiguous welfare improvement order agrees with the strict part of the preference relation  $\succsim$  which is not involved in any preference cycles. So, by the consistency axiom, this is included by the criterion offered by the transitive core. Moreover, we can show that, for any cyclic preference relation that is either a semiorder, a relative discounting time preference, or a regret preference defined as in Section 4, *no preference is free from cycles, and the unambiguous welfare improvement order is the empty relation*. Hence, it turns out that the unambiguous welfare improvement order is so cautious that it makes very little welfare judgement when the criterion is applied to nontransitive preference relations.<sup>25</sup>

## 6 Related literature

This paper relates to three branches of the economic literature. The first encompasses the descriptive studies that provide supporting evidence for nontransitive preference relations in practice. The second is literature in behavioral welfare economics, which attempts to obtain a welfare order for behavioral decision makers. The third is the social choice theory, which provides extensive welfare analysis for a group of individuals who do not share the same interest.

**Descriptive analysis.** The descriptive studies of nontransitive preference relations, in general, support the claim that decision makers are, in practice, endowed with a cyclic preference relation, often due to a certain decision process that is reasonable in the respective contexts. Several works in this branch were cited in Section 4, with the analysis of the performance of the transitive core. It is important to note that, while the history of the theory of nontransitive preference

---

<sup>25</sup>Note that the transitive core is not a method that merely takes preferences free from cycles as the welfare order. Preferences can be viewed as a part of the welfare order even if they form preference cycles. See the remark at the end of Section 3.

relations is quite long, there is still substantial interest in the model of nontransitive decision makers. In a recent paper, for example, Manzini and Mariotti [22] study choice behavior obtained by lexicographic application of semiorders and give its axiomatic characterization. With regard to the regret theory, Hayashi [15] examines a choice of prospects made by a decision maker whose motive is driven by anticipated ex post regrets. In particular, he proposes a model that is flexible enough to allow both regret aversion and nontrivial likelihood judgement.<sup>26</sup>

Importantly, the findings of this literature motivate the present paper: The method of welfare inference from nontransitive preference relations is of interest because of the evidence that, in contrast to the classical utility theory, a number of decision makers, in reality, suffer from cyclic preferences. In this sense, the current paper complements the descriptive analysis of nontransitive preferences.

**Behavioral welfare economics.** Bernheim and Rangel [6, 7] study an environment in which the decision maker's choice behavior is affected by ancillary conditions and develop a framework of behavioral welfare analysis that extends the classical approach. (See Section 5 for further discussion of their welfare criterion.) Likewise, Rubinstein and Salant [32] study an individual decision maker who behaves differently in different circumstances. They assume a set of preference relations as observable data (where observed preference relations account for the same decision maker's behavior under different frames) and seek an unobservable welfare order that underlies her behavior.<sup>27</sup> Moreover, Kőszegi and Rabin [17] study a method of distinguishing choice that reveals welfare of the decision maker from one that is made due to an error, whereas Chambers and Hayashi [11] investigate welfare criteria that apply to random choice data. We can, however, quickly note that these papers focus on aspects of behavioral decision making that are different from one this paper does. Due to this difference, the method proposed in this paper is not implied by the methods studied in the other works, and vice versa.

---

<sup>26</sup>In contrast to an objective probability model I adopt for the regret theory in this paper, Hayashi studies a subjective probability model *à la* Savage so that an aggregation of ex post regrets is a nontrivial issue.

<sup>27</sup>A significant difference between these two works is that Bernheim and Rangel seek an ambiguous welfare order by a model-free Pareto approach, while Rubinstein and Salant aim to make finer welfare judgments with the help of testable behavioral assumptions. The current paper, in this respect, adopts a similar approach to the former.



**Social choice theory.** Social choice theory has long been investigated, along with many proposed voting systems and a number of welfare criteria. The theory grows with welfare economics, providing a sound foundation for welfare analysis on policies that affect utilities of society members. The theory often adopts as its starting point a set of individuals who do not necessarily share a common interest and attempts to aggregate their opinions to a collective choice that maximizes the social welfare (at least in some measure). However, we should note that the methods developed in the social choice theory may not be applicable to welfare problems in other contexts due to the assumed structure of the approach. Many voting systems are defined on a basis of voters' preference relations and, thus, are of little use in eliciting the welfare order from an *individual* preference relation.

Having mentioned this, there are still some methods developed in the social choice theory that can be applied to other contexts. A notable example, mainly because of its close resemblance to the method of welfare inference proposed in this paper, is *the covering order* by Fishburn [13] and Miller [25]. For any preference relation  $\succsim$  on a set  $S$ , and for every  $x, y \in S$ , we say that  $x$  covers  $y$  if  $L(y, \succsim) \subseteq L(x, \succsim)$ ; and the set of alternatives in  $S$  not strictly covered by any other alternative is called *the uncovered set* of  $S$ .<sup>28</sup> They show that the uncovered set has many desirable properties and that several important voting systems choose a policy from this set. Interestingly, regardless of the obvious resemblance to the transitive core in the definition (cf. Fact 4), these two operations exhibit quite different features both inside and outside the social choice context. For example, Fishburn shows that the covering order satisfies the exclusive Condorcet principle, whereas the transitive core does not, as indicated in the previous section. Also, outside the social choice context, for any given semiorder on a commodity space  $\mathbb{R}^n$ , the uncovered set computed for a nontrivial budget set (a compact convex subset of  $\mathbb{R}^n$ ) never agrees with the set of maximizers of an associated utility function  $u$ , while the optimization of the transitive core always provides the set of utility maximizers.<sup>29</sup>

---

<sup>28</sup>Galaabaatar and Karni [14] also study an operation similar to the covering order for distinguishing indifference and incomparability in an environment where the decision maker expresses an incomplete strict preference relation.

