The Swaps Index for Consumer Choice

Mia Lu and Nick Netzer*

January 2022

Abstract

We extend the swaps index of rationality, introduced by Apesteguia and Ballester (2015) for a finite set of alternatives, to the standard consumer choice setting with infinite commodity spaces. Applications include consumer demand from competitive budget sets and the state-space approach to choice under uncertainty. We are primarily interested in Apesteguia and Ballester's result that the swaps index recovers the decision-maker's true preference from choice data for a large class of boundedly rational behavioral models. We show that this result still holds in the consumer choice setting under a suitably defined monotonicity condition. This condition is satisfied for various models of interest but violated for others.

Keywords: measures of rationality, revealed preference, behavioral welfare economics

JEL Classification: D01, D11, D60, D90

*Department of Economics, University of Zurich. Email: mia.lu@econ.uzh.ch, nick.netzer@econ.uzh.ch.
We are grateful for very helpful comments to Lasse Mononen, Tim Netzer, Jakub Steiner, and seminar participants at the University of Zurich. This research project and related results were made possible with the support of the Alfred P. Sloan Foundation and the NOMIS Foundation.

1 Introduction

Measuring the degree of rationality of choices has a long tradition in economics. The early literature on revealed preference theory (Samuelson, 1938; Houthakker, 1950; Arrow, 1959) provided conditions under which choices can or cannot be classified as perfectly rational. Subsequent authors have acknowledged that human decision-makers often deviate from perfect rationality, if only by mistake or by lapse of attention, and have proposed measures that quantify these deviations. The most prominent examples of rationality measures are Afriat's efficiency index (Afriat, 1973) and Varian's goodness-of-fit index (Varian, 1990), which measure how much a decision-maker's monetary budget would have to be reduced to rule out contradictory choices.[1] These measures have been applied in various empirical studies and have facilitated important insights, for example on the development of rationality during childhood (Harbaugh et al., 2001), the correlation between rationality and wealth (Choi et al., 2014), and the causal impact of education on rationality (Kim et al., 2018).

Recently, Apesteguia and Ballester (2015, henceforth AB) have proposed the *swaps index* of rationality, which is based on counting the minimal number of choices that would have to be swapped to reconcile behavior with the maximization of a preference. This approach not only quantifies deviations from rationality but additionally delivers a revealed preference, like the original revealed preference literature mentioned above. The *swaps preference* is the one closest to the observed behavior in terms of the number of required swaps. AB have shown that this approach recovers the decision-maker's true preference from choice data for a large class of boundedly rational behavioral models. The swaps preference can thus be used as a welfare measure, connecting the literature on measuring rationality with the literature on behavioral welfare economics (e.g. Kőszegi and Rabin, 2007; Bernheim and Rangel, 2009; Rubinstein and Salant, 2012; Benkert and Netzer, 2018).

AB obtained their results in an abstract choice setting with a finite number of alternatives. In this paper, we extend the swaps approach to the standard consumer choice setting with infinite commodity spaces, which is the setting for which the earlier rationality measures have been defined. Applications include consumer demand from competitive budget sets and the state-space approach to choice under uncertainty. We show that the swaps approach still recovers the true preference in this setting under a suitably defined monotonicity condition. We discuss the plausibility of the monotonicity condition by showing that it is satisfied for various models of interest but violated for others.

The swaps approach rests on the definition of a *swaps distance* between observed behavior

[1] Other rationality measures have been proposed, among others, by Echenique et al. (2011), Dean and Martin (2016), and Echenique et al. (2018).

and any candidate preference. In the finite case, if alternative x was chosen from budget set B and the candidate preference is $\succeq$, then the number of swaps required to reconcile this behavior with the maximization of $\succeq$ is simply the number of alternatives in B which are strictly preferred by $\succeq$ over x. This is the number of alternatives which a decision-maker with preference $\succeq$ must have "overlooked" when choosing x. The swaps distance between a stochastic choice function and a preference is the probability-weighted average of these numbers across all observations. A swaps preference is a strict preference minimizing the swaps distance to the stochastic choice function, and the swaps index is the minimum swaps distance.

For the infinite case, Apesteguia and Ballester (2015, Online Appendix D.3) proposed replacing the number of preferred alternatives in a budget set by their Lebesgue measure. We implement this proposal. Our approach is more general than the proposal by AB in two dimensions. First, we allow for arbitrary stochastic choice from a budget set, formalized by a probability measure on that set, not only finite randomization. Special cases are deterministic choice or finite randomization, but also continuous distributions like uniform trembles over the entire budget set or more complex singular distributions like continuous trembles on the budget frontier. Second, we allow for arbitrary collections of choice situations, not only finite collections of competitive budget sets. This allows us to embed conventional, deterministic Walrasian demand functions for all prices as a special case, and additionally is flexible enough to model non-linear prices, endowment points, and even frames which affect behavior without affecting the budget set.

The central property for the recoverability of preferences from choice data is monotonicity. For the finite case, AB define a stochastic choice function to be monotone with respect to a preference $\succeq$ if $x \succeq y$ implies that the probability of choosing x is larger than the probability of choosing y, on average across all budget sets in which both x and y are available. Several behavioral models generate choice data which satisfy $\succeq$-monotonicity with respect to the decision-maker's true preference, for example a simple trembles model in which uniform choice mistakes happen with small probability. AB show that $\succeq$-monotonicity implies that $\succeq$ is a swaps preference.

For our definition of monotonicity in the consumer choice setting, we first apply a Lebesgue decomposition of the choice probability measure on each budget set into its absolutely continuous and its singular component. Then we require that $x \succeq y$ implies that the density of the absolutely continuous component is weakly larger for x than for y. We furthermore require that all bundles in the support of the singular component are optimal with respect to $\succeq$. Based on this definition, we show that $\succeq$-monotonicity implies that $\succeq$ is a swaps preference. The proof technique for the infinite case is very different from the finite case. We first approximate the absolutely continuous component of the choice data by a weighted sum of uniform distributions on the upper contour sets of $\succeq$. We then derive a lower bound on the swaps index for the absolutely continuous component and show that this lower bound is attained by $\succeq$. Using that the singular component is optimal with respect to $\succeq$ implies that $\succeq$ also explains the singular component best and hence is a swaps preference.

We then study examples of boundedly rational behavioral models. Like in the finite case, many of them generate data satisfying monotonicity. The first is a uniform trembles model in which the decision-maker chooses optimally from a budget set with probability $1 - \epsilon$ and randomizes uniformly over the entire budget set with probability ϵ. The true preference is a swaps preference whenever choice mistakes take this form. Furthermore, and again in analogy to AB, the swaps preference becomes unique when the data are sufficiently rich. To establish this result, we exploit a connection to the problem of unique recoverability of preferences from conventional, deterministic Walrasian demand functions (Mas-Colell, 1977). A second example for which monotonicity obtains is a continuous version of the multinomial logit model. A third example is a model of framing, in which choice is influenced by defaults or presentation effects more generally. As long as this affects only the absolutely continuous component of choice and averages out across frames, a weaker but still sufficient version of monotonicity is satisfied.

There are more intricate models of choice in which monotonicity is less straightforward. For example, even a trembling decision-maker may understand that consumption bundles in the interior of the budget set are dominated and may randomize over the budget frontier instead of the entire budget set. As we will show, monotonicity with respect to the true preference breaks down in that case. Swaps preferences will be systematically distorted by the need to explain the choices on the budget frontier. However, even in that case there may be a solution based on reducing the dimensionality of the commodity space, so that the singular randomization on the budget frontier becomes absolutely continuous in the lower dimension. Not too surprisingly, there are also models in which monotonicity plainly fails, like a limited attention version of the logit model (Matejka and McKay, 2015) or a random parameter model in the context of choice under uncertainty (Apesteguia and Ballester, 2018).

To summarize, the swaps approach by AB is a bridge between the literatures on measuring rationality and on behavioral welfare economics. We show that this bridge is valid in the conventional consumer choice setting in which measures of rationality are typically defined. In empirical studies like Kim et al. (2018), experimental treatments will often have effects on both rationality and preferences simultaneously. The swaps approach is able to disentangle these separate effects.

2 Theoretical Framework

Consider the standard consumer choice setting. We assume that a decision-maker has conventional preferences but sometimes chooses differently from her most preferred option, for example due to random stochastic errors or lack of attention. We observe her choice behavior for a collection of budget sets, but not her preference.

Formally, the choice set is $X = \mathbb{R}^L_+$, where L denotes the number of commodities. The commodities could be different consumption goods, or they could refer to states of the world. Budget sets $B_\delta \subseteq X$ are parameterized by $\delta \in \Delta \subseteq \mathbb{R}^K$, where K denotes the number of parameters. The budget sets are assumed to be bounded, for each $\delta \in \Delta$. In the standard setting, $\delta \in \Delta \subseteq \mathbb{R}^L_{++}$ is a price vector and $B_\delta = \{x \in X : \delta \cdot x \leq 1\}$ is the competitive budget set.[2] Our approach is more general and δ could, for example, also encode kinks in the budget line generated by endowment points in an insurance setting, other non-linear budget lines, and even frames which affect choice without affecting the budget set.

