Abstract: This paper presents a model of choice induced by constrained optimization, where the constraint sets are subjective. The main idea is that the decision-maker (DM) faces an unobservable constraint, e.g. temptation, at the point of choice and this constraint distorts choices by rendering some objects in the menu infeasible. This creates a distinction between the (objective) physical set of choices and the subset of (subjectively) feasible choices. Whereas a standard (Arrowian) DM maximizes a fixed utility over the physical set of choices, our DM’s maximize a fixed utility on the set of subjectively feasible choices. Consequently, when the two sets don’t agree, choices from the overall opportunity set may look non-Arrowian even though, from the DM’s perspective, choices are completely standard. The model provides a concise explanation for two experimentally confirmed phenomena, (i) the compromise effect and (ii) (a version of) the attraction effect, both of which are well-documented examples of WARP violations, but neither of which can be explained by (most) existing models of menu choice. Moreover, the parameters of the utility are behaviorally identified.

Keywords: Menu Choice, Preference for Commitment, Compromise Effect, Attraction Effect.

*I thank Andrea Cann Chandrasekher, Hector Chade, Eddie Dekel, Juan Dubra, Haluk Ergin, Amanda Friedenberg, Faruk Gul, Bart Lipman, Wolfgang Pesendorfer, and the audiences of the Southwest Economic Theory Conference (SWET) 2009, NASM 2009, ESWC 2010, and LACEA-LAMES 2011 for useful feedback. Previous drafts have circulated under the titles “A Theory of Temptation without Uncertainty” and “A Model of Temptation-Induced Compromise”. Any errors are my own.

†Mailing address: Department of Economics, W.P. Carey School of Business, P.O. Box 873806, Tempe, AZ 85287-3806; E-mail address: madhav.chandrasekher@asu.edu
## Contents

1 Introduction ................................................. 3  
  1.1 Experimental Evidence ................................... 5  
  1.2 Three Classes of Anomalies ............................... 7  
  1.3 Related Literature ...................................... 9  

2 Model .......................................................... 11  
  2.1 Model Description ....................................... 11  
  2.2 Axioms .................................................. 15  
  2.3 Main Results ........................................... 18  
  2.4 Identification .......................................... 23  

3 Comparison between Models ................................. 26  
  3.1 Comparison with the Strotz model ....................... 27  
  3.2 Identification Revisited ................................ 32  
  3.3 Comparison with Costly Self-Control models .......... 34  

4 Menu Choice vs. Choices from Menus ....................... 38  

5 Conclusion .................................................. 42  

6 Appendix .................................................... 43  
  6.1 Proofs for Section 2 ................................... 43  
  6.2 Proofs for Section 3 ................................... 58  
  6.3 Proofs for Section 4 ................................... 63
1 Introduction

Consumer theory is based on the assumption that choices are derived from utility maximization, hence can be used to conduct welfare analysis. However, there is a growing body of experimental evidence that challenges this hypothesis. The issue is well-illustrated with a simple anecdotal example. If choices reveal preferences, then how can we explain why consumers, on the one hand, declare certain choices, e.g. smoking, procrastinating, etc., to be harmful (i.e. welfare reducing) and, at the same time, make these very same choices? If we assume a decision-maker (DM) has a well-defined ranking on consumption options, then this behavior cannot be derived from maximization of this ranking.

The literature on temptation and self-control, initiated by the seminal piece Gul and Pesendorfer (2001), arose, in part, to provide a revealed-preference model that could reconcile observed conflicts between choice and welfare. These models show that we can recover behavior such as addiction, excessive risk taking, etc. via maximization of an alternative ranking on choices. This ranking need not coincide with the welfare (normative) ranking and is usually expressed as a difference between welfare utility and a cost function, representing self-control costs. Hence, whilst choices are non-classical in that they cannot be recovered from welfare maximization, they can, nonetheless, be recovered by maximization of an alternative “net welfare” ranking. An important feature of this approach is that one needs to observe not just the welfare ranking on choice objects, but the welfare ranking on choice problems, i.e. menus of choice objects.

This paper provides an alternative to the Gul and Pesendorfer (2001) framework. As in their approach, our goal is to reconcile the conflict between choice and welfare. However, a key distinction between our model and those that take the costly self-control approach, after Gul and Pesendorfer (2001), is that choices reveal true welfare (as opposed to net welfare). In other words, utility from an option in a given choice problem is the same as it would be if the choice problem consisted of that option alone. Whereas the costly self-control approach explains choices via (unconstrained) maximization of a net welfare ranking, we explain choices via constrained maximization of the true welfare ranking. Hence, while preferences over choice objects are standard, choice data can nonetheless be non-standard (i.e. violate WARP) because we do not maximize over the full domain of the choice problem.

---

1All of our analysis can be carried out with two observables, (i) a welfare ranking $u(\cdot)$, which is a ranking on consumption choices, and (ii) a choice correspondence, $C(\cdot)$. Moreover, in modeling the attraction effect we only need the choice correspondence. However, for the purpose of brevity and ease of comparison with existing models, we will write axioms using menu preferences. In a subsequent section, we show that there is a one-to-one correspondence between the described pair of observables, $(u(\cdot), C(\cdot))$, and the menu preference that this pair generates.
In our model, temptation acts as a (subjective) constraint on the set of choices available to the DM. The idea is that, while the choice problem describes the physical set of choices, the presence of temptation creates a tension between what is physically feasible and what the DM subjectively views as feasible. Let us illustrate with the following limiting case of a costly self-control preference. Imagine the DM is addicted to nicotine, e.g. he cannot resist a cigarette whenever one is offered. This can be expressed as the following preference over choice problems: \{accept cigarette, decline cigarette\} \sim \{accept cigarette\}. The first menu represents the option to either accept or decline a cigarette. The singleton menu represents a commitment to smoke, i.e. the DM does not have the option to decline. The fact that the DM is indifferent between these two choice problems reveals his addiction. In this example, the physical set of choices is to either accept or decline. However, in the presence of addiction it is as if the option to decline doesn’t even exist. Consequently, the only subjectively feasible choice is to accept the cigarette.

Choices from the subjective feasible set are standard, i.e. they maximize welfare. However, since the subjective and objective feasible sets need not agree, there can be a distinction between the unconstrained welfare optimum and the temptation-constrained welfare optimum. This is the source of non-standard choice behavior in our model. Hence, the key feature of the model is the manner in which it explains how objective choices are reduced to the subset of subjectively feasible choices. This “reduction” process works through an object called a category, denoted \(C\), and is the main ingredient of the model. We give a formal description of the model in the next section, but – in brief – it is summarized by two parameters \((u, C)\). First, a consumption ranking on alternatives \(u\) (we sometimes call this a welfare ranking) and, second, a collection of subsets, \(C_i \subseteq X\), of the grand set of alternatives with the property that \(\bigcup_i C_i = X\). We refer to a pair \((u, C)\) as a category model. The category \(C\) denotes the collection \(\{C_i\}\).

The idea behind the model is that choice objects are comprised of multiple subjective attributes and the sets \(C_i\) group together those objects which share a given attribute. Using these sets, choices are determined by the following two-stage constrained maximization process. First, within each set \(C_i\) temptation “chooses” the object with the lowest \(u\)-value. Aggregating across sets \(C_i\) yields the (subjectively) feasible set. Second, the DM maximizes \(u\) on the subjective feasible set. Since elements in the feasible set (disregarding ties for the moment) come from different sets \(C_i\), the latter step can be interpreted as the DM “choosing” an attribute. Hence, the DM optimally chooses an attribute knowing that, once this selection is made, his temptation will render only the \(u\)-minimal choices feasible among objects which share that attribute. The key point here is that the DM’s (subjective) choice variable is the set of attributes. When attributes finely describe choices, so that each set
\( C_i \) consists of one choice alone, then there is no choice distortion and we recover the standard model. However, when attributes coarsely describe choices, so that some sets \( C_i \) are non-singleton, then choices can be non-standard since optimal attribute choice is exercised only over the domain of feasible (temptation-admissible) objects which possess that attribute.

1.1 Experimental Evidence

The model provides a concise explanation for some important examples of experimentally confirmed choice “anomalies”\(^2\). The evidence we reference is described in Simonson and Tversky (1993).\(^3\) The authors report the results of two experiments. In the first, subjects are given two choice problems in sequence, where in both problems the objective is to select a camera from a menu of choices. In the first problem, the choices are (i) a very cheap 35mm or (ii) a standard 35mm. After making a selection from this menu, subjects are then asked to choose from (i) the same cheap 35mm, (ii) the standard 35mm, or (iii) an accessorized and expensive 35mm. The reported results are given in the following table:

<table>
<thead>
<tr>
<th>Choice Set</th>
<th>Market Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {m, m^+} )</td>
<td>50% 50% N/A</td>
</tr>
<tr>
<td>( {m, m^+, m^{++}} )</td>
<td>21.5% 57% 21.5%</td>
</tr>
</tbody>
</table>

Table 1: Compromise Effect, Simonson and Tversky (1993)

The symbols denote (resp.): \( m = \) Minolta X-370, \( m^+ = \) Minolta 3000i, \( m^{++} = \) Minolta 7000i. Here \( m \) is cheap, \( m^+ \) is standard, and \( m^{++} \) is accessorized. From left to right, columns denote respective market shares of \( m, m^+ \), and \( m^{++} \). This phenomenon has received the title of the “compromise effect”.

The second Simonson and Tversky (1993) experiment documents a phenomenon now known as the “attraction effect”. Their experiment is as follows. Subjects are offered two choice problems in sequence. In the first choice problem, they are asked to choose between two VCR players, (i) a generic (e.g. Emerson) or (ii) a well-known brand (e.g. Panasonic) that is higher priced than the generic, but is substantially discounted from its original price. In the second choice problem, a third option is added. The subjects are asked to choose from (i) the same Emerson, (ii) the well-known brand, or (iii) a generic at the discounted price. The reported results are given in the following table:

<table>
<thead>
<tr>
<th>Choice Set</th>
<th>Market Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>( {m, m^+} )</td>
<td>50% 50% N/A</td>
</tr>
<tr>
<td>( {m, m^+, m^{++}} )</td>
<td>21.5% 57% 21.5%</td>
</tr>
</tbody>
</table>

Table 1: Compromise Effect, Simonson and Tversky (1993)

By this we mean choice behavior which is natural (and has received experimental confirmation), but nonetheless cannot be recovered by the standard model, i.e. unconstrained maximization of a single preference relation on alternatives.

\(^3\)In Table 1, Simonson and Tversky (1993) do not report full percentages for the second experiment. What is reported is that the market share of \( m^+ \) is 57% and the other two choices have approximately equal market share. Based on this comment we have imputed 21.5%.
(ii) the same Panasonic, and (iii) a second, lower-quality Panasonic whose sticker price is lower than the sticker price of the other Panasonic but which is not on sale, so that the higher-quality Panasonic is cheaper once the discount is factored in. Choices are abbreviated are as follows: \( e \) = Emerson, \( p_+ \) = high quality Panasonic w/ good discount, \( p_- \) = good quality Panasonic w/ lesser discount. The data from the experiment is given in the following table.

<table>
<thead>
<tr>
<th>Choice Set</th>
<th>Market Share</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( { e, p_+ } )</td>
<td>57%</td>
<td>43%</td>
<td>N/A</td>
</tr>
<tr>
<td>( { e, p_+, p_- } )</td>
<td>27%</td>
<td>60%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Table 2: Attraction Effect, Simonson and Tversky (1993)

Notice that both effects are illustrations of “choice distortion”. In other words, there are two initial options, e.g. the Emerson and the Panasonic, where one option is more popular than the other. Once we add a third option, a majority of the sample now switches to what was originally the less popular option, e.g. the Panasonic. These effects are considered behaviorally distinct since the reasons for the switch are distinct. In both cases, there are two attribute dimensions that can describe the choice objects, (i) price and (ii) relative quality. In the case of the attraction effect, sometimes referred to as the “asymmetric dominance” effect, the existence of a dominated option (e.g. \( p_- \)) distorts choice towards the dominating, and previously unselected, option (e.g. \( p_+ \)). In the case of the compromise effect, the existence of a more extreme option (e.g. \( m_{++} \)) distorts choice towards a middle-price, and previously unselected (but not unambiguously dominated), option (e.g. \( m_+ \)).