<sup>29</sup>I mean by a nontrivial budget set  $B$ , one meeting the condition  $\sup u(B) - \inf u(B) > \epsilon$  as in Proposition 3. However, in general, the uncovered set always coincides with the maximizer of the *original* semiorder, and, thus, it agrees with the maximizer of the utility function  $u$  only if  $u$  is constant on  $B$ .

## 7 Concluding remarks

This paper studies a method of inferring welfare orders from nontransitive preference relations of a decision maker, so that the welfare analysis can proceed even when an observed preference relation is cyclic. The welfare evaluation rule, a function that maps a complete but nontransitive preference relation to a transitive welfare order, is defined as a representation of such a method, and certain desirable properties of the rule are discussed. The transitive core is obtained as a unique welfare evaluation rule that satisfies all the proposed properties of the rule at the same time. Furthermore, a simple criterion to compute the transitive core is provided to facilitate its applications.

The transitive core is applied to a variety of nontransitive preference relations. It is shown that, when the transitive core is applied to a semiorder, it infers an underlying utility function that measures the values of alternatives for the decision maker. Also, the transitive core of a relative discounting time preference turns out to be represented by multi general discounting functions, thereby re-discovering the classical theory of time preference within the descriptive discounting model. The same rule of welfare inference is verified to make intuitive welfare judgment for regret preference relations, justifiable preference relations, and collective preference relations obtained through the majority criterion.

As a final remark, I again note that this paper targets cyclic preferences as the source of behavioral aspects in decision making. This feature distinguishes the current paper from the works of Bernheim and Rangel [7] and many others which focus on different types of behavioral decision making procedures such as one under the framing effect. The cause of behavioral decision making is, however, by no means exhausted by the cyclic preference relations or the framing effects. Finding an intuitive and justifiable method of welfare inference for other sources of boundedly rational choice behavior is an open problem of great interest in behavioral welfare economics.<sup>30</sup>

---

<sup>30</sup>To illustrate the problem, for example, an application of the transitive core to a choice behavior under the limited attention studied by Masatlioglu et al. [23], where a preference relation is revealed by the choice behavior on binary sets, may make a strict mistake, in a sense that the transitive core judges an alternative  $y$  to be a strict welfare improvement over another alternative  $x$  even when  $u(x) > u(y)$  for any utility representation of the choice behavior with an attention filter. (They also indicated that the welfare criterion proposed by Bernheim and Rangel [7] makes a strict mistake in the same context. See [23, Example 1]).

## Appendix A: Proofs

*Proof of Theorem 1.* The proof consists of proving the following four steps.

- 1) Let  $\sigma$  and  $\sigma'$  be two WERs that admit Axioms 1-6. If  $\sigma(\succsim) = \sigma'(\succsim)$  for all  $\succsim \in \mathcal{P}$  with  $|D_{\succsim}| = 3$ , then  $\sigma = \sigma'$ .
- 2) Let  $\sigma$  be a WER that admits Axioms 1-6. Take any  $x, y, z \in X$ , and suppose that  $\succsim$  is such that  $x \succ y \succ z \succ x$ . Then,  $\sigma(\succsim) = \emptyset$ .
- 3) Let  $\sigma$  be a WER that admits Axioms 1-6. Take any  $x, y, z \in X$ , and suppose that  $\succsim$  is such that  $x \succ y \sim z \sim x$ . Then,  $\sigma(\succsim) = \{(x, y), (y, z), (x, z)\}$ .
- 4) Let  $\sigma$  be a WER that admits Axioms 1-6. Take any  $x, y, z \in X$ , and suppose that  $\succsim$  is such that  $x \succ y \succ z \sim x$ . Then,  $\sigma(\succsim) = \{(x, z)\}$ .

The first step claims that two welfare evaluation rules are identical if they coincide on every complete preference relation on a set of three alternatives. This is an immediate implication from the reduction and extension axioms, and the proof is omitted. Having the result of the first step, all we have to prove is that the axioms uniquely identify the welfare order  $\sigma(\succsim)$  for every complete preference relation  $\succsim$  defined on a triangle. Moreover, we can restrict our attention into cyclical preference relations, for  $\sigma(\succsim) = \succsim$  by Fact 1 if  $\succsim$  happens to be transitive. There are only three types of cyclical preference relations on a triangle, and the rest of the steps handle each of these cases. Step 2 considers a preference relation  $\succsim_1$  in Figure 2. The prudence and anonymity axioms immediately conclude that the welfare evaluation rule can preserve none for this preference relation. Step 3 takes a preference relation  $\succsim_1$  in Figure 3. This step was already verified in a remark in Section 3.3. We shall prove Step 4 in what follows.