The observed behavior of the decision-maker is summarized by her choice data, here defined as a collection D of probability measures, $D = (\nu, (\mu_\delta)_{\delta \in \Delta})$. First, ν is a full support probability measure on the sample space Δ, representing how often the decision-maker faces a choice situation indexed by δ.[3] We think of this as exogenous to the decision-maker; for example, the measure ν could describe the frequency of budget problems as determined by an experimenter. Since Δ can be finite or infinite, this encompasses the case of choice from finitely many different budget sets but also allows observation of a demand function for all prices, like in the conventional Walrasian approach to consumer demand. Second, the choice behavior of the decision-maker for any choice situation $\delta \in \Delta$ is captured by a probability measure μ_δ on the sample space B_δ. That is, for a measurable subset of alternatives $A \subseteq B_\delta$, $\mu_\delta(A)$ is the probability that the decision-maker chooses a consumption bundle in the subset A conditional on facing the choice situation δ. Again, we consider general probability measures here. This includes the cases of deterministic choice, discrete randomization, and stochastic choice described by a density function.

Unlike in the finite case, indifferences cannot be assumed away in the consumer choice setting with infinite commodity spaces. Instead, we only impose two weak regularity conditions on the decision-maker's preferences. First, we assume that her (complete and transitive) preference relation $\succeq$ on X can be represented by some utility function $u : X \to \mathbb{R}$. A suf-

[2]The implicit assumption then is that the decision-maker's demand satisfies homogeneity of degree zero. Hence we do not need to distinguish between demand on $\{x : \delta \cdot x \leq w\}$ and $\{x : (\delta/w) \cdot x \leq 1\}$ and can normalize wealth to 1.

[3]Throughout, all probability measures are defined on the Borel algebra. Expressions like "for almost all," "absolutely continuous," or "singular" refer to the Lebesgue measure, if not mentioned otherwise.

ficient condition for this would be the conventional assumption of continuity of $\succeq$. Second, we assume that $\succeq$ has thin level sets, i.e., the sets $L(z) = \{x \in X : u(x) = z\}$ have Lebesgue measure zero for all $z \in \mathbb{R}$. This amounts to the conventional assumption of local non-satiation. One consequence is that we can use $\lambda(\{y \in A : y \succeq x\})$ and $\lambda(\{y \in A : y \succ x\})$ interchangeably, where λ denotes the Lebesgue measure and $\succ$ the strict relation associated with $\succeq$. Denote the set of preference relations that have thin level sets and for which utility representations exist by $\mathscr{R}$.

We can now define the swaps index in this setting. This is a natural extension of the swaps index for finite sets, and it has been proposed by Apesteguia and Ballester (2015, Online Appendix D.3). We generalize it here by allowing for infinite collections of observed choice situations and for unconstrained stochastic choice in each of those situations. The *swaps distance* between preference $\succeq$ and choice data D is defined as

$$W(\succeq, D) = \int_{\Delta} \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ x\}) d\mu_\delta(x) d\nu(\delta).$$

It measures the inconsistency of the given choice data D with the given preference $\succeq$. For any chosen bundle x in a budget set B_δ, inconsistency with $\succeq$ is measured by the (Lebesgue) volume of the upper contour set of x in B_δ with respect to $\succeq$. That is, the inconsistency is captured by how many available alternative bundles were strictly preferred to x but not chosen, and which would thus have to be "swapped" with x to make choice consistent with the maximization of preference $\succeq$. The overall inconsistency of D with $\succeq$ is then computed by first taking the weighted average of this volume over all bundles in the budget set, and then taking the weighted average over all choice situations, where the respective weights are the observed frequencies in the data.[4]

The *swaps index* for the given choice data D is defined as

$$I(D) = \inf_{\succeq \in \mathscr{R}} W(\succeq, D).$$

If a preference attains the infimum, we refer to it as a *swaps preference.* A swaps preference is a preference with the smallest possible inconsistency with the choice data. Note that, unlike in the finite case in AB, existence of a swaps preference is not always guaranteed. However, the monotonicity condition we establish below implies existence.

[4]A similar measure was proposed by Heufer (2008). The main difference is that Heufer (2008) does not use the Lebesgue measure of the upper contour set of a chosen bundle with respect to a candidate preference, but instead uses the Lebesgue measure of the bundles which are revealed to be preferred over the chosen bundle by the decision-maker's other choices (and monotonicity). This generates a measure of rationality but not a complete welfare preference.

3 Main Result

We will now show that the true preference is a swaps preference as long as the choice data satisfy a monotonicity condition with respect to the true preference. We begin by providing a general technical result.

Lemma 1. *For any bounded set $B \subseteq X$, the unweighted average volume of the upper contour sets is the same for all preferences $\succeq \in \mathscr{R}$. More specifically:*

$$\int_B \lambda(\{y \in B : y \succ x\})dx = \frac{\lambda(B)^2}{2} \quad \forall \succeq \in \mathscr{R}. \tag{1}$$

Consider bundle x in Figure 1. Obviously, the upper contour set of x within the depicted budget set is smaller for the Leontief preference (blue indifference curve) than for a preference where the commodities are not perfect complements (green indifference curve). This could lead to the conjecture that Leontief preferences generate smaller average volumes of the upper contour sets than other preferences. However, for a bundle like y in Figure 1, the ranking is reversed for the same two preferences, with the Leontief preference generating the larger upper contour set. Maybe surprisingly, on average across all consumption bundles in a budget set, the size of the upper contour sets is the *same for all* preferences.

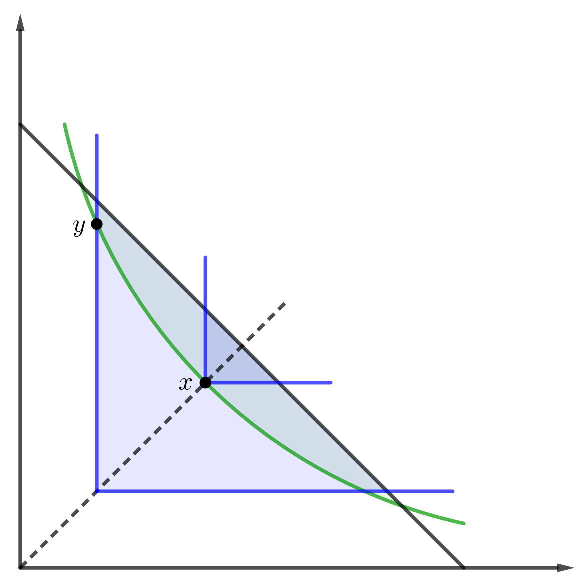

Figure 1: Volume of Upper Contour Sets.

The result is best understood by considering an arbitrary preference which partitions B into n indifference classes of equal volume $\lambda(B)/n$. While such a preference is not in $\mathscr{R}$, it can be used to approximate preferences in $\mathscr{R}$ for large n. The average size of the upper contour sets for this preference is

$$\sum_{i=1}^{n} \frac{\lambda(B)}{n} \sum_{j=i+1}^{n} \frac{\lambda(B)}{n} = \frac{\lambda(B)^2}{n^2} \sum_{i=1}^{n} (n-i) = \lambda(B)^2 \left(1 - \frac{n+1}{2n}\right).$$

Equation (1) can be seen as the limit of such a summation formula for $n \to \infty$. On a more intuitive level, the Lebesgue measure of the upper contour set of x within a budget set gives the probability (up to a normalizing constant) that a uniformly drawn bundle from that budget set is preferred over x. Averaging over all x in the budget set then gives the probability that among two bundles that are sequentially drawn uniformly and independently from the budget set, the second is preferred over the first. This probability is independent of the preference and equals 1/2.

Lemma 1 implies that if the decision-maker randomizes uniformly over all bundles in a budget set, then all preferences in $\mathscr{R}$ are equally distant to the choices in that budget set as measured by the swaps distance. Note that the statement holds for general bounded subsets of X, in particular also for subsets of budget sets.

Recall that our goal is to establish a condition which ensures that the true preference is a swaps preference. To this end, we decompose each probability measure μ_δ into two components. By Lebesgue's Decomposition Theorem, we can always decompose μ_δ into an absolutely continuous measure μ_δ^c and a singular measure μ_δ^s such that

$$\mu_\delta = \mu_\delta^c + \mu_\delta^s.$$

Moreover, this decomposition is always unique. Due to the absolute continuity, we can then write the measure μ_δ^c in terms of a density f_δ, such that

$$\mu_\delta^c(A) = \int_A f_\delta(x) dx$$

for every measurable subset $A \subseteq B_\delta$.[5] The singular measure μ_δ^s captures the possibility of deterministic choice, discrete randomization, or stochastic choice on other subsets of B_δ that have Lebesgue measure zero.