Our model explains both of these effects as stemming from “temptation”, where temptation denotes a willingness to pay for additional features. In the case of the compromise effect this is expressed as a willingness to pay for superfluous amenities (frills), and in the case of attraction a willingness to swap a generic for a higher priced brand name. The two examples indicate different levels of willingness to make this tradeoff. In the case of compromise, the DM is evidently tempted to splurge on extravagances when these options are available. Moreover, the DM realizes this tendency and, hence, switches to the middle priced option when all three options are present to mitigate this urge. In the case of attraction, he also has a willingness to trade higher price for more amenities, but this tendency is less severe than in the compromise example. In particular, he only makes the switch to the middle-priced option if he is convinced that this is a good deal. “Temptation” refers to the fact that he can be persuaded to spend money on accessories so long as choices are presented in a manner that makes them seem reasonably priced.

\[4\] Neither of these examples can be explained by the benchmark (i.e. Gul and Pesendorfer (2001),
For both effects, choosing among options involves tradeoffs between prices and other attributes of the choice objects, e.g. brand name. To appropriately capture these tradeoffs, we take a subjective approach, so that the attributes of the choice object that the DM uses as a basis for his choice are not outwardly observable to the modeler. To motivate this approach, note that the attraction and compromise effects both arise from relative comparisons between choice objects. For instance, while everyone sees the same list of features that describe the Minolta 7000i, choice is determined only by the subset of features (attributes) that the DM uses in his evaluation. A flash with a built-in timer might be considered a relevant feature by one DM and completely superfluous by another. Hence, the set of attributes that are used to determine choice is subjective and needs to be inferred from choice data. Accordingly, the observables of the Simonson and Tversky (1993) experiments are described by a pair \((u, C)\) consisting of a price function (actually, we take \(u = \text{negative price}\)) and a choice correspondence. More generally, we will take \(u\) to be a representation of some observable ranking on consumption choices. Having said this, most analysis in this paper is cast in the language of a preference order on menus. In most models of temptation, e.g. Gul and Pesendorfer (2001), this is strictly stronger than taking both a consumption ranking and a choice correspondence as a primitive. However, for the model in this paper it turns out that the two primitives are equivalent, so that by taking a menu preference as the observable we are (equivalently) taking the same observables as in the Simonson and Tversky (1993) experiment.

### 1.2 Three Classes of Anomalies

Anomalous choice data refers to pairs \((u, C)\) where \(C(\cdot)\) is not recovered as the arg max of \(u(\cdot)\), i.e. any pairs \((u, C)\) which are not generated by the Arrowian model \(U(A) = \max_{x \in A} u(x)\) are “anomalous”. Consider pairs \((u, C)\) where \(u\) represents the consumption ranking and \(C(\cdot)\) is a choice correspondence. The Simonson and Tversky (1993) model describes two types of choice anomalies involving triples \(\{x, y, z\}\). There is a third possible anomaly which, together with the compromise and attraction effect, yields the following trichotomy. Put \(u(x) > u(y) > u(z)\) and assume \(C(\{y, z\}) = \{y\}\).

1. (Compromise) \(C(\{x, y\}) = \{x\}, C(\{x, z\}) = \{z\}, C(\{x, y, z\}) = \{y\}\).
2. (Attraction) \(C(\{x, y\}) = \{x\}, C(\{x, z\}) = \{x\}, C(\{x, y, z\}) = \{y\}\).

Dekel et al. (2009)) models in the temptation literature. However, the model in Noor and Takeoka (2010) can explain the compromise effect. To our knowledge, no models in this literature can explain the attraction effect, or more precisely, the Simonson and Tversky (1993) version of the attraction effect.

5This idea – taking \(u\) to equal negative price – is taken from an example of Efe Ok.
3. (Strotzian Choice)

(a) \( C(\{x, y\}) = \{y\}, C(\{x, z\}) = \{z\}, C(\{x, y, z\}) = \{y\} \).

(b) \( C(\{x, y\}) = \{y\}, C(\{x, z\}) = \{x\}, C(\{x, y, z\}) = \{y\} \).

Imagine we observe the following choice data. First, the ranking on singleton menus is such that \( u(x) > u(y) > u(z) \). Second, \( y \) is chosen from \( \{y, z\} \) and also from \( \{x, y, z\} \). That is, the middle (welfare-wise) option is picked both over the welfare-minimal option and from the menu containing the minimal and maximal options. Now consider the ways we could fill out the rest of the choice data. The listed patterns taxonomize choices from the pairs \( \{x, y\} \), \( \{x, z\} \) that are consistent with the choice of \( y \) from \( \{x, y, z\} \) (and \( \{y, z\} \)). We have labeled the first two choice patterns according to the Simonson and Tversky (1993) experiment. The final group is what we have called Strotzian choice. While it not covered by the Simonson and Tversky (1993) experiment, it is an example of non-Arrowian choice and admits a representation by the well-known Strotz model of menu choice.

An important feature of the model in this paper is that it reflects this trichotomy. To each choice pattern listed above we attach a sub-class of the category model (with an accompanying behavioral characterization). Each sub-class has the property that deviations from WARP are due solely to the type of anomaly to which that sub-class is associated. Hence, there is a sub-class of models for which deviations from WARP occur only due to compromise cycles, another sub-class for which deviations are only due to attraction, and a third for which WARP deviations are due solely to Strotzian choices. Moreover, the full model exhibits deviations from WARP only due to one of these types of choice patterns. In this way, the model provides a concise explanation for these three classes of choice anomalies.

The remainder of the paper is organized as follows. In section 2, we present the model, axioms, and main representation results. Section 3 carries out a comparison between our model, the Strotz model of menu choice, and the costly self-control models (after Gul and Pesendorfer (2001)). There are important connections between these two models. For example, our model nests the Strotz model and the manner in which this nesting occurs is itself an object of interest. It shows that we

\[6\] Note that the difference between compromise and attraction is in the choice from the menu \( \{x, z\} \) (consisting of the normative best and normatively worst option). This matches our intuition that the difference between these effects lies in how willing the DM is to trade off price against other attributes. However, there is no reported data in Simonson and Tversky (1993) on choices from the problem \( \{e, p_+\} \) (resp. \( \{m, m_{++}\} \)). Hence, when we impute choices from \( \{x, z\} \) we are adding data to the observed choices. This is, however, forced upon us if we want these effects to be observably distinguishable. For this we must impute \( C(\{x, z\}) = \{x\} \) for attraction and \( C(\{x, z\}) = \{z\} \) for compromise.

\[7\] For future reference, we refer to this triple as a behavioral trichotomy.
can equivalently interpret Strotzian preferences as arising from (constrained) maximization of a single preference relation as opposed to the two-stage procedure of the Strotz utility, which uses two preference relations. These results also have relevance for what we called the “behavioral trichotomy” of cycles. They connect each class of cycles to a corresponding subclass of the category model. Finally, section 4 shows that we can equivalently recast the main results of the paper in terms of ex post choices. Proofs are collected in the appendix.

1.3 Related Literature

This paper connects two strands in the recent choice theory literature. On the one hand, the literature on temptation-driven preferences and, on the other, the literature on quasi-rational (i.e. boundedly rational) choice. Our menu preferences are “temptation-driven” in the spirit of Dekel et al. (2009). The connections to the temptation literature and, in particular, the Gul and Pesendorfer (2001) and Dekel et al. (2009) models are examined in greater detail in section 3 of the paper, after the formal model is introduced. The recent literature on quasi-rational choice, e.g. Cherepanov et al. (2008), Lleras et al. (2008), Masatlioglu et al. (2012), de Clippel and Eliaz (2012), Manzini and Mariotti (2012), Ok et al. (2010) generally considers two-stage models of constrained optimization. In the first stage, the physical choice set is reduced to a subset of options. In the second stage, some preference relation is maximized on the reduced set of options. de Clippel and Eliaz (2012) develop an axiomatic model of intra-personal bargaining that captures versions of the compromise and attraction effects.

We share the experimental motivation with de Clippel and Eliaz (2012), but there are important formal and conceptual differences in the way in which we explain the experiments. Moreover, the physical evidence we explain is formally distinct. For instance, the specific data of the Simonson and Tversky (1993) experiments cannot be fitted in the model of de Clippel and Eliaz (2012). Following their paper, define $S_{\succeq i}(x, A) := |\{y \in X : x \succ_i y\}|$ to be the score of $x$ in the menu $A$ (w.r.t. the relation $\succeq_i$). For a fixed pair $(\succeq_1, \succeq_2)$ define $U(A) = \max_{x \in A} \min_{i=1,2} S_{\succeq i}(x, A)$. This is the de Clippel and Eliaz (2012) model. To illustrate the difference with our model, let us consider how the two models explain the attraction effect. Note that if we want $C(\{x, y, z\}) = \{y\}$, then we must have $C(\{x, y\}) = \{x, y\}$ in de Clippel and Eliaz (2012) (unless $x$ is never chosen). The story is that the ranking on the pair $(x, y)$ depends on the relation $\succeq_i$, where these relations are interpreted as attribute dimensions. To induce a choice of $y$ by adding $z$ (as, say, the attraction effect requires), we must have (i) $z$ beat $x$ under the same relation $\succeq_i$ for which $y$ beats $x$.

---

8More precisely, we define an ex post observable to be a pair $(u, C)$, consisting of a welfare ranking on choices and a choice correspondence. The equivalence is between pairs $(u, C)$ and menu preferences which induce such pairs.
(so that the minimum score of $x$ in the menu $\{x, y, z\}$ is 0), and (ii) $y$ must beat $z$ under both relations $\succeq_1, \succeq_2$. It follows that $y$ is chosen from the triple $\{x, y, z\}$ because it is now less worse than $x$, from a Rawlsian (max-min) perspective. In this sense, it is a compromise between $x$ and $z$.

This explanation for the attraction effect is conceptually distinct from the one we give.\footnote{The comparison with de Clippel and Eliaz (2012) is a little strained since the actual data we associate with the compromise and attraction effect is different.} In our model $y$ (the unchosen option from $\{x, y\}$) is unambiguously worse than $x$ where, with the Simonson and Tversky (1993) experiments in mind, worse refers to the normative rank $u(\cdot)$ (= negative of price). The switch to $y$ occurs because the introduction of $z$ constrains the DM from choosing on the basis of price alone. With $z$ and $y$ in the option set, he gives in to the temptation of buying a brand name and, conditional on brand, selects the most affordable option. This is how we interpret the attraction effect.

Our model falls under the rubric of the “reduce then maximize” approach that describes many of the listed quasi-rational choice models. The differences are in how we model quasi-rationality and in the specificity of our approach. Some papers we have listed are “multiple-self” models, where choice is determined by aggregation amongst multiple (possibly conflicting) preference relations. In these models, choice is not rational since it is not determined by a single relation. However, it is quasi-rational in the sense that it can be recovered in two steps by fixing (i) a collection of preferences and (ii) a menu-dependent selection rule which determines the preference in the collection that governs choice from the given menu. In our model, there is only one preference relation. This is a key difference with multi-self models since recovery of welfare from choice is an important desideratum of the category model.

There are more similarities with the recent literature on choice under consideration. For instance, the models in Lleras et al. (2008), Ok et al. (2010), Manzini and Mariotti (2012), and Masatlioglu et al. (2012) describe, in novel ways, a general process where objective choices are filtered down to a set of subjective choices and preferences are then maximized on the filtered set. This paper also describes such a process, but the innovation is in producing a model that explicitly describes how and why choice filtering occurs. In existing models the choice filter is endowed with some properties\footnote{We will have much more to say about the comparison between consideration set-based models and the category model in section 2.} that gives the model its empirical content, but the main contribution of these papers is in showing the (surprising) fact that “reduce then maximize” models are, in general, characterized by a natural weakening of the WARP axiom. Our goal is not to provide a behavioral foundation for choice filtering processes per se, but to demonstrate that a particular class of filters (the category model) is characterized by
a criterion for subjective feasibility – as described in our axioms. Choice objects are categorized according to their subjective attributes and the DM optimally selects an attribute which – on account of temptation – coarsely determines his actual choice. Hence, the category can be thought of as a specific type of (implicit) choice filter. The additional structure possessed by categories over an abstract filter makes for a more narrow model. For instance, categories are identifiable from choice. Moreover, compromise and attraction effects, along with Strotzian choice, are the only kinds of non-standard choice patterns admitted by our model. If the preference does not exhibit either of these patterns, then the DM is Arrowian, i.e. there is no filter and he maximizes a fixed preference over the full menu.

There is also a recent axiomatic literature on random choice which provides a rationalization of the compromise and attraction effects. Gul et al. (2010) develop a multi-attribute generalization of the well-known Luce (logit) model of random choice which can also account for these effects. The connection with our model is through the idea that these choices patterns are induced by (random) choice rules that use multiple (subjective) attributes to assign probability weights to options. For example, in the Gul et al. (2010) model the presence of more brand name products makes the brand name more salient in the mind of the DM, hence shifting overall choice probability towards brand name options. Another recent random choice model that can explain the attraction effect is Natenzon (2010). In this model, the DM receives information that informs his preferences over choice objects. The author characterizes a class of random choice rules in which the attraction effect emerges as a consequence of information acquisition over time.

2 Model

The choice environment is described as follows.