Figure 5 gathers all preference relations used in the course of the proof. I denote  $\succsim_{ij}$  to specify a preference relation of the  $i$ th row and the  $j$ th column in the figure. So,  $\succsim_{14} = \{(x, z)\}$ , for example. We wish to show that  $\sigma(\succsim_{11}) = \succsim_{14}$  for arbitrary  $x, y, z \in X$ . To this end, take and fix any  $x, y, z \in X$ . First, we show that  $\sigma(\succsim_{11})$  must be either  $\succsim_{14}$ ,  $\succsim_{21}$  or  $\succsim_{22}$ . To see this, note that a preference  $(x, z)$  of  $\succsim_{11}$  is not involved in any cycle of  $\succsim_{11}$ , and thus  $x \sigma(\succsim_{11}) z$  by the consistency axiom. Moreover, it is implied that

$$x \sigma(\succsim_{11}) y \stackrel{\text{symmetry}}{\iff} y \sigma(\succsim_{12}) x \stackrel{\text{anonymity}}{\iff} y \sigma(\succsim_{13}) z$$

by the symmetry and anonymity axioms. But  $\succsim_{11} = \succsim_{13}$  and, thus, preferences  $(x, y)$  and  $(y, z)$  are either both accepted or both rejected by the welfare order

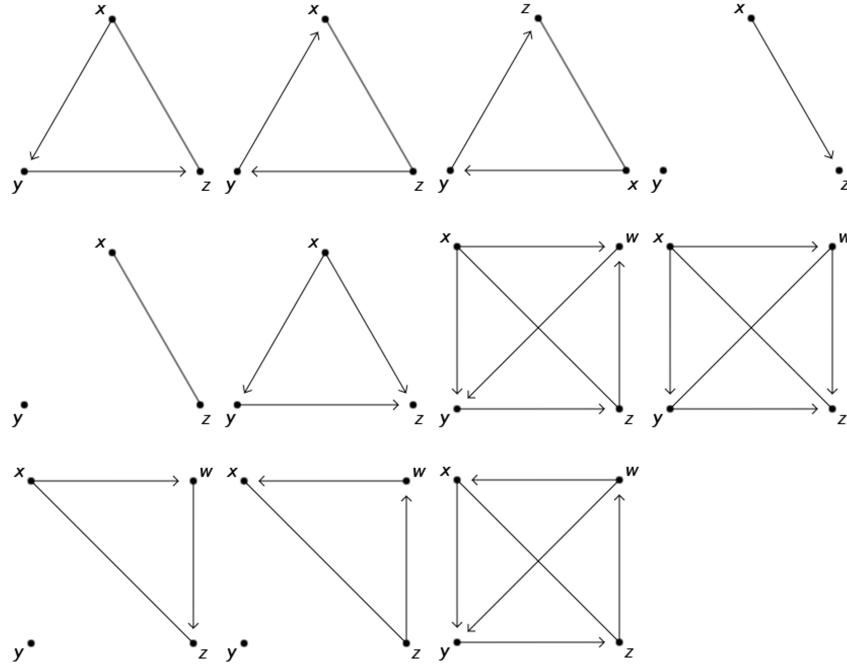


Figure 5: Proof of Theorem 1

$\sigma(\succsim_{11})$ . These two observations along with the prudence axiom and transitivity of  $\sigma(\succsim_{11})$  together imply that  $\sigma(\succsim_{11}) \in \{\succsim_{14}, \succsim_{21}, \succsim_{22}\}$ .

Next, we show that  $\sigma(\succsim_{11})$  cannot be  $\succsim_{21}$ . Suppose the contrary. Pick any  $w \in X$  distinct from  $x, y, z$ , and we shall consider a preference relation  $\succsim_{23}$ .<sup>31</sup> It follows from the extension axiom that  $z \sigma(\succsim_{23}) x$  and  $x \sigma(\succsim_{23}) w$ .<sup>32</sup> Since  $\sigma(\succsim_{23})$  is transitive, this implies  $z \sigma(\succsim_{23}) w$ . Now, consider the restriction of  $\succsim_{23}$  on  $\{y, z, w\}$ , and observe that  $(z, w)$  is preserved by the welfare order of this restriction by the reduction axiom, contradicting the result from Step 2.

Lastly, we show that  $\sigma(\succsim_{11})$  cannot be  $\succsim_{22}$ . Assume the contrary. Pick any  $w \in X$  distinct from  $x, y, z$ , and we shall prove the next claim.

*Claim.*  $\sigma(\succsim_{32}) = \{(z, x), (z, w), (w, x)\}$ .

The claim says that, when we assume  $\sigma(\succsim_{11}) = \succsim_{22}$ , the welfare order of the same form must be revealed from  $\succsim_{32}$ . (Note that this is not an immediate im-

<sup>31</sup>Recall the assumption that  $|X| > 3$ . Theorem 1 is false when  $|X| = 3$ .