We can now introduce our monotonicity property. Informally speaking, it requires that more preferred bundles (with respect to the true preference generating the data) are chosen more often.

[5]Note that, while μ_δ is a probability measure, μ_δ^c and μ_δ^s are generally not probability measures, unless μ_δ itself is already absolutely continuous or singular. Hence f_δ is not a probability density function, because $\int_{B_\delta} f_\delta(x)dx = \mu_\delta^c(B_\delta)$ can be smaller than one. This does not have to concern us in the following.

Definition 1. *Let* $D = (\nu, (\mu_\delta)_{\delta \in \Delta})$ *be some choice data and* $\succeq_* \in \mathscr{R}$ *be some preference. We say that* D *is* $\succeq_*$*-monotone if it satisfies the following two conditions for every* $\delta \in \Delta$*:*

(*i*) *The absolutely continuous component* μ_δ^c *has a density* f_δ *that is monotone in* $\succeq_*$*:*

$$\forall x, y \in B_\delta : x \succeq_* y \Longrightarrow f_\delta(x) \geq f_\delta(y).$$

(*ii*) *The support of the singular component* μ_δ^s *is optimal with respect to* $\succeq_*$*:*

$$\forall x \in supp(\mu_\delta^s) : x \succeq_* y \;\; \forall y \in B_\delta.$$

Condition (*i*) requires that more preferred bundles according to $\succeq_*$ have higher density in the absolutely continuous component of the data, and in this sense are observed more frequently. Bundles in the support of the singular component (e.g., mass points) are observed even "more frequently," so condition (*ii*) requires them to be optimal with respect to $\succeq_*$.[6]

We will discuss the plausibility of the monotonicity property in the next subsection. Here, we only point out that $\succeq_*$-monotonicity is conceptually stronger than AB's monotonicity for the finite case, because the conditions in Definition 1 have to hold for each choice situation $\delta \in \Delta$ separately, not just on average across all choice situations. We can now state our main result.

Proposition 1. *If* D *satisfies* $\succeq_*$*-monotonicity, then* $\succeq_*$ *is a swaps preference:*

$$\succeq_* \in \arg\min_{\succeq \in \mathcal{R}} W(\succeq, D).$$

The intuition for Proposition 1 is that $\succeq_*$-monotonicity ensures that the choice data are sufficiently informative about the true preference $\succeq_*$. The true preference then minimizes the weighted inconsistency with the data and is a swaps preference. In the proof of the proposition, we first approximate the absolutely continuous component of the choice data by a weighted sum of uniform distributions on the upper contour sets of the true preference. We can then use Lemma 1 for the absolutely continuous part to derive a lower bound on $W(\succeq, D)$ and show that this lower bound is attained by $\succeq_*$. Second, using that the singular component of the choice data is optimal with respect to the true preference yields directly that $\succeq_*$ explains the singular component best.

[6] We could relax Definition 1 in two ways but refrain from doing so for ease of notation. First, we could replace condition (*ii*) by the weaker requirement that $\lambda(\{y \in B_\delta : y \succ_* x\}) = 0$ for all $x \in supp(\mu_\delta^s)$, thus allowing for non-optimal bundles in the support of μ_δ^s provided that the Lebesgue measure of better bundles is zero. This would give some additional degrees of freedom when $\succeq_*$ is not a continuous preference. Second, we could relax the requirements in Definition 1 on null sets of the product measures $\nu \times \mu_\delta$.

Monotonicity implies the existence of a swaps preference. However, the swaps preference is not necessarily unique, so the true preference cannot necessarily be recovered uniquely from the data. We will return to this issue in the next section, where we will illustrate that uniqueness can be achieved with sufficiently rich data sets.

4 Applications

To illustrate the plausibility of the monotonicity condition and to elaborate on the unique recoverability of the true preference, we now discuss several examples of behavioral models for which monotonicity is satisfied or violated.

4.1 Uniform Trembles

A first example satisfying $\succeq_*$-monotonicity is a uniform trembles model, which AB have studied for the finite case. In choice situation $\delta \in \Delta$, the decision-maker chooses optimally in budget set B_δ according to her true preference $\succeq_*$ with probability $1-\epsilon > 0$. For example, Figure 2 illustrates a case where the unique optimal bundle is x^*. With probability $\epsilon \geq 0$, she trembles uniformly over all bundles in B_δ, as indicated by the red area in Figure 2.

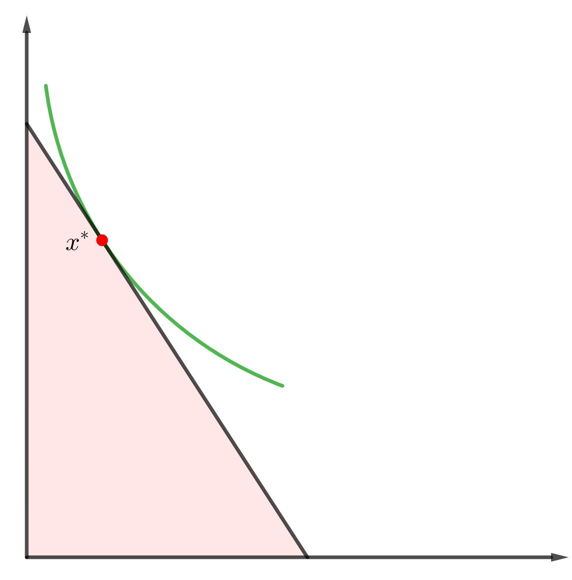

Figure 2: Uniform Trembles.

Formally, the choice data $D = (\nu, (\mu_\delta)_{\delta\in\Delta})$ are given by an exogenous distribution ν over choice situations and, for each $\delta \in \Delta$, a probability measure μ_δ that can be decomposed into an absolutely continuous component μ^c_δ with constant density $f_\delta(x) = \epsilon/\lambda(B_\delta)$ and a singular component μ_δ^s that satisfies $\mu_\delta^s(A) = 1 - \epsilon$ if $x_\delta^* \in A$ and $\mu_\delta^s(A) = 0$ if $x_\delta^* \notin A$, where x_δ^* denotes the optimal bundle in B_δ according to $\succeq_*$ (assumed to be unique for simplicity). It is easy to see that these data satisfy $\succeq_*$-monotonicity. It follows from Proposition 1 that the true preference $\succeq_*$ is a swaps preference for D.

The swaps preference will generally not be unique. For example, if only a single budget set B is observed, like in Figure 2, then any preference for which bundle x^* is optimal is a swaps preference (the same would be true in a finite setting). Observing choice from additional budget sets puts more constraints on the set of swaps preferences, because all mass points have to be explained as optimal to achieve the same minimum swaps index as the true preference $\succeq_*$. Our next result gives conditions under which the swaps preference becomes unique when the data set becomes complete.

We make use of an existing recoverability result for conventional, deterministic Walrasian demand functions by Mas-Colell (1977). We restrict attention to preferences $\succeq$ that are continuous, monotone, and strictly convex. As defined in Mas-Colell (1977, p. 1411), such a preference is *lipschitzian* if, for every $0 < r < \infty$, the function $V_{\succeq,r} : K_r \to \mathcal{C}$ given by $V_{\succeq,r}(x) = \{y \in K_r : y \succeq x\}$ is Lipschitz continuous, where $K_r = \{y \in \mathbb{R}_+^L : 1/(1+r) \leq y_i \leq 1 + r, \forall i = 1, \ldots, L\}$ and $\mathcal{C}$ is the set of non-empty, compact, and convex subsets of $\mathbb{R}_+^L$.

Proposition 2. *Let $\Delta = \mathbb{R}_{++}^L$ and $B_\delta = \{x \in X : \delta \cdot x \leq 1\}$. Let $D = (\nu, (\mu_\delta)_{\delta\in\Delta})$ be choice data generated by a uniform trembles model with a true preference $\succeq_*$ that is continuous, monotone, and strictly convex. If $\succeq_*$ is additionally lipschitzian, then it is the unique swaps preference in the class of continuous, monotone, and strictly convex preferences.*

Under the assumptions of the proposition, minimizing the swaps distance to the data yields the true preference as a unique solution. This way, the true preference can be recovered from the noisy data. To understand the logic of the result, it is instructive to consider the conventional, deterministic Walrasian demand function x^* generated by preference $\succeq_*$. Under the lipschitzian assumption, Mas-Colell (1977) has shown that any other preference $\succeq \neq \succeq_*$ generates a Walrasian demand function x that differs from x^* for at least once price vector δ_0. Continuity of Walrasian demand then implies that x^* and x must differ on an open set of prices. Furthermore, the set of bundles that $\succeq$ prefers strictly over x^* has strictly positive Lebesgue measure on that open set of prices. Since the singular part of the choice data generated by the uniform trembles model is described by x^*, its swaps distance to $\succeq_*$ is zero but strictly positive to $\succeq$. Since both $\succeq_*$ and $\succeq$ explain the absolutely continuous part of the choice data equally well, by Lemma 1, it follows that $\succeq$ is not a swaps preference.