- Let $X = \{x_1, \ldots, x_n\}$ be an enumeration of the consumption space.
- Let $\mathcal{M}$ denote the collection of non-empty subsets of $X$ (menus).
- Let $P(X)$ be the set of complete, transitive preference relations defined on $\mathcal{M}$.

Henceforth, preference relations $\succeq$ will be assumed to be complete and transitive. For any menu $A$ we take $\sup(A)$, $\inf(A)$ to be (resp.) the set of $(\succeq)|_X$-maximal (resp. $(\succeq)|_X$-minimal) elements in the menu $A$.

2.1 Model Description

The model has two parameters:
• Take $\mathcal{C} := \{C_i\}_i$ to be a collection of subsets of the consumption space $X$ (call this collection a \textit{category}).

• Let $u(\cdot) : X \to \mathbb{R}$ be a representation of the ranking on consumption choices, i.e. the welfare ranking.

Using these two ingredients we construct the following utility on menus (i.e. choice problems):

$$U(A) := \max_{C_i, \mathcal{C}, \cap A \neq \emptyset} \min_{x \in C_i \cap A} u(x)$$

Hereafter, in writing the above formula we will suppress the requirement that the maximum is taken only over those $i$ such that $C_i \cap A \neq \emptyset$\footnote{This can be avoided if we declare, by fiat, the minimum of a function over the empty-set to be some $-K$ where $-K < u(x), \forall x$.}. In this paper, a \textit{category model} is defined to be any pair $(u, \mathcal{C})$ as given above. Break up the utility formula given above into three steps, respectively labelled (i) categorization, (ii) reduction, and (iii) maximization. We associate the category utility with the following decision procedure: (put $A^* := \bigcup_i \arg \min_{x \in A \cap C_i} u(x)$)

\[
A \xrightarrow{\text{categorize}} \{C_i \cap A\}_i \xrightarrow{\text{reduce}} A^*(= \bigcup_i \arg \min_{x \in A \cap C_i} u(x)) \xrightarrow{\text{maximize}} \max_{x \in A^*} u(x) = U(A)
\]

The key feature of our model is the first step, which determines how the menu $A$ of objectively feasible options is reduced to the set $A^*$ of subjectively feasible options. Once the categories are determined then temptation reduces the objective set of options, say $A$, to the “reduced” set of feasible options $A^*$. The DM then maximizes consumption utility over the reduced set, which we equivalently think of as maximizing a choice of attribute since (unless there are ties – in which case he exhibits indifference between some attributes for the given menu) elements of $A^*$ correspond one-to-one with elements of $\bigcup_i \inf(A \cap C_i)$.

The map $\min : A \cap C_i \mapsto \inf(A \cap C_i)$, which determines feasibility conditional on attribute selection captures one – seemingly stark – manner in which the “reduction” process might occur. The important thing to keep in mind is that the collection $\{C_i\}_i$ is a choice variable of the DM. There is no restriction on how fine or coarsely the attributes describe the choices, so that when we aggregate the arg minima across all attributes the minimum operator may not have reduced any options at all\footnote{This is what happens if we take attributes to perfectly describe choices, i.e. take $C_x = \{x\}$ and put $\mathcal{C} \equiv \{C_x\}_{x \in X}$. In this case $\bigcup_{x \in A} \inf(A \cap C_x) = A$ and we recover the standard model.}. The minimum operator has bite precisely when attributes coarsely describe choice objects. Since the DM can only choose attributes, he can only affect his consumption up to choices which share the selected attribute. In this sense, we consider a set $C_i$ as a (potential) constraint on the choice of $x$, where $x \in \sup(C_i)$.
The idea being that any \( y \in \mathcal{C} \) where \( u(x) > u(y) \) obstructs the DM from selecting \( x \), unless there is an attribute present that separates \( x \) from \( y \).

It is useful to rewrite the category model in the following way. For each \( x \in X \), put

\[
\mathcal{C}(x) := \{ \mathcal{C}_i \in \mathcal{C} : x \in \sup(\mathcal{C}_i) \}
\]

That is, \( \mathcal{C}(x) \) is the (possibly empty) sub-collection of \( \mathcal{C} \) consisting of sets whose \( u \)-maximal element is \( x \). Note that we have \( \mathcal{C} = \bigcup_{x \in X} \mathcal{C}(x) \). We can hence re-express the menu utility associated to the category model in the following form:

\[
U(A) = \max_{\mathcal{C}_i \in \mathcal{C}} \min_{z \in \mathcal{C}_i \cap A} u(z) = \max_{x \in X} [\max_{\mathcal{C}_i \in \mathcal{C}(x)} \min_{z \in \mathcal{C}_i \cap A} u(z)]
\]

The full maximization problem breaks up into a collection of smaller problems where, in each sub-problem, the DM’s choice domain is the collection \( \mathcal{C}(x) \). Notice that \( x \) is the unconstrained welfare optimum over all elements in the sets \( \mathcal{C}_i \in \mathcal{C}(x) \). Hence, the objective of this (sub)problem is to select the option whose \( u \)-value is as close to \( x \) as possible. When \( \{x\} = \mathcal{C}_i \cap A \), for some \( \mathcal{C}_i \in \mathcal{C}(x) \) then \( x \) is selected. However, if there is no such set then the DM is genuinely constrained from choosing \( x \). In this case, the indices on the sets \( \mathcal{C}_i \) label the attributes that tempt the DM away from \( x \) and the DM can only coarsely select a choice via selecting a \( \mathcal{C}_i \) that contains that choice, which we interpret as selecting an attribute.

Let us now see how this works through the compromise and attraction examples. The elements \( x, y, z \) denote (resp.) a low price alternative \((e)\), a brand name alternative with a discount \((p_+)\), and a brand name alternative without a discount \((p_-)\). For the compromise example, we put \( y = (m_+) \) as the middle-range Minolta and \( z = (m_{++}) \) as the Minolta with all the frills, and \( x = (m) \) is the regular Minolta. Let \( u(\cdot) \) rank options based on affordability, i.e. negative of the posted price, so that \( u(x) > u(y) > u(z) \).

**Example 1** (Compromise). Put \( \mathcal{C}_1 := \{y\}, \mathcal{C}_2 := \{x, z\} \) and \( \mathcal{C} \equiv \{\mathcal{C}_1, \mathcal{C}_2\} \). Let \( U(\cdot) \) denote the associated menu utility, i.e. \( U(A) := \max_{\mathcal{C}_i} \min_{z \in A \cap \mathcal{C}_i} u(z) \). Notice that:

- \( U(\{x, z\}) = u(z), U(\{x, y\}) = u(x) \).
- \( U(\{x, y, z\}) = u(y) \).

Let us interpret the sets \( \{\mathcal{C}_1, \mathcal{C}_2\} \) comprising the category. Imagine the DM wants to choose \( x \) (the “no frills” camera). This object might be described by two attributes (i) price and (ii) shutter speed. Imagine the \( z \) option (the “frills” camera) is described by its price and the fact that it has an adjustable shutter speed. The DM’s photography needs require only a bare bones camera such as \( x \), but when a
camera that allows for more serious photography is present he is tempted to splurge on this option. Now consider the third option, \( y \), which is described by price and some other attribute, say, battery life. The labels on the sets \( C_1, C_2 \) are (resp.) taken to be (i) shutter speed and (ii) battery life. These are (subjective) choice variables of the DM. When both \( x, y \), and \( z \) are present in the menu, on account of the DM’s temptation he only “sees” the options \( y \) and \( z \). Hence, it is as if he is choosing between a high price camera with adjustable shutter speed or a moderately priced camera with generous battery life. Since these objects are separated by their attributes (evidently, the middle price camera does not also have adjustable shutter speed) choosing between objects is the same as choosing between attributes. Note that when the choice problem is \( \{x, y, z\} \) attributes only coarsely map to choices. This tension between the de facto domain of choice, i.e. the space of subjective attributes, and the actual choice domain is how the category model captures the idea of subjective feasibility.

We now turn to the attraction effect. Put \( x = e, y = p_+, z = p_- \) and recall the implied menu preference from the choices: \( \{x, y\} \sim \{x\}, \{x, z\} \sim \{x\}, \{x, y, z\} \sim \{y\} \). Here there are three attributes that can be used to describe the choices, (i) price, (ii) brand name and (within-brand) model type, and (iii) whether the object is on discount. Hence, \( x \) is a low price generic, \( y \) is a high-end model with a name brand that is on discount, and \( z \) is high-end name brand that is not on discount (it is also not as high-end as \( y \), though this is not a critical component of the representation or in the evidence from Simonson and Tversky (1993)). Construct the following category model representation.

**Example 2** (Attraction). Let \( C_1 := \{x, z\}, C_2 := \{x, y\}, C \equiv \{C_1, C_2\} \). Let \( U(\cdot) \) denote the associated menu utility. Notice that:

- \( U(\{x, z\}) = u(x) = U(\{x, y\}) \).
- \( U(\{x, y, z\}) = u(y) \).

Interpret the indices on the sets \( C_1, C_2 \) as (i) brand name and (ii) brand name with a discount. These are the attributes that can be used to describe the non-generic options \( y, z \). When the menu of options is \( \{x, y, z\} \) the DM cannot select \( x \) (note that \( C(x) = \{C_1, C_2\} \) in this example), hence his subjective choice domain is the collection of tempting attributes – brand name without discount, which describes \( z \), and brand name with a discount which describes \( y \). Note that the DM’s choice in this case is to select the attribute \( C_2 \). This matches our intuition for attraction that the DM switches to the discounted Panasonic only once he is persuaded that the discount is a good deal. When the menu is \( \{x, y\} \) the DM selects the set \( C_1 \) and obtains \( x \). Here it is helpful to keep the decomposition \((*)\) in mind. When choosing between \( C_1 \) and \( C_2 \), the DM is trying to get as close to \( u(x) \) as possible (since \( C(x) = \{C_1, C_2\} \)). When either \( y \)
or $z$ is not on the menu, then his choice of set $C_i$ maps exactly to the choice of $x$ since he is not tempted at the menu $\{x, y\}$ (resp. $\{x, z\}$). However, when both $y, z$ are present he can only coarsely select $x$ by selected one of the sets $C_i$. In this case, he selects the (tempting) attribute whose associated choice has $u$-value closest to $x$.

### 2.2 Axioms

The central concept of the paper is the idea of subjective feasibility. By this we mean that there is a tension between the physical options on the menu and the subset of options that the decision-maker is able to choose. The set of choosable elements is the (subjective) feasible set. How is this set to be revealed to the modeler? To answer this, the first step is to formulate a revealed preference criterion for whether or not an object is feasible, from the viewpoint of the decision-maker. The following definition suggests such a criterion.

**Definition 1.** An element $x \in A$ is subjectively feasible if whenever $A' \subseteq A$ and $x \in A'$, then $A' \succeq \{x\}$.

Via the category model, welfare from a menu equals the value of the element chosen from the menu. More generally, consider the class of models:

$$ (*) \ U(A) = \max_{x \in \Theta(A)} u(x) $$

where $\Theta(A) \subseteq A$ is interpreted to be the set of subjectively feasible elements and is required to satisfy contraction consistency, i.e. $\Theta(B) \cap A \subseteq \Theta(A)$. For the category model we have $\Theta(A) = \bigcup_i \inf(C_i \cap A)$. We keep this more general model in mind to interpret the criterion in the definition. For brevity’s sake, we refer to the criterion in Definition 1 as the *welfare criterion* for feasibility.

The welfare criterion says that the value of the maximizer in $A'$, for any $A'$ containing $x$, never falls below the value of $x$. Put another way, the DM’s welfare is...

---

13\ footnotemark\ This is probably a good place to confess our unease with the word “feasibility” or, for that matter, “choosability”. When we say an object on the menu is infeasible, we do not mean the DM is physically unable to select it. Rather we simply mean that the DM is indifferent between having that object in or out of the existing option set. In other words, it is as if the object is infeasible. Nevertheless, in the absence of a more appropriate catchphrase we revert to the term (in)feasibility.