<sup>32</sup>This part also uses Fact 1:  $\sigma(\succ) = \succ$  for any transitive preference relation  $\succ$ .

plication from the anonymity axiom.) To show this, consider  $\succsim_{24}$ . It follows that  $x \sigma(\succsim_{24}) y$  and  $y \sigma(\succsim_{24}) z$  by the extension axiom, which implies that  $x \sigma(\succsim_{24}) w$  and  $w \sigma(\succsim_{24}) z$  by the anonymity axiom.<sup>33</sup> The reduction axiom then implies that  $x \sigma(\succsim_{31}) w$  and  $w \sigma(\succsim_{31}) z$ , and in addition  $x \sigma(\succsim_{31}) z$  by transitivity of  $\sigma(\succsim_{31})$ , concluding that  $\sigma(\succsim_{31}) = \{(x, w), (x, z), (w, z)\}$ . Now, applying the anonymity axiom on  $\succsim_{31}$  and  $\succsim_{32}$  verifies the desired claim.

To conclude the proof, consider  $\succsim_{33}$ . It then follows that  $w \sigma(\succsim_{33}) x$  and  $x \sigma(\succsim_{33}) y$  by the extension axiom, and  $w \sigma(\succsim_{33}) y$  by transitivity of  $\sigma(\succsim_{33})$ . The reduction axiom implies that  $(w, y)$  is preserved in the welfare order revealed for the restriction of  $\succsim_{33}$  on  $\{y, z, w\}$ , contradicting Step 2. Hence,  $\sigma(\succsim_{11})$  cannot be  $\succsim_{22}$ .

As a result,  $\sigma(\succsim_{11}) = \succsim_{14}$ , and Step 4 is verified. The proof is now complete.<sup>34</sup>

□

*Proof of Proposition 2.* Define a map  $\sigma$  on  $\mathcal{P}$  by

$$\sigma(\succsim) = \{(x, y) \in \succsim : U(x, \succsim) \subseteq U(y, \succsim) \text{ and } L(y, \succsim) \subseteq L(x, \succsim)\}$$

for every  $\succsim \in \mathcal{P}$ . By Theorem 1 and the definition of the transitive core, all we have to show is that  $\sigma$  is a welfare evaluation rule satisfying Axioms 1-6. Take any  $\succsim \in \mathcal{P}$ . Then,  $\sigma(\succsim)$  is a preference relation on  $D_{\succsim}$ , for  $\sigma(\succsim) \subseteq \succsim \subseteq D_{\succsim} \times D_{\succsim}$ . Also, it is obvious that  $\sigma(\succsim)$  is reflexive. To show that  $\sigma(\succsim)$  is transitive, let  $x \sigma(\succsim) y \sigma(\succsim) z$ . It follows that  $U(x, \succsim) \subseteq U(z, \succsim)$  and  $L(z, \succsim) \subseteq L(x, \succsim)$ . As  $\succsim$  is complete,  $x \in U(x, \succsim) \subseteq U(z, \succsim)$  and thus  $x \succsim z$ . So,  $x \sigma(\succsim) z$ , verifying transitivity of  $\sigma(\succsim)$ . This proves that  $\sigma$  is indeed a welfare evaluation rule. The rest verifies each of the axioms for the rule  $\sigma$ . Take any  $S, T \subseteq X$  with  $T \subseteq S$  and any complete preference relation  $\succsim, \succsim'$  on  $S$ .

*Anonymity.* Suppose that there exists a permutation  $\pi$  on  $S$  such that  $x \succsim y$  if and only if  $\pi(x) \succsim' \pi(y)$  for any  $x, y \in X$ . If  $U(x, \succsim) \subseteq U(y, \succsim)$  for some  $x, y \in X$ , we can follow a series of implications

$$\begin{aligned} z \in U(\pi(x), \succsim') &\Rightarrow z \succsim' \pi(x) \Rightarrow \pi^{-1}(z) \succsim x \\ &\Rightarrow \pi^{-1}(z) \succsim y \Rightarrow z \succsim' \pi(y) \Rightarrow z \in U(\pi(y), \succsim') \end{aligned}$$

<sup>33</sup>Use relabeling that swaps  $y$  and  $w$ , and note that the relabeled relation is again  $\succsim_{24}$ .

<sup>34</sup>Formally speaking, the proof shows only that a contradiction arises whenever  $\sigma(\succsim_{11}) \neq \succsim_{14}$  and does not verify that the condition  $\sigma(\succsim_{11}) = \succsim_{14}$  implies no contradiction. If it does, then it means that a welfare evaluation rule satisfying Axioms 1-6 does not exist. In either case, the uniqueness is guaranteed.

to verify  $U(\pi(x), \succ') \subseteq U(\pi(y), \succ')$ . The converse follows by symmetry, and the same equivalence holds for the lower contour set by duality. It thus follows that

$$U(x, \succ) \subseteq U(y, \succ) \Leftrightarrow U(\pi(x), \succ') \subseteq U(\pi(y), \succ')$$

and

$$L(y, \succ) \subseteq L(x, \succ) \Leftrightarrow L(\pi(y), \succ') \subseteq L(\pi(x), \succ'),$$

from which we can easily show that  $x \sigma(\succ) y$  if and only if  $\pi(x) \sigma(\succ') \pi(y)$ .