Using Lemma 1, we can also calculate the value of the swaps index for the data generated by a uniform trembles model:

$$\begin{aligned}
I(D) &= \int_\Delta \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ_* x\}) d\mu_\delta(x) d\nu(\delta) \\
&= \int_\Delta \frac{\epsilon}{\lambda(B_\delta)} \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ_* x\}) dx d\nu(\delta) \\
&= \int_\Delta \frac{\epsilon}{\lambda(B_\delta)} \frac{\lambda(B_\delta)^2}{2} d\nu(\delta) \\
&= \epsilon \int_\Delta \frac{\lambda(B_\delta)}{2} d\nu(\delta).
\end{aligned}$$

The degree of irrationality is increasing linearly in the noise parameter ϵ. It also depends on the decision-maker's environment described by ν. If she is confronted with larger budget sets on average, then her uniform trembles over those sets imply a larger degree of irrationality.

4.2 Uniform Trembles on the Budget Frontier

Now consider a variant of the uniform trembles model. With probability $1 - \epsilon_1 - \epsilon_2 > 0$, the decision-maker chooses optimally in budget set B_δ according to $\succeq_*$. With probability $\epsilon_1 \geq 0$, she trembles uniformly over all bundles in B_δ. With probability $\epsilon_2 > 0$, she trembles uniformly over all bundles on the frontier of B_δ. For example, the decision-maker may find it easier to understand that bundles in the interior of the budget set are dominated than to determine the exact location of the optimal bundle on the budget frontier. Trembles on the frontier would then arise when the decision-maker is mildly distracted, while trembles on the entire budget set would arise if distraction is strong.

Formally, the measure μ_δ is decomposed into the absolutely continuous component μ_δ^c that captures trembles on the entire budget set and the singular component μ_δ^s that captures trembles on the budget frontier and the optimal choice. A simple case is illustrated in Figure 3, where the optimal bundle is x^* and trembles take place sometimes on the budget line and sometimes on the entire budget set (as indicated in red).

Suppose for simplicity that we only observe one budget set B, like in Figure 3. It follows immediately that the choice data are best explained by an indifference between all bundles on the budget line. With an indifference curve like the red line, all singular choices are perfectly explained. Since all preferences explain the uniform trembles on the entire budget set equally well, by Lemma 1, it follows that some perfect-substitutes preference is a swaps preference. In fact, the choice data are monotone with respect to that perfect-substitutes preference. Whenever the true preference is different, like the green indifferences curves depicted in Figure 3, its swaps distance to the data is strictly larger, because it generates

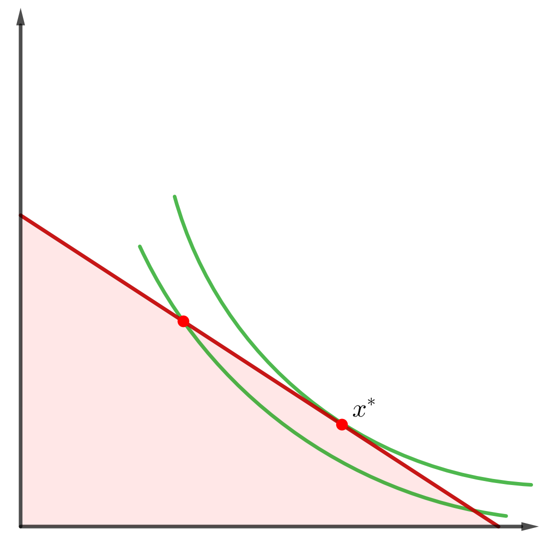

Figure 3: Uniform Trembles on Budget Frontier.

upper contour sets with strictly positive volume for all non-optimal choices on the budget line. Hence, the true preference is not a swaps preference and recoverability fails. The value of the swaps index is $I(D) = \epsilon_1 \lambda(B)/2$, missing the trembles that arise with probability ϵ_2.

Indifferences cannot be ruled out in the consumer choice setting, in contrast to the finite setting. Note that the red indifference curve in Figure 3 is also compatible with strongly monotone preferences, a restriction proposed by Apesteguia and Ballester (2015, Online Appendix D.2) to deal with indifferences. Note furthermore that the recoverability problem would not disappear when imposing the additional restriction of strictly convex preferences, such that an indifference among all bundles on the budget line is no longer possible. Preferences with indifference curves closer to the red line still perform better than the true preference. The problem is also not due to the assumption of observing only one budget set. Figuring out the swaps preference is more difficult when the environment is richer, but the previous arguments imply that preference elicitation will be systematically distorted by the need to explain the singular choices on the budget frontiers.

An interesting special case arises when $\epsilon_1 = 0$, so that trembles take place entirely on the budget frontier. We can now attempt to solve the problem by reducing the dimensionality of the commodity space. Consider again an example with two commodities, $x = (x_1, x_2)$, and a single budget set, as in Figure 3. Write the decision-maker's problem as a choice of $x_1 \in B = [0, x_1^{max}]$ and determine x_2 as a residual on the budget line. The data are then given by the choice of x_1^* with probability $1 - \epsilon_2$ (singular component) and a uniform distribution on $[0, x_1^{max}]$ with probability ϵ_2 (now the absolutely continuous component). We can look for a swaps preference on $B = [0, x_1^{max}]$. Note that $\mathscr{R}$ now excludes the complete indifference, which would not have thin level sets. Since Lemma 1 holds regardless of the number of dimensions, and hence also for $L = 1$, we obtain that all preferences in $\mathscr{R}$ explain the uniform randomization equally well. The same arguments as in Subsection 4.1 then imply that any preference for which x_1^* is optimal in B is a swaps preference, generating a swaps index of $I(D) = \epsilon_2 x_1^{max}/2$. In particular, the decision-maker's true preference on B induced by her true preference $\succeq_*$ on $\mathbb{R}_+^2$ satisfies that condition. Hence, reducing the dimensionality of the commodity space solves the problem when the observed choice is entirely singular.

We conclude this subsection by showing that there are also behavioral models for which recoverability fails even after a reduction of the dimensionality of the commodity space. The following example is a random parameter model in the context of choice under uncertainty.

Example 1. Let $x = (x_1, x_2) \in X = \mathbb{R}_+^2$ denote consumption in two states of the world. The decision-maker is endowed with bundle $(w, w - d) \in X$, reflecting that her wealth is w if no accident occurs (state 1) but is reduced by $d > 0$ if an accident occurs (state 2). The probability of an accident is $0 < p < 1$. The decision-maker can buy insurance. Each unit of insurance costs q, to be paid in both states of the world, and pays 1 in case of an accident. We assume $q > p$ so that the insurance company earns profits. The resulting budget set B is illustrated in Figure 4. The 45-degree line indicates perfect insurance bundles.

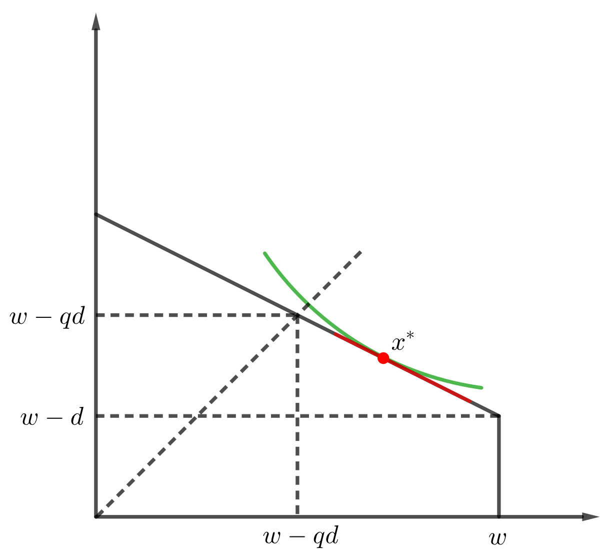

Figure 4: Random Parameter Model.

The decision-maker is an expected utility maximizer with CARA utility $u(x_i) = -e^{-\beta x_i}$

where $\beta > 0$. It is an easy exercise to show that the optimal bundle is

$$(x_1, x_2) = \left(w - qd + \frac{q\delta}{\beta}, w - qd - \frac{(1-q)\delta}{\beta} \right),$$

where

$$\delta = \ln\left(\frac{1-p}{p}\right) - \ln\left(\frac{1-q}{q}\right) > 0,$$

if we assume $\beta > \delta/d$ to rule out corner solutions in the endowment point.