14\ footnotemark\ This more general model is very similar to the representation in, Lleras et al. (2008), Masatlioglu et al. (2012). The main difference is that the set $\Theta(A)$ (which they call a *consideration filter*) satisfies the following relaxation of Houthaker’s axiom, $\Theta(A \setminus x) = \Theta(A), \forall x \in A \setminus \Theta(A)$. The contraction consistency property on $\Theta(\cdot)$ neither implies nor is implied by this special case of Houthaker’s axiom. The intuition for contraction consistency is that if we pass to a smaller menu and keep the original choice intact, then we have possibly removed some temptations that rendered better (higher welfare) choices infeasible. Moreover, if $x$ was feasible in the presence of these temptations, then it should remain feasible when we remove them from the menu.
bounded below by \( u(x) \) so long as \( \{x\} \) is in the option set. What does this reveal about the DM’s constraints on choice? The inference we make is that this criterion (Definition 1) reveals whether \( x \) is choosable, i.e. it is within the (subjective) constraint set over which the DM maximizes his welfare ranking. However, this is just an inference, it need not hold for all models. If the DM uses a model like (*) to make choices, then the criterion in the definition is a merely a necessary condition for \( x \) to be subjectively feasible – it need not be sufficient. Indeed, if \( x \in \Theta(A) \) (the putative feasible set of \( A \)), then \( x \in \Theta(A') \) by contraction consistency of \( \Theta(\cdot) \). Hence, \( U(A') \geq u(x) \) for all \( A' \subseteq A \) with \( x \in A' \). If this is to be a criterion for feasibility, then we require that the condition in the definition is not just necessary, but also sufficient for feasibility. At the level of generality of (*) this need not be true. The contraction consistency condition does not, by itself, place enough structure on the map \( \Theta(\cdot) \) to ensure that \( \Theta(A) \) coincides with the set of feasible elements in \( A \) under the welfare criterion for feasibility. However, once we narrow our scope to the category model, the welfare criterion is necessary and sufficient for feasibility. That is, fixing any category model \((u, C)\) and putting \( \Theta(A) := \bigcup_i \inf(C_i \cap A) \) we have that \( x \in \Theta(A) \) if and only if it satisfies the welfare criterion.

This discussion leaves an extant question. The category model is just one model for which the welfare criterion reveals feasibility. Fixing this criterion as our revealed preference test for feasibility, a natural question is whether there is a larger class of models, which (i) nests all category models, (ii) is a subclass of (*), and (iii) for which the welfare criterion also reveals feasibility. Finding the maximal such class of models is important since one of the goals of this paper is to develop a model of choice which reconciles choice with welfare. To this end, we turn to our first axiom titled “CRW” – which stands for “Choice Reveals Welfare”.

**Axiom 1**: (CRW) Every menu \( A \) possesses an \( x \in A \) with the property that:

- \( \{x\} \sim A \).
- If \( A' \subseteq A \) with \( x \in A \), then \( A' \succeq A \).

The element \( x \in A \) is a putative choice from the menu \( A \) and has two properties that reveal that it is chosen from the menu. First, it satisfies the welfare criterion of feasibility. Second, welfare from the choice problem \( A \) in which \( x \) is allegedly chosen is the same as it would be if the DM had only \( x \) to choose from, i.e. \( x \) is a welfare-equivalent for the choice problem. We parse the axiom as follows: *every menu \( A \) possesses a welfare-equivalent choice \( x \) and this choice is revealed via the welfare criterion*. It is also straightforward to see that CRW characterizes the class of models defined by (*), i.e. \( U(A) = \max_{x \in \Theta(A)} u(x) \).

\(^{15}\)Necessity is obvious. For sufficiency we just need to define a contraction-consistent map \( \Theta(\cdot) \).
There are good reasons to refine this model. First, the contraction consistency condition does not – by itself – guarantee that the set \( \Theta(A) \) equals elements revealed to be feasible under the welfare criterion. That is, putting \( \succeq (u, \Theta) \) equal to the menu preference induced by the pair \( (u, \Theta) \), we want \( \Theta(A) \) to be exactly those elements in \( A \) which satisfy the welfare criterion w.r.t \( \succeq (u, \Theta) \),

\[
\Theta(A) = \{ x \in A : A' \succeq (u, \Theta) \{ x \}, \forall A' \subseteq A \text{ s.t. } x \in A' \}
\]

Second, without any further structure the map \( \Theta(\cdot) \) is a black box that doesn’t tell us why an object is designated as feasible or infeasible.\(^{16}\) We show (Theorem 2) that the CRW axiom (coupled with order) characterizes a generalization of the category model, called the local category model, which refines the class (\( \ast \)) to address these two issues. This result also vindicates our focus on the category model as it shows that these models form a maximal class of models with the property that the welfare criterion characterizes subjective feasibility. In addition to CRW, we require one more axiom for our other characterization, which is a representation result for the (global) category model \( (u, C) \). To state this axiom, let \( A^* \) denote the subset of subjectively feasible elements in a menu \( A \). From Definition 1, we have \( A^* := \{ x \in A : \forall A' \subseteq A \text{ s.t. } x \in A', A' \succeq \{ x \} \} \).

Axiom 2: (Strong Reduction) If \( A^* \subseteq A' \subseteq A \), then \( A' \sim A \).

The axiom combines two separate hypotheses. First, the set of feasible elements \( A^* \) determines the indifference class of the menu. Second, if there is a subset \( A' \) sandwiched in between, then everything feasible in \( A \) is feasible in the subset \( A' \). Moreover – and this is the key part – nothing in \( A' \) is feasible that was not already feasible in \( A \), i.e. \( (A')^* = A^* \). The first part of this (the second hypothesis) is common to the CRW axiom. Since it has now appeared twice, let us refer to it as a contraction consistency condition on feasibility. That is, if \( A' \subseteq A \) and \( x \) is feasible in \( A \), then \( x \) remains feasible in the contracted menu \( A' \). The second condition can be thought of as a restricted expansion consistency condition: if the contracted menu \( A' \) has the property that it contains all the feasible elements of \( A \), then anything that is feasible in \( A' \), i.e. \( x \in (A')^* \), remains feasible in \( A \).

Axiom 1 (CRW) and Axiom 2 (Strong Reduction) are the two main axioms in the paper. In later sections we will refine the category model, which will involve such that the pair \( (u, \Theta) \) generates \( \succeq \) via the equation (\( \ast \)). Let \( \Sigma(A) \) denote the set of welfare-equivalents which satisfy the welfare criterion, i.e. \( \Sigma(A) = \{ x \in A : (i) \{ x \} \sim A, (ii) \forall A' \subseteq A, x \in A' \Rightarrow A' \succeq \{ x \} \} \). Put \( \Theta(A) := \cup_{B : B \supseteq A} (\Sigma(B) \cap A) \) and note that (i) \( \Theta(\cdot) \) is contraction consistent and (ii) the pair \( (u, \Theta) \) represents \( \succeq \).

\(^{16}\)In contrast to (\( \ast \)), the model of Lleras et al. (2008) and Masatlioglu et al. (2012) is defined by the property \( \Theta(A) = \Theta(A \setminus x) \), if \( x \notin \Theta(A) \). This model also has the issue that the set \( \Theta(A) \) is not the subset of subjectively feasible elements under the induced menu preference \( \succeq (u, \Theta) \).

17
strengthened versions of these axioms. However, the main model is characterized by these two axioms alone (in addition to order). The axioms distill two behavioral principles. CRW expresses the first: *choices determine welfare*. The two requirements defining this axiom are the revealed preference indicators that the welfare equivalent of the menu is actually the element that is chosen from the menu. The Strong Reduction axiom expresses a second principle: *the set of feasible elements in a menu determines the value of the menu*. Note that the concept of temptation barely receives mention in this two axioms.\textsuperscript{17} Our interpretation of both axioms is based on the notion of subjective feasibility which, by itself, makes no mention of temptation per se. Nevertheless, temptation is visibly present in the model via the “min” operator. We now present the two utility representations which show that this structure is characterized by the two axioms introduced in this section.

### 2.3 Main Results

We first present a lemma which captures a key recursive feature of the construction used to prove the main representation theorem. The intuition behind the lemma is that (in)feasibility is a transitive concept. To express this formally first introduce an axiom which is a weakening of Strong Reduction. Introduce notation which is shorthand for the statement that \( x \) is infeasible in the menu \( \{x, y\} \). Say that \( x \rightarrow_t y \) if \( \{x\} \succ \{y\} \) and \( \{x, y\} \sim \{y\} \). This will be used to concisely express the axiom below and at several other points in the paper. To distinguish this from the notion of an attraction set (defined below) we will refer to the \( y \) such that \( x \rightarrow_t y \) as a “singleton temptation” (or, equivalently, a singleton attraction).

**Axiom 2\textsuperscript{*}:** (Reduction) If \( x \rightarrow_t y \) and \( y \in A \), then \( A \sim A \setminus x \).

The interpretation is that if \( x \rightarrow_t y \), then \( x \) is not feasible in \( \{x, y\} \) and, hence, is not feasible in \( A \) either, so this can be thought of as an “expansion consistency” condition regarding the concept of feasibility. Following the intuition of Strong Reduction, this means that the set of subjectively feasible elements is a subset of \( A \setminus x \), implying indifference between \( A \) and \( A \setminus x \). It is straightforward to check that Strong Reduction implies Reduction. Moreover, if preferences satisfy Reduction, then they satisfy the following “transitivity of infeasibility” property: if \( x \rightarrow_t y \) and \( y \rightarrow_t z \), then \( x \rightarrow_t z \).\textsuperscript{18} This property generalizes to non-doubleton menus.

**Definition 2.** Call a menu \( A(x) \) an *attraction set* for \( x \) if the following three properties hold:

\textsuperscript{17}It is lurking in the background of the CRW axiom, which itself implies a well-known axiom in the temptation literature: Positive Set-Betweenness. See the proof of either of the main representations for further discussion on this point.

\textsuperscript{18}Proof: If \( x \rightarrow_t y, y \rightarrow_t z \), then by Reduction we have (putting \( A = \{x, y, z\} \) \( \{y, z\} = A \setminus x \sim A \) and \( A \sim A \setminus y = \{x, z\} \). Since \( \{y, z\} \sim \{z\} \), we obtain \( \{x, z\} \sim \{z\} \), so that \( x \rightarrow_t z \).
1. \( \{x\} \succ \{y\}, \forall y \in A(x) \).
2. \( \{x\} \succ \{x\} \cup A(x) \).
3. \( \{x\} \sim \{x\} \cup (A(x) \setminus y), \forall y \in A(x). \)

The first two conditions say that \( x \) is not subjectively feasible in the menu \( A(x) \cup \{x\} \). The last condition says that \( A(x) \) is a minimal (w.r.t. set inclusion) set at which \( x \) is not subjectively feasible. When preferences satisfy Strong Reduction (Axiom 2), transitive infeasibility still holds when we replace doubleton menus \( \{x, y\} \) with attraction sets for \( x \). We summarize the fact that \( A(x) \) is an attraction set for \( x \) with the notation \( x \rightarrow t A(x) \). We then have the following generalization of the preceding observation.

**Lemma 1.** Assume \( \succeq \in \mathcal{P}(X) \) satisfies CRW and Strong Reduction. If \( x \rightarrow_t y \) and \( y \rightarrow_t A(y) \), then \( \{x\} \succ \{x\} \cup A(y) \).

We can iterate the above argument, so that if we have any chain \( \{x\} \succ \{x\} \cup A(x) \) and \( y \rightarrow_t A(y) \) for some \( y \in A(x) \), then we have \( \{x\} \succ \{x\} \cup A(y) \cup (A(x) \setminus y) \).

Similarly, we can inductively extend this result to arbitrary chains of singleton temptations/attraction sets. The lemma plays an important part in the proof of the following theorem.

**Theorem 1.** A preference \( \succeq \in \mathcal{P}(X) \) satisfies Axiom 1 (CRW) and Axiom 2 (Strong Reduction) if and only if it admits a category representation.

Let us sketch a proof of this result. The construction of the categories which represent a given \( \succeq \) relies on an object that we call an “\( x \)-tree”. We use the preceding lemma to provide an intuition for the construction. The construction of the categories is recursive and (implicitly) invokes the lemma at each level of the recursion. The moral of the lemma can be expressed simply as “chains of infeasibility are transitive”. Hence, if we have \( x \rightarrow_t y \) and \( y \rightarrow_t A(y) \). Then, \( A(y) \) is also an attraction set for \( x \). Similarly, if we have \( x \rightarrow_t y, y \rightarrow_t A(y) \) and \( z \rightarrow_t z' \) for some \( z \in A(y) \), then we have that \( (A(y) \setminus z) \cup \{z'\} \) is also an attraction set for \( y \), implying that it is also an attraction set for \( x \).

This suggests how to create a category model that represents the menu preference. We create a “tree” whose nodes are subsets of \( X \) and with the element \( x \) at the (top) root of the tree, call this an \( x \)-tree. Take any attraction set \( A_1(x) \) for \( x \) and at the first level of the \( x \)-tree (i.e. the branches immediately attached to the root of the tree) create a new node for every element of the attraction set. Whereas

---

\(^{19}\)Proof: Note that \( x, y \notin (\{x\} \cup \{y\} \cup A(y) \cup A(x) \)^* \), so that \( (\{x\} \cup \{y\} \cup A(y) \cup A(x) \)^* \subseteq A(y) \cup (A(x) \setminus y) \). Strong Reduction then implies \( \{x\} \cup (A(y) \cup (A(x) \setminus y)) \sim A(y) \cup (A(x) \setminus y) \), so that \( \{x\} \succ \{x\} \cup A(y) \cup (A(x) \setminus y) \).
the root node denotes the singleton \( \{ x \} \), each of the new nodes denotes a doubleton \( \{ x, x_1 \} \) for some \( x_1 \) in the attraction set \( A_1(x) \). Now consider another attraction set \( A_2(x) \) and repeat the previous part. That is, for each node \( \{ x, x_1 \} \) create \( |A_2(x)| \) branches. The new nodes at the ends of these branches denote triples of elements, \( \{ x, x_1, x_2 \} \), where \( x_2 \) is in the attraction set \( A_2(x) \). Keep adding branches and new nodes until we exhaust all \( x \)-attraction sets. Now for each node at the end of this process pick any \( z (\neq x) \) in the node and repeat this procedure.