*Symmetry.* Suppose that  $x \succ y$  if and only if  $y \succ' x$  for all  $x, y \in X$ . Obviously, we have  $U(z, \succ) = L(z, \succ')$  and  $L(z, \succ) = U(z, \succ')$  for any  $z \in X$ , which implies that

$$U(x, \succ) \subseteq U(y, \succ) \Leftrightarrow L(x, \succ') \subseteq L(y, \succ')$$

and

$$L(y, \succ) \subseteq L(x, \succ) \Leftrightarrow U(y, \succ') \subseteq U(x, \succ').$$

So,  $x \sigma(\succ) y$  if and only if  $y \sigma(\succ') x$ , and the axiom is verified.

*Prudence.* By definition,  $\sigma(\succ) \subseteq \succ$ , and this is what we want to show.

*Consistency.* Let  $(x, y) \in \succ$  be involved in no cycle of  $\succ$ , and suppose by contradiction that  $x \sigma(\succ) y$  does not hold. Then, either  $U(x, \succ) \not\subseteq U(y, \succ)$  or  $L(y, \succ) \not\subseteq L(x, \succ)$ . If the former holds, then there exists a  $z \in X$  such that  $x \succ y \succ z \succ x$ . If the latter holds, then there exists a  $z \in X$  such that  $x \succ y \succ z \succ x$ . In either case,  $(x, y)$  is involved in a cycle  $(x, y, z)$  of  $\succ$ , a contradiction.

*Reduction.* Note that  $U(z, \succ_T) = U(z, \succ) \cap T$  and  $L(z, \succ_T) = L(z, \succ) \cap T$  for any  $z \in T$ , which implies that whenever  $x, y \in T$ ,

$$U(x, \succ) \subseteq U(y, \succ) \Rightarrow U(x, \succ_T) \subseteq U(y, \succ_T)$$

and

$$L(y, \succ) \subseteq L(x, \succ) \Rightarrow L(y, \succ_T) \subseteq L(x, \succ_T).$$

So, if  $x \sigma(\succ) y$  and  $x, y \in T$ , then  $x \sigma(\succ_T) y$ , confirming the reduction axiom.

*Extension.* Suppose that  $x \sigma(\succ_{xyz}) y$  for every  $z \in S$ . Let  $z$  be any alternative in  $X$  such that  $z \succ x$ . Then,  $z \succ_{xyz} x$ , and hence the hypothesis implies that  $z \succ_{xyz} y$  and  $z \succ y$ . As  $z$  is arbitrary,  $U(x, \succ) \subseteq U(y, \succ)$ . Dually, it follows that  $L(y, \succ) \subseteq L(x, \succ)$ . The two inclusions show that  $x \sigma(\succ) y$ , as required.

So,  $\sigma$  is a welfare evaluation rule satisfying Axioms 1-6, and we are done.  $\square$

*Proof of Proposition 3.* Let  $X$  be a connected metric space and  $\succcurlyeq$  be a semiorder on  $X$  with a representation  $(u, \epsilon)$ , where  $\sup u(X) - \inf u(X) > 2\epsilon$ . Take any  $x, y \in X$  with  $u(x) \geq u(y)$ . If  $z \in X$  is such that  $z \succcurlyeq x$ , then  $u(z) \geq u(x) - \epsilon \geq u(y) - \epsilon$ , and hence  $z \succcurlyeq y$ . If  $z \in X$  is such that  $y \succcurlyeq z$ , then  $u(x) \geq u(y) \geq u(z) - \epsilon$  and thus  $x \succcurlyeq z$ . So,  $x \text{ core}(\succcurlyeq) y$  by Proposition 2. For the converse, take any  $x, y \in X$  with  $u(y) > u(x)$ , and we shall show that not  $x \text{ core}(\succcurlyeq) y$ . Note that, by the hypothesis on the width of  $u(X)$ , 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 with the latter inequality is similar and thus omitted.) 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  $u$ . It follows that  $y \succcurlyeq z > x$ , negating  $x \text{ core}(\succcurlyeq) y$  by Proposition 2. The proof is complete.  $\square$