Suppose the true risk aversion coefficient is β_* but choice is governed by a uniform distribution of β on $[\beta_* - \epsilon/2, \beta_* + \epsilon/2]$ for some $\epsilon > 0$, i.e., trembles arise due to a noisy preference parameter. This model generates singular stochastic choice on the budget line, as indicated by the red line segment in Figure 4. Hence, by our previous arguments, the swaps approach applied in $\mathbb{R}^2_+$ does not recover the true preference correctly but predicts an indifference among all chosen bundles.[7]

Reducing the dimensionality of the commodity space does not solve that problem. The decision-maker can be conceptualized as choosing x_1 from $B = [x_1^{min}, x_1^{max}]$, where

$$x_1^{min} = w - qd + \frac{q\delta}{\beta_* + \epsilon/2} \quad \text{and} \quad x_1^{max} = w - qd + \frac{q\delta}{\beta_* - \epsilon/2}.$$

Due to non-linearity in the risk aversion coefficient, choice is not uniform on B. It is described by the density

$$f(x_1) = \frac{q\delta}{\epsilon\,(x_1 - w + qd)^2},$$

which is strictly decreasing in x_1. These data are monotone with respect to preferences that always favor smaller values of x_1 in $B = [x_1^{min}, x_1^{max}]$, i.e., higher levels of insurance. The decision-maker's true preference on B, induced by risk aversion coefficient β_*, is unimodal with interior peak at $x_1^* = w - qd + q\delta/\beta_*$ and therefore does not satisfy that condition. Within the parametric class of CARA utility functions, swaps preferences are the preferences with $\beta \geq \beta_* + \epsilon/2$. The swaps approach therefore overestimates risk aversion systematically.

[7]This is compatible with subjective expected utility theory, predicting that the decision-maker is risk neutral and entertains a subjective accident probability equal to q.

4.3 Logit Models

In a conventional multinomial logit model (or Luce model) for a finite set of alternatives, the probability of choosing an alternative x from a set B is given by

$$p(x, B) = \frac{e^{u(x)/\gamma}}{\sum_{y \in B} e^{u(y)/\gamma}}, \tag{2}$$

where $\gamma > 0$ is a noise parameter. As $\gamma \to 0$, choice concentrates on the utility maximizing alternatives in B. As $\gamma \to \infty$, choice converges to a uniform randomization. The model can be given a justification as a random utility model with Gumbel error terms (see e.g. Anderson et al., 1992).[8]

We can translate this idea to our consumer choice setting by postulating that each μ_δ is absolutely continuous and the decision-maker chooses from B_δ in choice situation $\delta \in \Delta$ according to density

$$f_\delta(x) = \frac{e^{u(x)/\gamma}}{\int_{B_\delta} e^{u(y)/\gamma} dy}.$$

It is immediate that the resulting data D are monotone with respect to the preference represented by the utility function u, which is hence a swaps preference.

There are modifications of the logit model for which monotonicity fails. Matejka and McKay (2015) study a rational inattention model in which the decision-maker has to learn about her true utility function and pays an entropy-based cost of information. Consider their setting with a finite set of alternatives and assume that utility function u_i arises with prior probability q_i, where $i = 1, \ldots, m$. Matejka and McKay (2015) show that the rationally inattentive decision-maker chooses stochastically according to a modified multinomial logit formula. When the true utility function is u_i, her conditional choice probabilities are

$$p(x, B|u_i) = \frac{p(x, B) e^{u_i(x)/\gamma}}{\sum_{y \in B} p(y, B) e^{u_i(y)/\gamma}}, \tag{3}$$

where

$$p(y, B) = \sum_{i=1}^{m} q_i p(y, B|u_i)$$

[8]See Alós-Ferrer et al. (2021) for a critique of the distributional assumptions made in conventional random utility models like the multinomial logit, and for a constructive way how to get rid of these assumptions by using response time data.

are the unconditional choice probabilities. When all alternatives are ex-ante homogeneous, this simultaneous system of equations yields the conventional logit formula (2) from above for the conditional choice probabilities. Otherwise, conditional choice probabilities are distorted towards unconditional choice probabilities. We show in the Appendix that the conditional choice probabilities are not always monotone with respect to the preference $\succeq_i$ represented by u_i; if $p(x, B)$ is larger than $p(y, B)$, we can have that $p(x, B|u_i)$ is larger than $p(y, B|u_i)$ even though $u_i(x) < u_i(y)$. We also show that the unconditional choice probabilities are not always monotone with respect to the prior expected preference; we can have $p(x, B) > p(y, B)$ even though $\sum_{i=1}^{m} q_i u_i(x) < \sum_{i=1}^{m} q_i u_i(y)$. It follows that there is not necessarily a plausible relation between swaps preferences and true preferences for data generated by the rational inattention model of Matejka and McKay (2015).

4.4 Framing

Consider finally a situation in which choice can be affected by frames (Salant and Rubinstein, 2008). For example, one of the bundles in a budget set could be marked as the default choice, and different defaults could lead to different choice behaviors. A choice situation could then be described by $\delta = (p, d) \in \Delta$, where $p \in \mathbb{R}^L_{++}$ are the commodity prices and $d \in \{x \in X : p \cdot x \leq 1\}$ is the default bundle. More generally, we can always combine budget sets with different frames that describe the way in which the budget set is presented to the decision-maker, or that encode other relevant features of the environment such as the level of distraction for the decision-maker.

Formally, given an arbitrary set of choice situations Δ, we can define an equivalence relation $\simeq$ on Δ by letting $\delta \simeq \delta'$ if and only if $B_\delta = B_{\delta'}$. In words, $\delta \simeq \delta'$ means that the two choice situations δ and δ' involve the same budget set and hence can differ only in their framing. In our introductory example, we would have $\delta = (p, d) \simeq (p', d') = \delta'$ if and only if $p = p'$. Denote the equivalence class of δ by $[\delta]$ and the quotient set of all equivalence classes by $\Delta/\simeq$.

We could proceed like before and ignore the specific structure of frames. There are applications in which this comes without loss. For example, if we have a uniform trembles model like in Subsection 4.1 and different frames only affect the level of noise ϵ, because they encode different levels of distraction, then the data still satisfy $\succeq_*$-monotonicity for the true preference $\succeq_*$, and the true preference can be recovered from the data just like before. In other applications, like the one with defaults, frames may have a systematic directional effect on choices and destroy $\succeq_*$-monotonicity. We will now show that monotonicity can be weakened to allow for such systematic effects, as long as they arise only in the absolutely continuous part of the data and average out across frames.

We say that $D = (\nu, (\mu_\delta)_{\delta\in\Delta})$ is *weakly* $\succeq_*$*-monotone* if it satisfies Definition 1 but with property (i) replaced by the weaker requirement that for every equivalence class $[\delta] \in \Delta/\simeq$:

$$(i)' \ \forall x, y \in B_\delta : x \succeq_* y \Longrightarrow \int_{[\delta]} f_{\delta'}(x) d\nu(\delta') \geq \int_{[\delta]} f_{\delta'}(y) d\nu(\delta').$$

It is no longer required that the density is larger for x than for y in every choice situation, only on average across all choice situations that share the same budget set (in which both x and y are available). While $\succeq_*$-monotonicity clearly implies weak $\succeq_*$-monotonicity, the weaker property can be satisfied even if some frames induce an opposite ranking of the densities. The following result is therefore a generalization of Proposition 1.

Proposition 3. *If D satisfies weak $\succeq_*$-monotonicity, then $\succeq_*$ is a swaps preference.*

This result may be useful for the analysis of nudging (Thaler and Sunstein, 2008). In the spirit of the behavioral welfare approach by Benkert and Netzer (2018), one could first check weak monotonicity of frame-dependent stochastic choice data and then evaluate the quality of choices under different frames using as a welfare criterion the respective preference for which monotonicity obtains, with the goal of ranking frames according to their induced decision quality. We leave such an approach to future research.

5 Conclusions

We have studied the swaps index by Apesteguia and Ballester (2015) in a consumer choice setting with infinite commodity spaces. We have shown that the swaps approach has a solid foundation in that setting, because many (but not all) plausible boundedly rational models of behavior generate choice data for which the true preference of the decision-maker is a swaps preference.

Our "data sets" are assumed to be ideal, in the sense that all probability measures are perfectly observed. We share this admittedly heroic assumption with the textbook analysis of Walrasian demand or the theory of stochastic choice which assumes that exact probabilities are observable. Real-world data is necessarily finite. A pragmatic approach in that case would be to specify a parametric class of preferences and minimize the swaps distance to the data on that class.

References

Afriat, S. (1973). On a system of inequalities in demand analysis: An extension of the classical method. *International Economic Review*, 14:460–472.

Alós-Ferrer, C., Fehr, E., and Netzer, N. (2021). Time will tell: Recovering preferences when choices are noisy. *Journal of Political Economy*, 129:1828–1877.

Anderson, S., de Palma, A., and Thisse, J. (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge: MIT Press.

Apesteguia, J. and Ballester, M. (2015). A measure of rationality and welfare. *Journal of Political Economy*, 123:1278–1310.