Formally, this is where the recursive property of the construction enters. We construct \( x \)-trees by inducting upwards on the singleton ranking and recursively re-construct \( x \)-trees for higher singletons by attaching (to each node) \( x \)-trees for the lower singletons. This is the formal argument, but the informal logic for why these should yield a representation is the preceding lemma. Namely, elements in nodes of \( z \)-trees, where \( z \) is an element of an \( x \)-attraction set, are themselves elements of \( x \)-attraction sets. Hence, they potentially constrain the DM from choosing \( x \) and should be contained in any category that contains \( x \). This describes the construction in a nutshell.

Axiom 1 suggests that choices satisfy a weakened form of contraction consistency, where choices in supersets remain feasible (but not necessarily chosen) in contracted subsets. Axiom 2 adds to this by requiring that feasible elements in contracted sets agree with those in the superset, conditional on the contracted set containing all originally feasible elements. This can be thought of as an expansion consistency condition on the set of feasible elements. Since the category model requires both axioms, it conflates together these consistency conditions. This makes it difficult to disentangle the separate restrictions that the axioms place on the utility model. For instance, it is of interest to know which features of the category model are implied by Axiom 1 and what features are implied by Axiom 2. Moreover, we have yet to answer the question we raised prior to the statement of Axiom 1. Namely, while the category model provides one class of models of subjective feasibility, i.e. models of the general form, \( U(A) = \max_{x \in \Theta(A)} u(x) \), in what sense is this the “right” class to examine? To answer these questions, we now present a generalization of the category model that is characterized by Axiom 1 alone.

**Definition 3.** For each menu \( A \) let \( C_A \) be a category.\(^{20}\) The set of categories \( \{ C_A \}_{A \in \mathcal{M}} \) is coherent if whenever \( A' \subseteq A \), then for each \( C'_{A'} \in C_{A'} \), \( \exists C'_A \in C_A \) with \( C'_{A'} \subseteq C'_A \).

To understand the coherence condition, fix a menu of options \( A' \) and think of the collection \( C_{A'} \) as a grouping of elements of \( A' \) by shared (temptation) characteristics. The indices of the sets \( C'_i \in C_A \) label the (subjective) set of characteristics that the

\(^{20}\) \( C_A \) denotes a collection of subsets of \( A \), \( C'_A \subseteq A \), such that \( \bigcup_i C'_i = A \).
DM is constrained by when the option set is $A'$. Now enlarge the option set to $A$. Since there may be new temptations added to the menu, the set of constraining characteristics (i.e. object attributes that the DM finds tempting) is possibly larger. Moreover, constraints from the smaller option set are still present. Thus, when he groups objects by shared characteristics, there should be still be a bin, $C'_A \in C_A$, assigned to every characteristic that the DM was constrained by when the option set was $A'$. This bin contains the original group along with any new elements that have been added to the menu that share this characteristic. To fix ideas, imagine objects are described by their color composition and that when the menu is $A' = \{x, y, z\}$ the collection of subjective characteristics is $\{\text{red, white}\}$. Let’s say that this leads to the following grouping

$$C_{\text{red}} = \{x, y\}, C_{\text{white}} = \{z\}$$

Now enlarge the option set to $A = \{x, y, z, w\}$. The collection of characteristics might also be enlarged, viz. $\{\text{red, white, blue}\}$. The following is an example of a grouping that is a coherent extension of the grouping for the menu $A'$:

$$C_{\text{red}} = \{x, y, w\}, C_{\text{white}} = \{z\}, C_{\text{blue}} = \{z, w\}$$

Imagine that the color characteristics are types of temptations, so that the presence of blue temptation $w$ creates a temptation for $z$. Note that the coherence restriction is silent on the composition of bins that correspond to new temptations. It only requires that bins used to evaluate the smaller menu are carried over to the larger menu, where possibly more elements are added.

The coherence condition is a cross-sectional restriction on the family of categories $\{C_A\}$. This adds some structure to the forthcoming representation, but not enough to imply Axiom 1. For this we have the following additional restriction, which when coupled with coherence is the behavioral content of Axiom 1.

**Definition 4.** A collection $\{C_A\}_{A \in M}$ is **downwards rigid** if for each $C'_A \in C_A$ we have $\inf(C'_A) \nsubseteq \cup_{j \neq i} C'_j$.

This condition requires some non-redundance in the collection of tempting characteristics. Namely, for each characteristic it requires that there is a feasible object in the menu that possesses that characteristic only. Consider the following condition which more transparently formalizes the notion of non-redundance:

**Non-Redundance**: $C_i \nsubseteq \cup_{j \neq i} C_j$, $\forall C_i \in C$

That is, each set $C_i$ in the collection $C$ contains a distinguished element that is not in any of the other sets in the collection. With the preceding example in mind, this requires that for each color type there is an object that is composed of
that color alone (note that the example fails this requirement). Downwards rigidity strengthens this by requiring that the distinguished element be subjectively feasible.

**Definition 5.** A collection \( \{C_A\}_{A \in M} \) is called a local category if it is coherent and downwards rigid. A menu utility \( U(\cdot) \) is a local category model if it is generated by the pair \( (u, \{C_A\}) \), i.e. \( U(A) = \max_{C_A \in C_A} \min_{x \in C_A} u(x) \).

The following is our representation for the local category model.

**Theorem 2.** A preference \( \succeq \in P(X) \) satisfies Axiom 1 (CRW) if and only if it admits a local category representation.

The requirements of coherence and (downwards) rigidity are both required in order to imply Axiom 1. Without either one, one can construct an example of a local category that does not satisfy the CRW axiom. Finally, let us connect the local category model with the utility defined in (\( \ast \)) earlier in the section. We want to first show that the local category model is in the class of models of the form, (\( \ast \)) \( U(A) = \max_{x \in \Theta(A)} u(x) \). Recall that the set \( \Theta(A) \) denotes the subjective domain of utility maximization when the objective option set is \( A \). We would like this map to have two properties: (i) \( \Theta(B) \cap A \subseteq \Theta(A) \) (Sen’s \( \alpha \)) and (ii) \( \Theta(A) \) equals the set of elements satisfying the welfare criterion under the induced menu preference \( \succeq_{(u,\{C_A\})} \). To each local category model \( (u, C_A) \) we have an induced “feasibility” map \( \Theta(\cdot) \), where \( \Theta(A) = \bigcup_i \inf(C_i^A) \). All of these maps will possess the Sen’s \( \alpha \) property, but not all will possess the second property. To formalize this property we consider the following refinement on the class of local category models \( (u, C_A) \) which represent \( \succeq \).

**Definition 6.** A local category \( (u, \{C_A\}) \) is reflexive if the induced map \( \Theta_{(u,C_A)}(\cdot) \) satisfies:

1. If \( A \subseteq B \) and \( x \in \Theta(B) \cap A \), then \( U(A) \geq u(x) \), and
2. If \( A \succeq_{(u,C_A)} \{x\}, \forall A \subseteq B \) s.t. \( x \in A \), then \( x \in \Theta(B) \).

We now ask for a characterization of reflexive local category models \( (u, \{C_A\}) \). However, it turns out this task has already been completed. The model we construct in our sufficiency argument for Theorem 2 is reflexive (in the proof of the theorem we check this after presenting the construction). The representation result says that CRW (in addition to order) is equivalent to having a reflexive local category representation. We now claim that the class of category models is a maximal class of models of the form (\( \ast \ast \)) \( U(A) = \max_{x \in \Theta(A)} u(x) \), where \( \Theta(A) \) is exactly the set of subjectively feasible elements (w.r.t the induced menu preference \( \succeq_{(u,\Theta)} \)). The footnote below checks that preferences that admit a representation (\( \ast \ast \)) are a subclass of preferences which admit a representation of the form (\( \ast \)). Let \( \Sigma_r, \Sigma_{\ast \ast}, \Sigma_{(u,C_A)} \)
denote (resp.) menu preferences which admit representations of the form \((\ast), (\ast\ast)\), and \((u, C_A)\). Note that we have the following containments 21,

\[\Sigma_* \supseteq \Sigma_{\ast\ast} \supseteq \Sigma_{(u, C_A)}\].

Axiom CRW is implied by models in all three classes. Theorem 2 shows that the reflexive local category model, which is more specialized than the utilities in \((\ast)\) and \((\ast\ast)\), is observationally equivalent to these models. In this sense, the local category is a maximal class of models in which subjective feasibility is characterized by the welfare criterion.

2.4 Identification

It is possible to obtain an identification result for the category model, but to do so we need to refine the class of category models we consider. First, note that the category model does not have a non-redundance restriction, e.g. there can be duplicate sets in the category. This introduces a trivial source of multiplicity in the set of models \((u, C)\) which represent a given preference \(\succsim\). For example, let \((u, C)\) be a non-redundant category representation of some preference \(\succsim\). Let \(x_*\) be a \(\succsim\)-minimal singleton and consider the following category. Let \(C' := \{C, \{x_*\}\}\). That is, we have simply added the singleton set \(\{x_*\}\) to the original collection \(C\). It is easy to see that if the original pair \((u, C)\) represents \(\succsim\), then the pair \((u, C')\) also represents \(\succsim\). Moreover, the category \(C'\) is obviously redundant. A second source of trivial multiplicities in the set of representations comes from ties in the singleton ranking. For example, say \(\{x\} \sim \{x'\}\) and put \(A(x) = \{y, z\} = A(x')\), so that \(x, x'\) share a common attraction set. Consider the pair of category models, \((u, C(1))\) where \(C(1) \equiv \{x, x', y, z\}\), and \((u, C(2))\) where \(C(2) \equiv \{C_1(2), C_2(2)\}\), \(C_1(2) = \{x, y, z\}\), \(C_2(2) = \{x', y, z\}\). Note that both models are (i) non-redundant and (ii) represent the same menu preference. To preclude representations which are duplicates as in this example, for the identification result (only) we will restrict attention to menu preferences in which the singleton ranking is strict. Towards identification, we now restrict attention to category models \((u, C)\) that are sharp, i.e. every set \(C_i \in C\) is necessary for the comparison of some pair of menus \((A, B)\).

**Definition 7.** A model \((u, C)\) is sharp if for any sub-collection \(C' \subseteq C\), the induced menu preferences \(\succeq_{(u, C)}\) and \(\succeq_{(u, C')}\) are unequal. 22

Readers familiar with Dekel et al. (2001) will note that this is analogous to the notion of a “relevant state”, which is a necessary refinement in their identification

---

21 The inclusion \(\Sigma_{\ast\ast} \supseteq \Sigma_{(u, C_A)}\) is obvious. To check \(\Sigma_* \supseteq \Sigma_{\ast\ast}\), let \((u, \Theta)\) denote a model of class \((\ast\ast)\) and put \(A \subseteq B, x \in \Theta(B) \cap A\). By hypothesis, \(\forall B' \subseteq B\) with \(x \in B'\) we have \(B' \succeq \{x\}\). Hence, \(\forall A' \subseteq A, x \in A'\) we have \(A' \succeq \{x\}\), implying \(x \in \Theta(A)\). This implies \(\Theta(\cdot)\) satisfies Sen’s \(\alpha\), so that \((u, \Theta)\) is of class \((\ast)\).

22 The sub-collection satisfies \(\cup_i C'_i = X\), else the statement that \(\succeq \neq \succeq'\) is obvious.
result as well. Our identification result for the category model shows that when we restrict attention to sharp category models \((u, \mathcal{C})\), then category representations can be weakly identified in the following sense.

**Definition 8.** Given two models \((u, \mathcal{C}_1)\), \((u, \mathcal{C}_2)\) say that \((u, \mathcal{C}_2)\) is a *prolongation* of \((u, \mathcal{C}_1)\) if there is a bijection \(\pi : \mathcal{C}_1 \rightarrow \mathcal{C}_2\) where for every set \(\mathcal{C}_1(i) \in \mathcal{C}_1\) we have \(\pi(\mathcal{C}_1(i)) \supseteq \mathcal{C}_1(i)\).