*Proof of Theorem 4.* Let  $\succcurlyeq$  be a relative discounting time preference with an associated representation  $(u, \eta)$ . Define  $\mathcal{D} := \{\eta(\cdot, r) : r \in [0, \infty)\}$ . Then, every  $\delta \in \mathcal{D}$  is a continuous function from  $\mathbb{R}_+$  to  $\mathbb{R}_{++}$ . Suppose that  $(x, t), (y, s) \in X$  are dated outcomes with  $(x, t) \text{ core}(\succcurlyeq) (y, s)$ . Take any  $\delta \in \mathcal{D}$ , and let  $r \in [0, \infty)$  be such that  $\delta(\cdot) = \eta(\cdot, r)$ . Since  $u$  is a homeomorphism from  $Z$  to  $\mathbb{R}_{++}$ , there exists a  $z \in Z$  such that  $u(z) = \eta(s, r)u(y)$ . Then,  $(y, s) \succcurlyeq (z, r)$ , which in turn implies that  $(x, t) \succcurlyeq (z, r)$  by Proposition 2. Hence,  $\delta(t)u(x) = \eta(t, r)u(x) \geq u(z) = \eta(s, r)u(y) = \delta(s)u(y)$  as desired. For the converse, take any  $(x, t), (y, s) \in X$  with  $\delta(t)u(x) \geq \delta(s)u(y)$  for every  $\delta \in \mathcal{D}$ . Let  $(z, r)$  be any dated outcome in  $X$  with  $(y, s) \succcurlyeq (z, r)$ . Then,  $\eta(s, r)u(y) \geq u(z)$ , while  $\eta(t, r)u(x) \geq \eta(s, r)u(y)$  as  $\eta(\cdot, r) \in \mathcal{D}$ . So,  $\eta(t, r)u(x) \geq u(z)$  and  $(x, t) \succcurlyeq (z, r)$ . Similarly, if  $(z, r) \succcurlyeq (x, t)$ , then  $(z, r) \succcurlyeq (y, s)$ . By Proposition 2, it follows that  $(x, t) \text{ core}(\succcurlyeq) (y, s)$ .  $\square$

*Proof of Proposition 5.* Let  $\succcurlyeq_J$  and  $\succcurlyeq_K$  be a justifiable preference relation and a Knightian preference relation on  $X$ , respectively, with  $(u, P)$  being a representing pair of a utility function and a set of priors. In this section, to reduce notation, I 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  $\text{core}(\succcurlyeq_J) = \succcurlyeq_K$ . First, take any  $f, g \in X$  with  $f \succcurlyeq_K g$ , that is,  $\varphi(f, p) \geq \varphi(g, p)$  for all  $p \in P$ . If  $h \in X$  is such that  $h \succcurlyeq_J f$ , then there exists a  $p \in P$  with  $\varphi(h, p) \geq \varphi(f, p) \geq \varphi(g, p)$ , and thus  $h \succcurlyeq_J g$ . If  $h \in X$  is such that  $g \succcurlyeq_J h$ , then there exists a  $p \in P$  with  $\varphi(f, p) \geq \varphi(g, p) \geq \varphi(h, p)$ , so  $f \succcurlyeq_J g$ . Therefore, by Proposition 2, we have  $f \text{ core}(\succcurlyeq_J) g$ . Next, take any  $f, g \in X$  with  $f \text{ core}(\succcurlyeq_J) g$ , and suppose that  $f \succcurlyeq_K g$  does not hold by contradiction. Then, there exists a  $p^* \in P$  with  $\varphi(g, p^*) > \varphi(f, p^*)$ . For a positive real number  $\epsilon > 0$ , let  $f^\epsilon$  be an

act that gives an outcome larger than that of  $f$  by  $\epsilon$  for every realization of state  $s \in S$  and every resolution of lottery  $f(s)$ .<sup>35</sup> By continuity of  $u$ , we can choose an  $\epsilon > 0$  small enough so that  $\varphi(g, p^*) > \varphi(f^\epsilon, p^*)$ . 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 stochastically dominates  $f(s)$  at every state  $s \in S$ . Hence,  $g \succcurlyeq_J f^\epsilon \succ_J f$ , contradicting  $f \text{ core}(\succcurlyeq_J) g$ . So,  $f \text{ core}(\succcurlyeq_J) g$  implies  $f \succcurlyeq_K g$ .  $\square$

*Proof of Proposition 6.* Let  $X = \mathbb{R}^n$  be the set of prospects and  $\succcurlyeq$  be a regret preference on  $X$  with an associated representation  $(u, Q)$ . Let  $x, y \in X$  be two prospects such that  $x \geq y$ . If  $z \in X$  is such that  $y \succcurlyeq z$ , then by monotonicity of  $u$  and  $Q$ ,

$$\sum_{i=1}^n p_i Q(u(x_i) - u(z_i)) \geq \sum_{i=1}^n p_i Q(u(y_i) - u(z_i)) \geq 0$$

and hence  $x \succcurlyeq z$ . Similarly, if  $z \succcurlyeq x$ ,  $\sum_{i=1}^n p_i Q(u(z_i) - u(y_i)) \geq \sum_{i=1}^n p_i Q(u(z_i) - u(x_i)) \geq 0$  and  $z \succcurlyeq y$ . By Proposition 2,  $x \text{ core}(\succcurlyeq) y$ . If  $x > y$ , then  $\sum_{i=1}^n p_i Q(u(x_i) - u(y_i)) > 0$ , for  $u$  and  $Q$  are strictly increasing and  $Q(0) = 0$ .<sup>36</sup> It follows that  $x > y$  and not  $y \text{ core}(\succcurlyeq) x$  by the prudence axiom. We have  $x \text{ core}^*(\succcurlyeq) y$  as desired.  $\square$