Apesteguia, J. and Ballester, M. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126:74–106.

Arrow, K. (1959). Rational choice functions and orderings. *Economica*, 26:121–127.

Benkert, J.-M. and Netzer, N. (2018). Informational requirements of nudging. *Journal of Political Economy*, 126:2323–2355.

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

Choi, S., Kariv, S., Müller, W., and Silverman, D. (2014). Who is (more) rational? *American Economic Review*, 104:1518–1550.

Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. *Review of Economics and Statistics*, 98:524–534.

Echenique, F., Imai, T., and Saito, K. (2018). Approximate expected utility rationalization. Mimeo.

Echenique, F., Lee, S., and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119:1201–1223.

Harbaugh, W., Krause, K., and Berry, T. (2001). Garp for kids: On the development of rational choice behavior. *American Economic Review*, 91:1539–1545.

Heufer, J. (2008). A geometric measure for the violation of utility maximization. Ruhr Economic Papers 69.

Houthakker, H. (1950). Revealed preference and the utility function. *Economica*, 17:159–174.

Kőszegi, B. and Rabin, M. (2007). Mistakes in choice-based welfare analysis. *American Economic Review, Papers and Proceedings*, 97:477–481.

Kim, H., Choi, S., Kim, B., and Pop-Eleches, C. (2018). The role of education interventions in improving economic rationality. *Science*, 362:83–86.

Mas-Colell, A. (1977). The recoverability of consumers' preferences from market demand behavior. *Econometrica*, 45:1409–1430.

Matejka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105:272–298.

Rubinstein, A. and Salant, Y. (2012). Eliciting welfare preferences from behavioural data sets. *Review of Economic Studies*, 79:375–387.

Salant, Y. and Rubinstein, A. (2008). (A,f): Choice with frames. *Review of Economic Studies*, 75:1287–1296.

Samuelson, P. (1938). A note on the pure theory of consumer's behavior. *Economica*, 56:61–71.

Thaler, R. and Sunstein, C. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven: Yale University Press.

Varian, H. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics*, 43:125–140.

A Appendix

A.1 Proof of Lemma 1

Consider a bounded set $B \subset X$ and a preference relation $\succeq \in \mathscr{R}$. Represent $\succeq$ on B by a utility function $u : B \to [0, 1]$. We denote by

$$U(x) = \{y \in B \mid u(y) \geq u(x)\}$$

the upper contour set of x in B, and by $V(x) = \lambda(U(x))$ the Lebesgue measure of the upper contour set.[9] We are interested in the average measure of these sets,

$$\bar{V} = \int_B V(x)\, dx.$$

We approximate function u from below by a sequence of step functions $(u_n)_{n\in\mathbb{N}}$ given by

$$u_n(x) = \frac{\lfloor nu(x) \rfloor}{n}.$$

Since $0 \leq u(x) - u_n(x) \leq 1/n$ for all $x \in B$, it holds that $u_n \to u$ uniformly as $n \to \infty$. Let

$$U_n(x) = \{y \in B \mid u_n(y) \geq u_n(x)\}$$

be the upper contour set based on u_n. Observe that $U(x) \subseteq U_n(x)$ for all $x \in B$ and $n \in \mathbb{N}$. Observe also that, if $y \notin U(x)$, there exists $N \in \mathbb{N}$ such that $y \notin U_n(x)$ for all $n > N$, because $u_n(y) < u_n(x)$ must hold for sufficiently large n. Hence we obtain the set-theoretic limit $U_n(x) \to U(x)$. For $V_n(x) = \lambda(U_n(x))$, continuity of the Lebesgue measure then implies $\lim_{n\to\infty} V_n(x) = V(x)$ for all $x \in B$, i.e., pointwise convergence of V_n to V.

Let us examine

$$\bar{V}_n = \int_B V_n(x)\, dx.$$

For each $k \in \{0, 1, \ldots, n\}$, define

$$B_n^k = \{x \in B \mid u_n(x) = k/n\},$$

so that $(B_n^k)_k$ forms a partition of B. For any $x \in B_n^k$ we obtain $U_n(x) = \bigcup_{j=k}^n B_n^j$ and $V_n(x) = \sum_{j=k}^n \lambda(B_n^j)$. By definition of the Lebesgue integral we can then write

$$\begin{aligned}
\bar{V}_n &= \sum_{k=0}^{n} \lambda(B_n^k) \sum_{j=k}^{n} \lambda(B_n^j) \\
&= \sum_{k=0}^{n} \lambda(B_n^k) \left[\lambda(B) - \sum_{j=0}^{k-1} \lambda(B_n^j) \right] \\
&= \lambda(B)^2 - \sum_{k=0}^{n} \sum_{j=0}^{k-1} \lambda(B_n^k) \lambda(B_n^j).
\end{aligned}$$

[9] Note again that it does not matter for $V(x)$ whether we use $u(y) \geq u(x)$ or $u(y) > u(x)$ in the definition of $U(x)$, because the level sets of u have Lebesgue measure zero.

To derive a lower bound on this expression, consider the auxiliary optimization problem

$$\max_{a_0,\dots,a_n\in\mathbb{R}_+} \sum_{k=0}^{n}\sum_{j=0}^{k-1} a_k a_j \quad \text{subject to} \quad \sum_{k=0}^{n} a_k = \lambda(B).$$

The first-order condition of the Lagrangian

$$\mathcal{L} = \sum_{k=0}^{n}\sum_{j=0}^{k-1} a_k a_j + \gamma\left(\lambda(B) - \sum_{k=0}^{n} a_k\right)$$

with respect to a_i is $\sum_{j=0}^{i-1} a_j + \sum_{k=i+1}^{n} a_k = \gamma$, or $\sum_{k\neq i} a_k = \gamma$, where γ is the multiplier of the constraint. Since this has to hold for all i, we obtain that $a_i = \lambda(B)/(n+1)$ in the optimum. (Verifying a second-order condition yields that this is indeed the unique local constrained maximum in the interior. By computing the maxima on the boundary of the constraint set, it can be shown that this is also the global constrained maximum.) Hence we know that

$$\begin{aligned}
\bar{V}_n &\geq \lambda(B)^2 - \sum_{k=0}^{n}\sum_{j=0}^{k-1}\left(\frac{\lambda(B)}{n+1}\right)^2 \\
&= \lambda(B)^2\left[1 - \sum_{k=0}^{n}\sum_{j=0}^{k-1}\frac{1}{(n+1)^2}\right] \\
&= \lambda(B)^2\left[1 - \sum_{k=0}^{n}\frac{k}{(n+1)^2}\right] \\
&= \lambda(B)^2\left[1 - \frac{n(n+1)}{2(n+1)^2}\right] \\
&= \lambda(B)^2\left[1 - \frac{1}{2}\left(\frac{n}{n+1}\right)\right].
\end{aligned}$$

To derive an upper bound, note that

$$\bar{V}_n = \lambda(B)^2 - \sum_{k=0}^{n}\sum_{j=0}^{k}\lambda(B_n^k)\lambda(B_n^j) + \sum_{k=0}^{n}\lambda(B_n^k)^2.$$

Let us assume that the regularity condition

$$\lim_{n\to\infty}\sum_{k=0}^{n}\lambda(B_n^k)^2 = 0 \tag{4}$$

holds. We will show below that this condition holds because the level sets of u have Lebesgue measure zero. For the limiting behavior of $\bar{V}_n$ we can then equivalently study the limiting behavior of

$$\tilde{V}_n = \lambda(B)^2 - \sum_{k=0}^{n}\sum_{j=0}^{k} \lambda(B_n^k)\lambda(B_n^j).$$

Analogously to above, the auxiliary problem

$$\min_{a_0,\ldots,a_n\in\mathbb{R}_+} \sum_{k=0}^{n}\sum_{j=0}^{k} a_k a_j \quad \text{subject to} \quad \sum_{k=0}^{n} a_k = \lambda(B)$$

has the solution $a_i = \lambda(B)/(n+1)$ for all i. (For the second-order condition, note that the Lagrangian is convex in $(a_o,\ldots,a_n)$.) Hence we obtain

$$\begin{aligned}\tilde{V}_n &\leq \lambda(B)^2 - \sum_{k=0}^{n}\sum_{j=0}^{k} \left(\frac{\lambda(B)}{n+1}\right)^2 \\ &= \lambda(B)^2 \left[1 - \frac{1}{2}\left(\frac{n}{n+1}\right) - \frac{1}{n+1}\right].\end{aligned}$$

Putting the upper and the lower bound together, we obtain that $\lim_{n\to\infty} \bar{V}_n = \lambda(B)^2/2$. We can now apply the Dominated Convergence Theorem to obtain

$$\bar{V} = \int_B V(x)dx = \int_B \lim_{n\to\infty} V_n(x)\,dx = \lim_{n\to\infty}\int_B V_n(x)dx = \lim_{n\to\infty} \bar{V}_n = \frac{\lambda(B)^2}{2}.$$