The identification result for the category model is now described as follows. Let \((u, \mathcal{T})\) denote the tree-category constructed in the sufficiency proof of the theorem. We show that there is a unique sharp sub-category \((u, \mathcal{C}_*)\) of the tree category with the property that *every* sharp category model that represents the menu preference is a (specific) prolongation of the model \((u, \mathcal{C}_*)\). To get a handle on the concepts of prolongation and sharpness we illustrate with an example. A warning to the reader: the forthcoming discussion uses notation from the proof of Theorem 1.

**Example 3.** Let \(X = \{x, y, z, p, q\}\) where \(x\) is attracted to \(\{y, z\}\), \(\{p, q\}\), \(\{y, q\}\), and \(\{p, z\}\). Also put \(y \rightarrow_t p\) and \(z \rightarrow_t q\), and assume there are no other constraints on choice. Consider the following category,

\[
\mathcal{C}_1 := \{x, y, p\}, \mathcal{C}_2 := \{x, z, q\}
\]

where \(\mathcal{C}_* \equiv \{\mathcal{C}_1, \mathcal{C}_2\}\). Notice that the pair \((u, \mathcal{C}_*)\) represents \(\succeq\). Now compare this with the category we obtain from the tree algorithm in the theorem. It is straightforward to see that the only relevant categories are those in the \(x\)-tree.

There are 4 levels in the \(x\)-tree corresponding to the four \(x\)-attraction sets, \(A_1(x) = \{y, z\}, A_2(x) = \{p, q\}, A_3(x) = \{y, q\}, A_4(x) = \{p, z\}\). A list of the categories on each level is as follows: (put \(B_t(x) := \{y \in X : x \rightarrow_t y\}\))

1. \(\mathcal{L}(0) \equiv \{\mathcal{C}_0^1 = \{x\}\}\) (since \(B_t(x) = \emptyset\), the root node of the \(x\)-tree contains only the singleton \(\{x\}\)).

2. \(\mathcal{L}(1) \equiv \{\mathcal{C}_1^1 = \{x, y, p\}, \mathcal{C}_1^2 = \{x, z, q\}\}\) (the two nodes are obtained by separately attaching \(B_t(y) \cup \{y\} = \{y, p\}\) and then \(B_t(z) \cup \{z\} = \{z, q\}\) to the initial node of the tree).

3. \(\mathcal{L}(2) \equiv \{\mathcal{C}_2^1 = \{x, y, p, q\}, \mathcal{C}_2^2 = \{x, z, p, q\}, \mathcal{C}_3^2 = \{x, z, q\}\}\) (separately attach \(p\) and \(q\) to each node in the preceding level – note that \(B_t(p) = \emptyset = B_t(q))\).

4. \(\mathcal{L}(3) \equiv \{\mathcal{C}_3^1 = \{x, y, p, q\}, \mathcal{C}_3^2 = \{x, y, p, q\}, \mathcal{C}_3^3 = \{x, y, p, q\}, \mathcal{C}_4^3 = \{x, y, p, q\}\}\) (separately attach \(B_t(y) \cup \{y\}\) and (resp.) \(\{q\}\) to each node in the preceding level – note that we include replicas in the tree construction).
5. $L(4) \equiv \{C_4^1 = \{x, y, p\}, C_4^2 = \{x, y, p, z, q\}, C_4^3 = \{x, y, p, z, q\}, C_4^4 = \{x, y, p, z, q\}, C_4^5 = \{x, y, p, q\}, C_4^6 = \{x, y, p, z, q\}, \ldots, C_4^{15} = \{x, z, p, q\}, C_4^{16} = \{x, z, q\}\}$ (the pattern at this stage should be clear, hence we omit the full enumeration of all 16 nodes in level 4 of the tree).

The example demonstrates that the sufficiency construction of the theorem does not typically yield a sharp category. In fact, far from it! Notice that several of the nodes in the terminal level of the $x$-tree are identical. However, notice also that the category model we started out with, $(u, C_*)$, is a sub-category of the terminal nodes of the $x$-tree. In particular, the sets in $C_*$ are exactly the minimal (w.r.t set inclusion) nodes in the $x$-tree. This points us to the general proof strategy for the uniqueness claim. We proceed in two main steps. Fix any category representation $(u, C)$. The first step is to use the concepts of prolongation and sharpness to realize this model as a prolongation of some sub-category $(u, C_*)$ of the tree category. This shows that there is a minimal sharp retraction of the model $(u, C)$ that sits inside the tree category, we call this the embedding step. The second and more complicated step, called pruning, shows that there is a unique sharp sub-category of the tree category. The proof is in the appendix. Given a pair of sets $(A, B)$ where $A \subseteq B$ (both contained in $X$) say that $A$ is a lower bound order interval in $B$ if there is some $z \in B$ such that $(-\infty, z] \cap B := \{x \in B : z \geq x\} = A$.

**Theorem 3** (Identification). Assume $(\succeq)|_X$ is strict. There is a unique, sharp model $(u, C_*)$ such that any other sharp representation, $(u, C)$, is a prolongation of $(u, C_*)$. Moreover, any prolongation $\pi : C_* \to C$ has the property that $C_i \in C_*$ is a lower bound order interval in $\pi(C_i)$.

The content of the model is in the sets $C_i$ comprising the category. The indices of these sets label the subjective attributes (constraints) on choice and a given set $C_i$ groups elements by a shared attribute. The identification result says that, controlling for sharpness – so that only relevant constraints are allowed in the model, the identity (i.e. the attribute labels) of the constraints is pinned down. Moreover, the content of the sets $C_i$ that is relevant for the representation is also pinned down: Given two models, $(u, C), (u, C_*)$ where $C$ prolongs $C_*$, each set $C'_i$ contains a unique set $C_{j_i} \in C_*$ as a lower bound order interval. Since $(u, C_*)$ represents the menu preference this means that the utility function on menus generated by the pair $(u, C)$ only lives on (i.e. for any set the maximum only occurs on the subset $C_{j_i}$) each of these lower bound order intervals. In this sense, both the attributes and the set of elements, $C_i$, grouped by this attribute are identified.

---

23 We say that a model $(u, C_*)$ is a retraction of $(u, C)$ if $C$ is a prolongation $C_*$, i.e. the domain of the bijection $\pi$ is the retracted category, via $\pi^{-1}$, and the range is the prolongation.

24 The unique, minimal category $C_*$ is constructed explicitly in the proof and is a sharp subcategory of the tree category, $(u, T)$, constructed in the proof of Theorem 1.
Both sharpness and prolongation are crucial for our proofs of the results, but this, nevertheless, begs the question: are they both necessary? Is it possible to obtain a stronger form of identification without the prolongation refinement (we have already explained why sharpness is necessary)? The following example shows that the answer to this question is negative, so that the identification we provide is the strongest form available.

**Example 4** (Sharpness \( \not\Rightarrow \) Uniqueness). Let \( X = \{ x \succ y \succ z \succ p \succ q \} \). Assume \( A_1(x) = \{ z, p \} \), \( A_2(x) = \{ z, q \} \), and \( p \rightarrow_t q, y \rightarrow_t p \). Put \( C_1 = \{ y, p, q \} \), \( C_2 = \{ x, z \} \), \( C_3 = \{ x, p, q \} \) and consider the category model \( (u, C) \) given by \( C \equiv \{ C_1, C_2, C_3 \} \). Note that \( (u, C) \) is a sharp representation of \( \succeq \). Now consider the following prolongation of \( C \). Put \( C'_1 = \{ x, y, p, q \} \), \( C'_2 = C_2 \), \( C'_3 = C_3 \) and let \( C' \equiv \{ C'_1, C'_2, C'_3 \} \).

Note that the model \( (u, C') \) also represents \( \succeq \) and, importantly, is also sharp. To see this, note that since \( y \) is not a member of any attraction set for \( x \) we cannot omit \( C'_3 \) from the representation. Thus, we need the prolongation refinement in order to pin down the model. Also note that the model \( (u, C') \) prolongs \( (u, C) \) in the manner described by the theorem: the sets comprising \( C \) are lower bound order intervals of the sets comprising \( C' \). This example shows that in order to model attraction effects we cannot use categories whose underlying menu preference can be pinned down on the basis of sharpness alone. Hence, there is a behavioral restriction on the menu preference implied by the fact that its category representation is identifiable based on the sharpness criterion alone. There is more to say on this issue and we will return to it when we revisit the question of identification in the next section, in which we compare the category model with existing models of temptation and self-control.

### 3 Comparison between Models

In this section we study the connections between the seminal Gul and Pesendorfer (2001) (GP) model, the Strotz model, and the category model.\(^2^5\) GP develop a model of costly self-control. Choices are distorted from (normative) preference maximization because temptation makes it costly to select the maximizer of normative preference. Consequently, an important feature of their model is that welfare cannot be recovered from choices alone. One needs the full observability of the menu preference to recover both the normative preference and the self-control cost function which distorts choice from maximization of this preference. At the other extreme of the GP model, we have the Strotz model of “no self-control”. In this model the DM gives in completely to temptation and selects according the the \( v \)-function alone (where \( (u, v) \) are the pair of functions in a GP representation – see

\(^{25}\)Recall that the Strotz model is defined by a pair of functions \( (u, v) \) assembled into a menu utility via \( U(A) = \max_{z \in A_v} u(x) \), where \( A_v := \arg \max_{z \in A} v(z) \).
below). Here welfare can be read off from choice, although menu preferences are still non-standard. The connections with the Strotz model are significant, so we begin here.

### 3.1 Comparison with the Strotz model

The category model nests the Strotz model. Moreover, there is an interesting connection between the axioms which characterize the category model and the ones which characterize the Strotz model. This is surprising because the Strotz model is a dual-self model, where the normative self (represented by $u$) maximizes his preference on the residual subset of maximizers of a second stage self’s (represented by $v$) choice. In contrast, the category model involves only a single preference relation over choices, represented by the cardinal $u$. Nevertheless, the nesting shows that there is a way to map the dual-self model into a model of (constrained) maximization by a single self. One may think of the residual set $A_v$ (consisting of the $v$-self’s arg max) as the constraint set, but the actual embedding of the Strotz model into the category model is more subtle. For instance, given a Strotzian pair $(u, v)$ we need to construct a category $C \equiv \{C_i\}$ such that $A_v = \bigcup_{i: C_i \in C} \inf(C_i \cap A)$.

Once we realize the Strotz model as a subclass of category models, we then analyze this subclass to better understand where it is situated within the constellation of category models. This takes us to a representation for a subclass of category models in which the only constraints on choice are due to compromise effects or “Strotzian” choices, i.e. this subclass precludes attraction effects.

**Definition 9.** An element $x \in A$ is a *Strotzian choice* if it satisfies the following two properties:

- $\{x\} \sim A$.
- If $A' \subseteq A$ and $x \in A'$, then $\{x\} \sim A'$.

Note that the first condition says that every menu possesses a welfare-equivalent, i.e. an option such that value of the menu equals the value of the option. The second condition says that if a sub-menu contains the original choice, then the DM is indifferent between the two menus, implying that $x$ remains a choice in the submenu. Contrast this with the CRW axiom, which stated that $A'$ is weakly preferred to $A$. An interpretation of this property is that feasibility satisfies a “contraction consistency” condition. Actually, the second part of the CRW axiom makes a slightly different claim. It says that if $x$ is a choice in $A$ (hence, feasible), then when we contract to $A'$ the choice remains feasible even though it may not be chosen. In other words, CRW implies a restricted form of contraction consistency, where choices contract to feasible elements but not necessarily to choices. The definition
given above strengthens this condition by saying that choices contract to choices, not just to feasible elements. The following axiom says that every menu contains such a choice, with the set of such elements denoted as $\Sigma_{ST}(A)$ for Strotzian choices.

**Axiom 3:** (Strong CRW) If $A \neq \emptyset$, then $\Sigma_{ST}(A) \neq \emptyset$.

**Axiom 3:** (DSB) For any menus $A, B$ either $A \cup B \sim A$ or $A \cup B \sim B$.

The latter restriction is referred to as “Degenerate Set-Betweenness” (DSB), after the well-known axiom in Gul and Pesendorfer (2001). To explain why we have given the axioms identical labels note that the first axiom clearly implies the latter. Moreover, the converse is true.

**Lemma 2.** DSB and Strong CRW are equivalent.

**Definition 10.** A menu preference $\succeq \in P(X)$ is a Strotz preference if it admits a representation: $U(A) := \max_{x \in A_v} u(x)$, where $A_v = \arg \max_{x \in A_v} v(x)$.

In Gul and Pesendorfer (2001) (Lemma 6 of their paper), it is shown that the Strotz model is characterized by order and DSB. Hence, the lemma above gives an equivalent characterization of the Strotz model which makes a transparent (behavioral) link with the local category model. This just means that menu preferences that admit a Strotz representation also admit an interpretation via the category model. It doesn’t, by itself, imply a connection between the functional form of the Strotz model and the category model. We establish this now.