*Proof of Proposition 7.* Let  $X$  be an arbitrary nonempty set, and  $\succcurlyeq$  be a majority preference induced by a set of linear orders  $\succcurlyeq_i$  for  $i = 1, \dots, n$ . To verify Pareto criterion, suppose that  $x, y \in X$  are two policies with  $x \succ_i y$  for all  $i$ . Observe that, for an arbitrary policy  $z \in X$ ,  $\{i : y \succcurlyeq_i z\} \subseteq \{i : x \succcurlyeq_i z\}$  and  $\{i : z \succcurlyeq_i x\} \subseteq \{i : z \succcurlyeq_i y\}$  as each  $\succcurlyeq_i$  is transitive. So, if  $y \succcurlyeq z$ , then

$$|\{i : x \succcurlyeq_i z\}| \geq |\{i : y \succcurlyeq_i z\}| \geq |\{i : z \succcurlyeq_i y\}| \geq |\{i : z \succcurlyeq_i x\}|$$

and thus  $x \succcurlyeq z$ . We can similarly show that  $z \succcurlyeq x$  implies  $z \succcurlyeq y$ . It follows that  $x \text{ core}(\succcurlyeq) y$  by Proposition 2. Obviously,  $x > y$  and not  $y \text{ core}(\succcurlyeq) x$  by the prudence axiom. So,  $x \text{ core}^*(\succcurlyeq) y$ . Next, to check Smith's principle, let  $S$  and  $T$  be a partition of  $X$  such that  $x > y$  for each  $x \in S$  and  $y \in T$ . Take any  $x \in S$  and  $y \in T$ . If  $z \in X$  is such that  $y \succcurlyeq z$ , then  $z \in T$  and  $x > z$  by construction. Similarly, if  $z \in X$  is such that  $z \succcurlyeq x$ , then  $z \in S$  and  $z > y$ . So, by Proposition 2,  $x \text{ core}(\succcurlyeq) y$ . Clearly,  $x > y$  and thus not  $y \text{ core}(\succcurlyeq) x$  by the prudence axiom. So,  $x \text{ core}^*(\succcurlyeq) y$  as required. Condorcet principle is implied by Smith's principle. (Set  $S$  as a singleton of a Condorcet winner.) A counterexample to the exclusive Condorcet principle is given in Section 4.5.  $\square$

<sup>35</sup>Formally, for any Borel probability measure  $\mu$  on  $\mathbb{R}$ , let  $\mu^\epsilon$  be a Borel probability measure such that  $\mu^\epsilon(B + \epsilon) = \mu(B)$  for all Borel set  $B$  in  $\mathbb{R}$ . Then,  $f^\epsilon$  is defined as an act that maps every state  $s \in S$  to a Borel probability measure  $f(s)^\epsilon$ .

<sup>36</sup>Note that an increasing homeomorphism from  $\mathbb{R}$  to  $\mathbb{R}$  must be strictly increasing. Also, the property  $Q(0) = 0$  follows from skew-symmetry.



## Appendix B: Supplementary results

**Independence of the axioms** This paragraph is devoted to showing that the axioms introduced in Section 3.2 are mutually independent. By Facts 2 and 3, the universally incomparable WER  $\sigma_0$  and the universally indifferent WER  $\sigma_1$  are examples that show the independence of the consistency axiom and the prudence axiom, respectively. Also, the covering order discussed in Section 6 induces a welfare evaluation rule that satisfies all except the symmetry axiom. For the anonymity axiom, fix any distinct alternatives  $x^*, y^* \in X$ , and define a welfare evaluation rule  $\sigma$  by the criterion below. If  $\succsim \in \mathcal{P}$  is such that  $|D_{\succsim}| \leq 3$ , set  $\sigma(\succsim) = \text{core}(\succsim)$  except when  $x \succ y \succ z \succ x$  with  $\{x^*, y^*\} \subset \{x, y, z\}$  for some  $x, y, z \in X$ , in which case set  $\sigma(\succsim) = \succsim \cap \{(x^*, y^*), (y^*, x^*)\}$ . If  $\succsim \in \mathcal{P}$  is such that  $|D_{\succsim}| > 3$ , let, for any  $x, y \in X$ ,  $x \sigma(\succsim) y$  if and only if  $x \sigma(\succsim_{xyz}) y$  for all  $z \in D_{\succsim}$ . Then,  $\sigma$  meets all the axioms except the anonymity axiom. For the reduction axiom, define a welfare evaluation rule  $\sigma$  by  $\sigma(\succsim) = \text{core}(\succsim)$  for all  $\succsim \in \mathcal{P}$  except when  $\succsim = \succsim_{33}$  (the third row and the third column) in Figure 5 for some  $x, y, z, w \in X$ . In the latter case, let  $\sigma(\succsim) = \{(x, z), (z, x), (y, w), (w, y)\}$ . Then,  $\sigma$  satisfies all but the reduction axiom. Lastly, for the extension axiom, define a binary relation  $R_{\succsim}$  for each  $\succsim \in \mathcal{P}$  by  $x R_{\succsim} y$  iff  $x \succsim y$  and there is no cycle of  $\succsim$  that involves  $(x, y)$ . In turn, define a welfare evaluation rule  $\sigma$  by assigning the transitive closure of  $R_{\succsim}$  to every  $\succsim \in \mathcal{P}$ . Then, we can show that  $\sigma$  satisfies all except the extension axiom. (For an example where  $\sigma$  violates the extension axiom, consider a preference relation  $\succsim_2$  in Figure 1. It follows that  $R_{\succsim_2} = \sigma(\succsim_2) = \{(y, z), (w, z)\}$ , while the extension axiom requires that the indifference  $(y, w)$  be preserved in  $\sigma(\succsim_2)$ .)