Note that this depends on u only to the extent that regularity condition (4) is satisfied. We now show that this condition holds because the level sets $L(z) = \{x \in B \mid u(x) = z\}$ have Lebesgue measure zero, for all $z \in [0,1]$. Letting $h(n) \in \arg\max_{k\in\{0,\ldots,n\}} \lambda(B_n^k)$, we obtain

$$\sum_{k=0}^{n} \lambda(B_n^k)^2 = \sum_{k=0}^{n} \lambda(B_n^k)\lambda(B_n^k) \leq \sum_{k=0}^{n} \lambda(B_n^k)\lambda(B_n^{h(n)}) = \lambda(B)\lambda(B_n^{h(n)}).$$

Therefore, if (4) does not hold, there exists $\epsilon > 0$ such that for each $N \in \mathbb{N}$ there exists $n > N$ for which $\lambda(B_n^{h(n)}) \geq \epsilon$. Given the sequence $(B_n^{h(n)})_n$, consider a subsequence $(B_{n_m}^{h(n_m)})_m$ with the properties that $\lambda(B_{n_m}^{h(n_m)}) \geq \epsilon$ for all m and $\lim_{m\to\infty} h(n_m)/n_m = z$ for some $z \in [0,1]$; such a subsequence exists because $h(n)/n \in [0,1]$ for all n and thus a converging subsequence exists.

For each m we have, by definition,

$$\begin{aligned} B_{n_m}^{h(n_m)} &= \{x \in B \mid u_{n_m}(x) = h(n_m)/n_m\} \\ &= \{x \in B \mid h(n_m)/n_m \leq u(x) < h(n_m)/n_m + 1/n_m\}. \end{aligned}$$

Partition the set $B_{n_m}^{h(n_m)}$ into three subsets S_m, E_m and G_m, so that each $x \in S_m$ satisfies $u(x) < z$, each $x \in E_m$ satisfies $u(x) = z$, and each $x \in G_m$ satisfies $u(x) > z$. It follows that $\lambda(B_{n_m}^{h(n_m)}) = \lambda(S_m) + \lambda(E_m) + \lambda(G_m)$. Since $\lim_{m\to\infty} h(n_m)/n_m = z$, we obtain the set-theoretic limits $S_m \to \varnothing$ and $G_m \to \varnothing$, implying $\lim_{m\to\infty} \lambda(S_m) = \lim_{m\to\infty} \lambda(G_m) = 0$ by continuity of the Lebesgue measure. Now fix any ϵ' with $0 < \epsilon' < \epsilon$. It follows that there exists $M \in \mathbb{N}$ such that for all $m > M$ we have $\lambda(E_m) \geq \epsilon'$. From this we can conclude that $h(n_m)/n_m \leq z < h(n_m)/n_m + 1/n_m$ must be satisfied for each $m > M$, which in turn implies $E_m = L(z)$ for all those m. It follows that $\lambda(L(z)) \geq \epsilon' > 0$, so the level set $L(z)$ has a strictly positive Lebesgue measure. □

A.2 Proof of Proposition 1

Note that it is sufficient to show the claim for the case that the choice data D consist of only one budget set (i.e., Δ is a singleton). Since the conditions in the monotonicity definition hold for every choice situation individually, the proof then immediately extends to the general case.

For ease of notation, we omit the subscripts from now on. Let $B \subseteq X$ be the budget set. Let μ describe the choice on B, decomposed as $\mu = \mu^c + \mu^s$, with density f of the absolutely continuous component. Suppose D is $\succeq_*$-monotone. Let $\succeq$ be an arbitrary preference. The swaps distance between $\succeq$ and D can then be decomposed as

$$W(\succeq, D) = \int_B \lambda(\{y \in B : y \succ x\}) d\mu^s(x) + \int_B \lambda(\{y \in B : y \succ x\}) f(x) dx. \tag{5}$$

For the first summand in (5) we have

$$\int_B \lambda(\{y \in B : y \succ_* x\}) d\mu^s(x) = 0 \leq \int_B \lambda(\{y \in B : y \succ x\}) d\mu^s(x),$$

as the singular component μ^s is rationalized by $\succeq_*$ per definition of $\succeq_*$-monotonicity.

We will now show that for the second summand in (5) it holds that

$$\succeq_* \in \arg\min_{\succeq \in \mathcal{R}} \int_B \lambda(\{y \in B : y \succ x\}) f(x) dx.$$

The statement of the proposition then immediately follows.

Approximate the density f by a sequence of simple functions given by

$$f_n(x) = \frac{\lfloor nf(x) \rfloor}{n},$$

so that $0 \leq f(x) - f_n(x) \leq 1/n$ for all $x \in B$ and $f_n \to f$ uniformly as $n \to \infty$. As f_n is a simple function, we can rewrite it as

$$f_n(x) = \sum_{i=1}^{N_n} f_i^{(n)} \mathbb{1}_{B_i^{(n)}}(x)$$

for some $f_i^{(n)} \geq 0$ and $B_i^{(n)} \subseteq B$, $i = 1, \ldots, N_n$, chosen such that $f_1^{(n)} > \ldots > f_{N_n}^{(n)}$ and $(B_i^{(n)})_i$ forms a partition of B. Note that $\succeq_*$-monotonicity and monotonicity of the floor function imply that, for all $x, y \in B$ and any $n \in \mathbb{N}$,

$$x \succeq_* y \Longrightarrow f(x) \geq f(y) \Longrightarrow f_n(x) \geq f_n(y). \tag{6}$$

We now show an auxiliary result.

Lemma 2. *For all $n \in \mathbb{N}$, $\succeq_*$ is a solution of*

$$\min_{\succeq \in \mathscr{R}} \int_B \lambda(\{y \in B : y \succ x\}) f_n(x) dx.$$

Proof (Lemma). Let $\succeq$ be some preference relation. We have

$$\begin{aligned}
&\int_B \lambda(\{y \in B : y \succ x\}) f_n(x) dx \\
&= \int_B \lambda(\{y \in B : y \succ x\}) \left(\sum_{i=1}^{N_n} f_i^{(n)} \mathbb{1}_{B_i^{(n)}}(x) \right) dx \\
&= \sum_{i=1}^{N_n} f_i^{(n)} \int_{B_i^{(n)}} \lambda(\{y \in B : y \succ x\}) dx \\
&= \left(\int_{B_{N_n}^{(n)}} \ldots dx \right) f_{N_n}^{(n)} + \left(\int_{B_{N_n-1}^{(n)}} \ldots dx \right) \underbrace{(f_{N_n}^{(n)} + (f_{N_n-1}^{(n)} - f_{N_n}^{(n)}))}_{= f_{N_n-1}^{(n)}} \\
&\quad + \ldots + \left(\int_{B_1^{(n)}} \ldots dx \right) \underbrace{(f_{N_n}^{(n)} + (f_{N_n-1}^{(n)} - f_{N_n}^{(n)}) + \ldots + (f_1^{(n)} - f_2^{(n)}))}_{= f_1^{(n)}}
\end{aligned}$$

$$
\begin{aligned}
=& f_{N_n}^{(n)} \left(\int_{B_{N_n}^{(n)}} \dots dx + \dots + \int_{B_1^{(n)}} \dots dx \right) + (f_{N_n-1}^{(n)} - f_{N_n}^{(n)}) \left(\int_{B_{N_n-1}^{(n)}} \dots dx + \dots + \int_{B_1^{(n)}} \dots dx \right) \\
&+ \dots + (f_1^{(n)} - f_2^{(n)}) \left(\int_{B_1^{(n)}} \dots dx \right) \\
=& \sum_{i=1}^{N_n} c_i^{(n)} \int_{U_i^{(n)}} \lambda(\{y \in B : y \succ x\}) dx, \qquad (7)
\end{aligned}
$$

where we define $c_i^{(n)} = f_i^{(n)} - f_{i+1}^{(n)}$ for $i = 1, \dots, N_n - 1$ and $c_{N_n}^{(n)} = f_{N_n}^{(n)}$, and $U_i^{(n)} = \bigcup_{j=1}^{i} B_j^{(n)} = \{y \in B : f_n(y) \geq f_i^{(n)}\}$. Note that $c_i^{(n)} \geq 0$ by construction and

$$
\int_{U_i^{(n)}} \lambda(\{y \in B : y \succ x\}) dx \geq \int_{U_i^{(n)}} \lambda(\{y \in U_i^{(n)} : y \succ x\}) dx = \frac{\lambda(U_i^{(n)})^2}{2},
$$

where the last equality follows from Lemma 1. This lower bound is constant in $\succeq$. Moreover, it is attained by the preference $\succeq_*$,

$$
\int_{U_i^{(n)}} \lambda(\{y \in B : y \succ_* x\}) dx = \int_{U_i^{(n)}} \lambda(\{y \in U_i^{(n)} : y \succ_* x\}) dx
$$

because

$$
y \in B \setminus U_i^{(n)} \Longrightarrow f_n(y) < f_i^{(n)} \Longrightarrow x \succ_* y \quad \forall x \in U_i^{(n)},
$$

where the last implication follows from (6). Thus, $\succeq_*$ minimizes each summand of (7) and hence also the whole sum. $\diamond$