Given a Strotz pair $(u, v)$, we now back out an explicit category $C_{(u,v)}$ where the model $(u, C_{(u,v)})$ generates the same utility function on menus as the pair $(u, v)$. That is, $\max_{x \in A_v} u(x) = U(A) = \max_{C_i \in C_{(u,v)}} \min_{x \in A \cap C_i} u(x)$. Fix $(u, v)$ and define the following sets

$$C_x := \{y \in X : v(y) > v(x), u(x) \geq u(y)\} \cup \{x\}$$

and let $C_{(u,v)} := \{C_x : x \in X\}$.

**Observation 1.** Fix a Strotz pair $(u, v)$ and a category model $(u, C_{(u,v)})$. These two pairs generate the same utility function on menus.

This fact is, perhaps, surprising but in and of itself not terribly useful. It just tells us that there is an embedding of the Strotz utility into the space of category models, but it doesn’t tell us anything more concrete than that. The model $(u, C)$ generates the same utility function on menus as the pair $(u, v)$. To this end, we seek general restrictions on the category $C$ that characterize Strotzian preferences. In other words, we find the subclass of models $(u, C_{ST})$ such that the associated menu
preference is Strotzian.

To answer this question, we consider three families of models, labeled \((u, \mathcal{C}_{\text{ST}})\), \((u, \mathcal{C}_{\text{nw}})\), and \((u, \mathcal{C}_{\text{NAT}})\). The first is the Strotzian class, the second is what we call the “narrow” category model, and the third is the “no attractions” category. The Strotz and narrow categories correspond to two of the three parts of the behavioral trichotomy described in the introduction. In particular, the narrow category captures the compromise effect alone, the Strotz category captures Strozian choice, and the NAT category is the maximal model that captures both Strozian choice and the compromise effect while precluding cycles due to the attraction effect.\(^{26}\) For brevity, let \((x, \infty)\) denote the order interval in \(X\), \(\{z \in X : \{z\} \succ \{x\}\}\). Similarly denote \((\infty, x), (\infty, x]\), and so on.

**Definition 11.** A pair \((u, \mathcal{C}_{\text{ST}})\) is a Strotzian category if the collection \(\mathcal{C}_{\text{ST}} \equiv \{\mathcal{C}_i\}\) satisfies the following three conditions:

1. \(\bigcup_i \mathcal{C}_i = X\).
2. Fix \(\{x\} \succ \{y\}\). If \(x \in \mathcal{C}_i, y \in \mathcal{C}_j \setminus \mathcal{C}_i\), then \((-\infty, y) \cap \mathcal{C}_i \subseteq \mathcal{C}_j\).
3. For each \(x \in X\), \(\exists \mathcal{C}_x \in \mathcal{C}\) s.t. \((-\infty, x) \cap \mathcal{C}_x \subseteq \mathcal{C}_i\), \(\forall \mathcal{C}_i\) s.t. \(x \in \mathcal{C}_i\).\(^{27}\)

**Definition 12.** A pair \((u, \mathcal{C}_{\text{nw}})\) is a narrow category if the collection \(\mathcal{C}_{\text{nw}} \equiv \{\mathcal{C}_i\}\) satisfies the following two conditions:

1. \(\bigcup_i \mathcal{C}_i = X\).
2. If \(x \in \mathcal{C}_i \cap \mathcal{C}_j\), then \((-\infty, x) \cap \mathcal{C}_i = (-\infty, x) \cap \mathcal{C}_j\).

**Definition 13.** A pair \((u, \mathcal{C}_{\text{NAT}})\) is a no attractions category if the collection \(\mathcal{C}_{\text{NAT}} \equiv \{\mathcal{C}_i\}\) satisfies the following two conditions:

1. \(\bigcup_i \mathcal{C}_i = X\).
2. For each \(x \in X\), \(\exists \mathcal{C}_x \in \mathcal{C}\) s.t. \((-\infty, x) \cap \mathcal{C}_x \subseteq \mathcal{C}_i\), \(\forall \mathcal{C}_i\) s.t. \(x \in \mathcal{C}_i\).

To state the axioms which characterize these models, we relax the notion of subjective feasibility. Recall that subjective feasibility requires, informally, that \(x\) is always “choosable” whenever it is available (in any sub-menu of a given menu \(A\)). The following definition relaxes this by requiring only that \(x\) is choosable from any binary sub-menu of \(A\).

\(^{26}\)There is some trivial overlap between the models. For example, part (b) of Strotzian choice in the taxonomy in section 1 can be represented by both the narrow and Strotz categories. Since \(z\) is irrelevant to the choice pattern in this case (note it is never chosen in any doubleton selected from \(\{x, y, z\}\)) the preference boils down to a non-standard ranking on \(\{x, y\}\), which is a less interesting phenomenon.

\(^{27}\)When there are no ties in the singleton ranking, property (3) is implied by the second property. We prove this in the appendix, when we prove the representation for the Strotzian category.
Definition 14. An element $x \in A$ is pairwise feasible in $A$ if $\{x, y\} \succeq \{x\}$, $\forall y \in A$. Say the menu $A$ is pairwise feasible if it consists entirely of pairwise feasible elements.

We now introduce the axioms which characterize these models. Recall the notation from section 2, where we put $x \rightarrow_t y$ if $x$ is not feasible in $\{x, y\}$. Similarly, put $x \not\rightarrow_t y$ if $x$ is feasible in $\{x, y\}$.

Axiom 4: (No Cycles) Fix $\{x\} \succ \{y\} \succ \{z\}$. If $x \rightarrow_t z, y \not\rightarrow_t z$, then $x \rightarrow_t y$.

Axiom 5: (Only Cycles) Fix $\{x\} \succ \{y\} \succ \{z\}$. If $x \rightarrow_t z, y \not\rightarrow_t z$, then $x \not\rightarrow_t y$.

Axiom 6: (No Attractions) If $A \cup B$ is pairwise feasible, then either $A \cup B \sim A$ or $A \cup B \sim B$.

The names for axioms 4 and 5 derive from an issue related to the Strotz model. Consider a triple of menus, (i) $\{x, y\}$, (ii) $\{x, z\}$, and (ii) $\{x, y, z\}$, where $\{x\} \succ \{y\} \succ \{z\}$. Imagine the DM chooses $x$ from the first menu, $z$ from the second, but $y$ from the third. Notice that these choices cannot be generated by a Strotz model. Else, if $(u, v)$ is a putative Strotz pair that generates these choices, then $v(\cdot)$ must contain a cycle – a contradiction.\(^{28}\) Holding choice of $z$ in (ii) constant, to admit a Strotzian representation we need to change either the choice in (i) or in (iii). If we keep $\{x, y\} \sim \{x\}$, then the only Strotzian choice from $\{x, y, z\}$ is $\{z\}$. Similarly, if we keep $\{x, y, z\} \sim \{y\}$, then we must have $\{x, y\} \sim \{y\}$. Put another way, if $z$ is chosen from $\{x, z\}$ and $y$ is chosen from $\{x, y, z\}$ (and from $\{y, z\}$), then the only way to be Strotzian is if $y$ is chosen from $\{x, y\}$ as well. Since Axiom 4 is consistent with a Strotzian representation we call it “no cycles”.

The companion logic explains why we call axiom 5 “only cycles”. In this case, $y$ is chosen from $\{x, y, z\}$, yet $x$ is chosen from $\{x, y\}$, which forces a cycle on the $v$-function in any putative Strotz representation $(u, v)$. Note that these two cases are complementary. For a given triple $(x, y, z)$ such that $\{x\} \succ \{y\} \succ \{z\}$, if $(x \rightarrow_t y) \land (y \not\rightarrow_t z)$, then we have $(x \rightarrow_t y) \lor (x \not\rightarrow_t y)$ – so long as choices from the menu reveal welfare, i.e. the menu preference satisfies CRW, which is in place for the entire paper. Axiom 4 says that for any such triple the first alternative always holds and Axiom 5 says that for any such triple, the second alternative always holds. Finally, note that Axiom 6 is just a restricted form of the DSB axiom, where we only apply DSB to menus which are pairwise feasible. The intuition is simply that pairwise feasibility determines feasibility so that, by implication, there can be no non-singleton attraction sets.

\(^{28}\)Concretely, note $\{x, y\} \sim \{x\}$ and $\{x, z\} \sim \{z\}$ implies (resp.) $v(x) \geq v(y), v(z) > v(x)$. Similarly, $\{x, y, z\} \sim \{y\}$ implies that $v(y) \geq \max\{v(x), v(z)\}$. Hence, $v(z) > v(x) \geq v(y) \geq v(z)$ – contradiction.
Now consider the three models described above. The NAT model is described by (essentially) the single condition that: 
\[(C_i \cap (-\infty, x) \subseteq C_j) \lor (C_j \cap (-\infty, x) \subseteq C_i).\]
In other words, for each \(x\) there is a minimal (w.r.t. set inclusion) constraint set \(C_x\) containing \(x\). Note that under this model \(x\) is choosable from a menu \(A\) if and only if 
\[((-\infty, x) \cap A) \cap C_x = \emptyset.\]
Hence, the intuition for this condition is that there is only one subjective attribute that tempts the DM away from selecting \(x\), with the set \(C_x\) grouping together all elements which share this attribute. The Strotz model is a subclass of this category with the following additional property. Take any pair \((x, y)\) with \(\{x\} \succ \{y\}\). The hypothesis that \(y \in C_j \setminus C_i\) says that (since \(x \in C_i\)) \(y\) is not a constraint on \(x\), i.e. if \(x\) is feasible in the menu \(A\), then \(x\) remains feasible in the menu \(A \cup \{y\}\) (using the assumption that there are no attractions). The restriction defining the Strotzian category, i.e. 
\[(-\infty, y) \cap C_i \subseteq C_j\] says that all potential constraints for \(y\) (i.e. elements of \((-\infty, y)\)) that constrain \(x\) must then also constrain \(y\), i.e. belong to \(C_j\). Note that this is just a translation of the dual-self model into the formalism of categories. If we have a pair \((x, y)\) where \(u(x) > u(y)\) and \(y\) does not constrain \(x\) (so that \(v(x) \geq v(y)\)), then since the set of \(x\)-constraints is summarized by the \(v\)-function, via \(\{z : v(z) > v(x), u(x) > u(z)\}\), it follows that anything (in \((-\infty, y)\)) that constrains \(x\) must also constrain \(y\). We now turn to the representation results for these models.

**Theorem 4.** A preference \(\succeq \in \mathcal{P}(X)\) is representable by a NAT category model if and only if it satisfies Axiom 2\(\,^*\) (Reduction) and Axiom 6 (No Attractions). Moreover, if \((u, C)\) is any category model that represents \(\succeq\), then \(C\) is a NAT category.

The representations of the narrow and Strotzian categories build on the representation of the NAT category.

**Proposition 1.** A preference \(\succeq \in \mathcal{P}(X)\) admits a narrow category representation if and only if it satisfies Axiom 2\(\,^*\), Axiom 5 (Only Cycles), and Axiom 6.

**Proposition 2.** A preference \(\succeq \in \mathcal{P}(X)\) admits a Strotzian category representation if and only if it satisfies Axiom 2\(\,^*\), Axiom 4 (No Cycles), and Axiom 6.\(^{29}\)

We now return to the behavioral trichotomy mentioned in the introduction. Fix the welfare ranking on the triple: \(\{x\} \succ \{y\} \succ \{z\}\). Consider the following triple of choice patterns: (in all cases, assume \(C(\{y, z\}) = \{y\}\))

- **Compromise (CMP)** \(C(\{x, z\}) = \{z\}, C(\{x, y\}) = \{x\}, C(\{x, y, z\}) = \{y\}\).
- **Attraction (ATT)** \(C(\{x, z\}) = \{x\}, C(\{x, y\}) = \{x\}, C(\{x, y, z\}) = \{y\}\).
- **Strotz (STZ)** \(C(\{x, z\}) = \{z\}, C(\{x, y\}) = \{y\}, C(\{x, y, z\}) = \{y\}\).

\(^{29}\)These axioms are equivalent to Strong CRW (equiv. DSB), as shown in the proof of the Proposition.
To each of these choice patterns we assign a specific subclass of category models. Each subclass has the property that deviations from WARP occur due to that choice pattern alone. In addition to narrow and Strotzian categories, the table lists a third class – the separated category \((u,C_{sp})\). We give a definition and a representation result for this subclass in section 4. Separated categories are those where deviations from WARP are solely due to the presence of attraction effects. That is, we strain out deviations due to Strotzian choices and compromise effects. It turns out that, when we do this, we can characterize the resulting (separated) category model from observing choices from menus alone. For this reason, we postpone the definition of \((u,C_{sp})\) to a later section, in which we show an observational equivalence between – on the one hand – pairs \((u,C)\) consisting of (i) a welfare ranking and (ii) a choice correspondence and – on the other – choices between menus.