## References

- [1] F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963.
- [2] W. E. Armstrong. The determinateness of the utility function. *Economic Journal*, 49(195):453–467, September 1939.
- [3] W. E. Armstrong. Uncertainty and the utility function. *Economic Journal*, 58(229):1–10, March 1948.

- [4] W. E. Armstrong. Utility and the theory of welfare. *Oxford Economic Papers, New Series*, 3(3):259–271, October 1951.
- [5] Kenneth J. Arrow. *Social Choice and Individual Values*. Wiley, New York, 1951.
- [6] B. Douglas Bernheim and Antonio Rangel. Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review*, 97(2):464–470, May 2007.
- [7] B. Douglas Bernheim and Antonio Rangel. Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, 124(1):51–104, February 2009.
- [8] Truman F. Bewley. Knightian decision theory. part i. *Decisions in Economics and Finance*, 25(2):79–110, November 2002.
- [9] Sushil Bikhchandani and Uzi Segal. Transitive regret. *Theoretical Economics*, 6:95–108, 2011.
- [10] Andrew Caplin, Mark Dean, and Daniel Martin. Search and satisficing. *American Economic Review*, 101(7):2899–2922, December 2011.
- [11] Christopher P. Chambers and Takashi Hayashi. Choice and individual welfare. *Journal of Economic Theory*, 147(5):1818–1849, 2012.
- [12] Peter C. Fishburn. Intransitive indifference with unequal indifference intervals. *Journal of Mathematical Psychology*, 7(1):144–149, February 1970.
- [13] Peter C. Fishburn. Condorcet social functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.
- [14] Tsogbadral Galaabaatar and Edi Karni. Objective and subjective expected utility with incomplete preferences. Working paper, 2010.
- [15] Takashi Hayashi. Regret aversion and opportunity dependence. *Journal of Economic Theory*, 139(1):242–268, March 2008.
- [16] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, March 1979.

- [17] Botond Kőszegi and Matthew Rabin. Mistakes in choice-based welfare analysis. *American Economic Review*, 97(2):477–481, 2007.
- [18] Ehud Lehrer and Roei Teper. Justifiable preferences. *Journal of Economic Theory*, 146(2):762–774, 2011.
- [19] Graham Loomes and Robert Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, 92(368):805–824, December 1982.
- [20] R. Duncan Luce. Semiorders and a theory of utility discrimination. *Econometrica*, 24(2):178–191, April 1956.
- [21] Paola Manzini and Marco Mariotti. Sequentially rationalizable choice. *American Economic Review*, 97(5):1824–1839, December 2007.
- [22] Paola Manzini and Marco Mariotti. Choice by lexicographic semiorders. *Theoretical Economics*, 7(1):1–23, January 2012.
- [23] Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y. Ozbay. Revealed attention. *American Economic Review*, 102(5):2183–2205, 2012.
- [24] David C. McGarvey. A theorem on the construction of voting paradoxes. *Econometrica*, 21(4):608–610, October 1953.
- [25] Nicholas R. Miller. A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, 24(1):68–96, February 1980.
- [26] Efe A. Ok and Yusufcan Masatlioglu. A theory of (relative) discounting. *Journal of Economic Theory*, 137(1):214–245, November 2007.
- [27] Daniel Read. Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty*, 23(1):5–32, 2001.
- [28] Peter H. M. P. Roelofsma and Daniel Read. Intransitive intertemporal choice. *Journal of Behavioral Decision Making*, 13:161–177, 2000.
- [29] Ariel Rubinstein. Ranking the participants in a tournament. *SIAM Journal on Applied Mathematics*, 38(1):108–111, February 1980.

- [30] Ariel Rubinstein. "Economics and psychology"? the case of hyperbolic discounting. *International Economic Review*, 44(4):1207–1216, November 2003.
- [31] Ariel Rubinstein and Yuval Salant. A model of choice from lists. *Theoretical Economics*, 1:3–17, 2006.
- [32] Ariel Rubinstein and Yuval Salant. Eliciting welfare preferences from behavioural data sets. *Review of Economic Studies*, 79:375–387, 2012.
- [33] Yuval Salant. Procedural analysis of choice rules with applications to bounded rationality. *American Economic Review*, 101(2):724–748, April 2011.
- [34] Yuval Salant and Ariel Rubinstein.  $(A, f)$ : Choice with frames. *Review of Economic Studies*, 75(4):1287–1296, October 2008.
- [35] John H. Smith. Aggregation of preferences with variable electorate. *Econometrica*, 41(6):1027–1041, November 1973.
- [36] Christopher J. Tyson. Cognitive constraints, contraction consistency, and the satisficing criterion. *Journal of Economic Theory*, 138(1):51–70, January 2008.
- [37] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 1944.