Recall that f is integrable and, for all $n \in \mathbb{N}$ and all $x \in B$, it holds that $\lim_{n\to\infty} f_n(x) = f(x)$ and $|f_n(x)| = f_n(x) \leq f(x)$. Since $\lambda(\{y \in B : y \succ x\})$ is constant in n, the Dominated Convergence Theorem implies, for any $\succeq$,

$$
\lim_{n\to\infty} \int_B \lambda(\{y \in B : y \succ x\}) f_n(x) dx = \int_A \lambda(\{y \in B : y \succ x\}) f(x) dx.
$$

Therefore, for all $\succeq$,

$$
\begin{aligned}
&\int_B \lambda(\{y \in B : y \succ_* x\}) f(x) dx \\
=& \lim_{n\to\infty} \int_B \lambda(\{y \in B : y \succ_* x\}) f_n(x) dx \\
\leq& \lim_{n\to\infty} \int_B \lambda(\{y \in B : y \succ x\}) f_n(x) dx
\end{aligned}
$$

$$= \int_B \lambda(\{y \in B : y \succ x\}) f(x) dx,$$

where the inequality follows from Lemma 2. Taking all results together we get

$$\succeq_* \in \arg\min_{\succeq \in \mathcal{R}} \int_B \lambda(\{y \in B : y \succ x\}) dx. \qquad \square$$

A.3 Proof of Proposition 2

Let $\succeq \neq \succeq_*$ be any continuous, monotone, and strictly convex preference. For each price vector $\delta \in \Delta = \mathbb{R}^L_{++}$, the optimal bundles in B_δ are unique so we can represent them by a Walrasian demand function, denoted by $x^*(\delta)$ for $\succeq_*$ and by $x(\delta)$ for $\succeq$. It follows from Berge's maximum principle that $x^*(\delta)$ and $x(\delta)$ are continuous functions.

As in the proof of Proposition 1, we can decompose the swaps distance into a singular part and an absolutely continuous part,

$$W(\succeq, D) = \underbrace{\int_\Delta \lambda(\{y \in B_\delta : y \succ x^*(\delta)\})(1-\epsilon) d\nu(\delta)}_{=:W^s(\succeq, D)} + \underbrace{\int_\Delta \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ x\}) \frac{\epsilon}{\lambda(B_\delta)} dx d\nu(\delta)}_{=:W^c(\succeq, D)},$$

where the definition of the uniform trembles model with true preference $\succeq_*$ has already been used. For uniqueness we need to show that $W(\succeq, D) > W(\succeq_*, D)$. It suffices to show that $W^s(\succeq, D) > 0$, as $W^s(\succeq_*, D) = 0$ and $W^c(\succeq, D) = W^c(\succeq_*, D)$ by definition of the uniform trembles model and Lemma 1.

Since $\succeq \neq \succeq_*$, it follows from Theorem 2 in Mas-Colell (1977) – using that $\succeq_*$ is lipschitzian – that there exists a price vector δ_0 such that $x(\delta_0) \neq x^*(\delta_0)$. Continuity of x and x^* then implies that there exists an open ball $U(\delta_0)$ around δ_0 such that $x(\delta) \neq x^*(\delta)$ and in particular $x(\delta) \succ x^*(\delta)$ for all $\delta \in U(\delta_0)$. Let u be a continuous utility representation of $\succeq$. Then $u(x(\delta)) > u(x^*(\delta))$ for all $\delta \in U(\delta_0)$.

It follows that $u^{-1}((u(x^*(\delta)), u(x(\delta)))) \subset \mathbb{R}^L_+$ is open and non-empty for all $\delta \in U(\delta_0)$. It furthermore follows that $u^{-1}((u(x^*(\delta)), u(x(\delta)))) \cap \text{Int}(B_\delta)$ is open and non-empty. Non-emptiness holds because $u(\alpha x(\delta) + (1-\alpha)x^*(\delta)) \in (u(x^*(\delta)), u(x(\delta)))$ for all $\alpha \in (0,1)$, by strict convexity and the fact that $x(\delta)$ uniquely maximizes u on B_δ. Thus by continuity there exists $\gamma \in (0,1)$ such that $u(\gamma(\alpha x(\delta) + (1-\alpha)x^*(\delta))) \in (u(x^*(\delta)), u(x(\delta)))$ and $\gamma(\alpha x(\delta) + (1-\alpha)x^*(\delta)) \in \text{Int}(B_\delta)$.

Therefore

$$W^s(\succeq, D) = \int_\Delta \lambda(\{y \in B_\delta : y \succ x^*(\delta)\})(1-\epsilon) d\nu(\delta)$$

$$\geq \int_{U(\delta_0)} \lambda(u^{-1}((u(x^*(\delta)), u(x(\delta)))) \cap \mathrm{Int}(B_\delta))(1-\epsilon)d\nu(\delta) > 0.$$

A.4 Non-Monotonicity of Rational Inattention

Consider an example in the finite setting of Matejka and McKay (2015) with $B = \{x, y\}$. Suppose the decision-maker has one of two possible utility functions. Utility function u_1 with $u_1(x) = 2$ and $u_1(y) = 0$ has prior probability q. Utility function u_2 with $u_2(x) = 0$ and $u_2(y) = 1$ has prior probability $1 - q$. Setting $\gamma = 1$ without loss of generality, the normalization condition (17) in Matejka and McKay (2015) becomes

$$\frac{q}{p(x,B) + (1-p(x,B))e^{-2}} + \frac{1-q}{p(x,B) + (1-p(x,B))e^{1}} = 1,$$

from which the unconditional choice probability $p(x, B)$ can be obtained. The two conditional choice probabilities $p(x, B|u_1)$ and $p(x, B|u_2)$ then follow from (3).

For $q = 0.35$, we obtain the numerical solution $p(x, B) \approx 0.45$. Hence $p(x, B) < p(y, B)$ even though $q2 + (1-q)0 = 0.7 > 0.65 = q0 + (1-q)1$, showing that unconditional choice probabilities are not monotone with respect to the prior preference.

For $q = 0.6$, we obtain $p(x, B) \approx 0.89$, from which it follows that $p(x, B|u_1) \approx 0.98$ and $p(x, B|u_2) \approx 0.74$. Hence $p(x, B|u_2) > p(y, B|u_2)$ even though $u_2(x) = 0 < 1 = u_2(y)$, showing that conditional choice probabilities are not monotone with respect to the true preference.

A.5 Proof of Proposition 3

As in the proof of Proposition 1, we can decompose the swaps distance into a singular part and an absolutely continuous part,

$$\begin{aligned} W(\succeq, D) = &\int_\Delta \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ x\}) d\mu^s_\delta(x) d\nu(\delta) \\ &+ \int_\Delta \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ x\}) f_\delta(x) dx d\nu(\delta). \end{aligned}$$

For the singular part (first line), it follows like in the proof of Proposition 1 that $\succeq_*$ achieves the minimum of 0 under weak $\succeq_*$-monotonicity (because weak monotonicity coincides with monotonicity concerning the singular part).

Now consider the absolutely continuous part (second line). As in the proof of Proposition 1, it is sufficient to show that $\succeq_*$ minimizes this expression in the case that the choice data D has only one equivalence class $[\delta] = \Delta$ (i.e., $\Delta/\simeq$ is a singleton). Since condition $(i)'$

in the weak monotonicity definition holds for every equivalence class individually, the proof then immediately extends to the general case.

Denote the unique budget set by B, i.e., we have that $B_\delta = B$ for all $\delta \in \Delta$. We can then rewrite the absolutely continuous part as

$$\begin{aligned}
\int_\Delta \int_{B_\delta} \lambda(\{y \in B_\delta : y \succ x\}) f_\delta(x) dx d\nu(\delta) &= \int_{[\delta]} \int_B \lambda(\{y \in B : y \succ x\}) f_{\delta'}(x) dx d\nu(\delta') \\
&= \int_B \int_{[\delta]} \lambda(\{y \in B : y \succ x\}) f_{\delta'}(x) d\nu(\delta') dx \\
&= \int_B \lambda(\{y \in B : y \succ x\}) \int_{[\delta]} f_{\delta'}(x) d\nu(\delta') dx \\
&= \int_B \lambda(\{y \in B : y \succ x\}) g(x) dx,
\end{aligned}$$

where $g(x) := \int_{[\delta]} f_{\delta'}(x) d\nu(\delta')$. Weak $\succeq_*$-monotonicity implies $g(x) \geq g(y)$ whenever $x \succeq_* y$. The result now follows exactly like in the proof of Proposition 1.