<table>
<thead>
<tr>
<th>Anomaly</th>
<th>Category Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMP</td>
<td>((u,C_{nw}))</td>
</tr>
<tr>
<td>ATT</td>
<td>((u,C_{sp}))</td>
</tr>
<tr>
<td>STZ</td>
<td>((u,C_{ST}))</td>
</tr>
</tbody>
</table>

Table 3: Behavioral Trichotomy

3.2 Identification Revisited

We turn now to identification of these category models. To this end, we recall the sets \(B_t(x) := \{y \in X : x \rightarrow_t y\}\). In words, these are the elements \(y\) which render \(x\) infeasible in the menu \(\{x,y\}\).

**Definition 15.** The set \(B_t(x) \cup \{x\}\) is said to be dominated by \(B_t(y) \cup \{y\}\) if (i) \(x \in B_t(y)\) and (ii) \(B_t(x) \cup \{x\} = B_t(y) \cap (-\infty, x]\).

Note that the notion of domination can be formalized as a well-defined, transitive relation on the collection of sets \(\{B_t(x) \cup \{x\}\}_{x \in X}\).\(^{30}\) Let \(\mathcal{B}^*\) denote the collection of maximal elements under this relation. That is, let \(\mathcal{B}^* = \{B_t(x_i) \cup \{x_i\}\}\) denote the collection of all undominated sets under this relation. We then have the following identification claim for the NAT category.

**Theorem 5** (Identification of NAT categories). Assume \((\succeq)\)|\(X\) is strict and let \((u,C)\) be a sharp NAT category representation of \(\succeq\). Then, \(C \equiv \mathcal{B}^*\).

\(^{30}\)Transitivity is obvious once we have well-definedness. For this, note that if \(x \neq y\) and we have \(B_t(x) \cup \{x\} = B_t(y) \cup \{y\}\), then \(y \in B_t(x)\), so that \(\{x\} \succ \{y\}\). But then, we cannot have \(x \in B_t(y)\). This shows that the association \(x \mapsto B_t(x) \cup \{x\}\) is a bijection from \(X\) to the collection of sets of the form \(B_t(x) \cup \{x\}\). Thus, we get a (well-defined) induced dominance relation on \(X\) which we can equivalently think of as a relation on the sets \(B_t(x) \cup \{x\}\).
The identification for the NAT category is stronger than for the full category model. As we showed by example in the previous section, we can only identify general category models up to (a specific kind of) prolongation. Hence, this identification result is distinct from and, moreover, does not build on Theorem 3. Let us conclude with a comment on the concept of sharpness. One problem with sharpness is that it allows redundancies in the category representation. For example, it is possible that, for some set $C_j$, we have $C_j \subseteq \bigcup_{i \neq j} C_i$. A natural question is whether we need to allow redundancies, in this sense, to represent the same classes of preferences. Notice that non-redundant categories are necessarily sharp. Hence, they cut out a (possibly) smaller class of menu preferences than sharp categories.

The relaxation from non-redundant categories to sharp categories is necessary to represent Strotzian preferences.\(^{31}\) Hence, to represent all Strotzian preferences with category models we need to allow that the categories are possibly redundant (but still sharp). Notice that this now gives us three increasingly stringent restrictions on the class of category models $(u, C)$:

i. Sharpness + Prolongation.

ii. Sharpness (by itself).

iii. Non-Redundance.

Sharpness plus prolongation is needed to represent all preferences with attraction effects. Sharpness alone (without prolongation) suffices to represent all NAT preferences (i.e. choice anomalies in groups (1) and (3) in Table 3), but we cannot tighten the restriction to non-redundance without simultaneously shrinking the domain of preferences that admit representation by a category model. This suggests the following question: fixing either of the restrictions (i), (ii), or (iii) what is the largest class of preferences that can be represented by categories satisfying (ii) but not (iii) and, similarly, what is the largest class that can be represented by (iii)? Moving from (i) to (ii) we lose the ability to model attraction effects. Moreover, a closer look at the example in the preceding section shows that this statement identifies the maximal class. The moment the preference exhibits an $x$ with a non-trivial attraction set, then the set of category models representing the preference are not uniquely pinned down by the sharpness criterion alone.

\(^{31}\)For an example, put $u(x_1) > u(x_2) > u(x_3); v(x_2) > v(x_3) > v(x_1)$ and let $\succeq$ denote the associated menu preference generated by the Strotz model $(u, v)$. If $(u, C)$ is a non-redundant category that represents $\succeq$, then let $C_{x_1} \in C$ be a set that contains $x_1$. Since $x_1 \rightarrow_t x_2, x_1 \rightarrow_t x_3$ we must have $x_2, x_3 \in C_{x_1}$. Non-redundance then implies that $C = \{C_{x_1}\}$. This implies $U(\{x_2, x_3\}) = u(x_3)$. OTOH, $\{x_2, x_3\} \sim \{x_2\}$, which contradicts representability.
We now consider the second part of the question, viz. pinning down the largest class of preferences that admit representation by models \((u, C)\), where \(C\) is non-redundant. Let \(\Sigma_{\text{NR}} \subseteq P(X)\) denote the sub-class of NAT menu preferences which admit representation by a non-redundant category model. It is easy to see that all preferences that are represented by narrow category models \((u, C_{\text{nw}})\) are in \(\Sigma_{\text{NR}}\). Moreover, we have seen by example that not all Strotzian preferences admit non-redundant category representations. It would interesting to find a concrete description of the class \(\Sigma_{\text{NR}}\). For example, this would tell us how much bigger the \(\Sigma_{\text{NR}}\) class is than the class of narrow categories.

### 3.3 Comparison with Costly Self-Control models

The max-min structure of the category model is shared between several models of temptation. We now clarify the connection between the category model and the Gul and Pesendorfer (2001) (GP) model of temptation and self-control. Fix a menu utility \(U : \mathcal{M} \to \mathbb{R}\) and consider the map

\[ u_{\min}(x, A) := \min\{U(A') : A' \subseteq A, x \in A'\} .\]

The number \(u_{\min}(x, A)\) is interpreted as the value of the most costly self-control problem faced by the DM when he is trying to commit to the option \(x\) and resist the temptations in the menu \(A\). For comparison, recall the GP menu utility:

\[ (*) \quad U(A) = \max_{x \in A} [u(x) - (\max_{y \in A} v(y) - v(x))] .\]

It is straightforward to verify that the formula \(\max_{x \in A} u_{\min}(x, A)\) recovers the utility on menus if and only if the underlying preference satisfies Positive Set-Betweenness.\(^{32}\)

Thus, the GP utility can be coarsely expressed as \(U(A) = \max_{x \in A} u_{\min}(x, A)\). The kernel \(u_{\min}(x, A)\) need not agree with the GP maximand \(u(x) - (\max_{y \in A} v(y) - v(x))\). It can be computed explicitly as follows. For a menu \(A\), define two subsets \(A_1(x), A_2(x)\) to be (resp.) \(A_1(x) := \{z \in A : (i) u(z) + v(z) > u(x) + v(x), (ii) u(z) \leq u(x)\}\), \(A_2(x) := \{z \in A : (i) u(z) + v(z) \leq u(x) + v(x), (ii) u(z) \leq u(x)\}\). Note that \(A_1(x)\) is the set of options in \(A\) that are (weakly) normatively worse than \(x\) yet are chosen over \(x\) head-to-head and \(A_2(x)\) consists of those normatively worse elements which lose head-to-head with \(x\). For the GP model we can check that \(u_{\min}(x, A)\) is given by:

\[ u_{\min}(x, A) = \min\{\min_{z \in A_1(x)} u(z), u(x) - (\max_{z \in A_2(x)} v(z) - v(x))\} .\]

\(^{32}\)Recall that Positive Set-Betweenness is the \((A \Rightarrow A \cup B)\) half of the Set-Betweenness axiom from Gul and Pesendorfer (2001). To prove the statement, for each \(x \in A\) let \(A_x\) denote the maximal menu (w.r.t. set inclusion) in the set \(\arg\min_{A' \subseteq A, x \in A'} U(A')\). Note that Positive Set-Betweenness implies a unique such maximal element. Now, notice that \(A = \cup_{x \in A} A_x\). By Positive Set-Betweenness, \(A_{x'} \supseteq A\) for a \(x' \in A\) such that \(A_{x'}\) is \(\succeq\)-maximal. OTOH, \(A \succeq A_{x'}\) by definition of \(A_{x'}\). Hence, \(U(A) = \max_{x \in A} u_{\min}(x, A)\).

\(^{33}\)The proof of this equality is found under Observation 2 in the appendix.
For intuition, consider the sets \( A_1(x), A_2(x) \). In the jargon of the temptation literature the first set consists of temptations (since \( u(z) \leq u(x) \) we consider these temptations) that “overwhelm” \( x \) in the sense that their presence in the menu prevents the DM from choosing \( x \). The second set, \( A_2(x) \), is the set of temptations which still allow the DM to choose \( x \), but at a (possible) cost. The formula for \( u_{\min}(x, A) \) is the (net) welfare value of the worst sub-problem of \( A \) which contains \( x \). In other words, we imagine the DM contemplates committing to \( x \) and asks: which sub-problem \( A_x \) of \( A \) makes it most difficult to maintain this commitment? The value of this sub-problem is \( u_{\min}(x, A) \). The formula for \( u_{\min}(x, A) \) is determined in two steps according to whether or not \( x \) is overwhelmed. First, consider the set of temptations which make it costly to commit to \( x \), but not overwhelming so. This is the GP value of the menu \( A_2(x) \). Now consider the temptations in \( A_1(x) \) which overwhelm \( x \). If there aren’t any, then the value of \( A_x \) coincides with \( A_2(x) \). On the other hand, if \( A_1(x) \) is non-empty then we compare the normative value of each overwhelming temptation against the most costly choice problem in which \( x \) is still chosen, i.e. \( A_2(x) \). Comparing the (normatively) worst such temptation against the value of \( A_2(x) \) yields the value \( u_{\min}(x, A) \). Note that this plainly recovers the GP function defined in (\( * \)), via \( U(A) = \max_{x \in A} u_{\min}(x, A) \).

Similarly, consider a category pair \((u, C)\) and let \( U(\cdot) \) be the utility given by the formula \( U(A) = \max_{C_i} \min_{z \in A \cap C_i} u(z) \). For each \( x \in X \), recall that \( C(x) := \{ C_i \in C : x \in C_i \} \). That is, \( C(x) \) is the sub-collection of sets in \( C \) that contain \( x \). Notice that

\[
\begin{align*}
    u_{\min}(x, A) &= \max_{C_i \in C(x)} \min_{z \in C_i \cap A} u(z).
\end{align*}
\]

Putting \( C \equiv \{ C_i \} \), we can simply reorganize the sets in \( C \) as \( \cup_{x \in X} C(x) \). It follows that

\[
\begin{align*}
    U(A) &= \max_{C_i} \min_{z \in C_i \cap A} u(z) \\
        &= \max_{x \in X} \left[ \max_{C_i \in C(x)} \min_{z \in C_i \cap A} u(z) \right] \\
        &= \max_{x \in X} u_{\min}(x, A).
\end{align*}
\]

Thus, the category model and the Gul and Pesendorfer (2001) model are both specializations of a more general max-min functional which is characterized by the Positive Set-Betweenness axiom.

Neither compromise nor attraction can be explained by the benchmark models in the temptation literature. There are extensions of these models, e.g. the model of Noor and Takeoka (2010), that can represent the compromise effect. However, to our knowledge none of these models can represent the attraction effect. Let us check the non-representability claim for the benchmark models. Put \( c(x, A) := \max_{z \in A} v(z) - v(x) \) and note that

35
This representation was generalized in Dekel et al. (2009) (DLR) who axiomatized the following model:

\[ U(A) = \sum_{i=1}^{n} q_i \left[ \max_{z \in A} u(z) - \sum_{k=1}^{n} c_k(z, A) \right]. \]

where \( c_k(z, A) := \max_{x \in A} v_k(x) - v_k(z). \) In the case of no uncertainty, the model reduces to the following

\[ U(A) = \max_{x \in A} [u(x) - \sum_{k=1}^{n} c_k(x, A)]. \]

Take this to be the benchmark “no uncertainty” model of temptation preferences. The compromise effect violates one of the DLR behavioral postulates (axiom AIT in their paper). To see that the attraction example cannot be represented consider each of the cost terms \( c_k(x, A) \) in the no-uncertainty utility and note that

\[ c_k(x, \{x, y, z\}) = \max\{c_k(x, \{x, y\}), c_k(x, \{x, z\})\} = 0 \]

Thus, the no-uncertainty DLR model cannot model the attraction example. Similarly reasoning on a “state-by-state” basis one checks that the full DLR model cannot represent the example either. Fixing a state, the cost functions \( c_k(\cdot, \cdot) \) can be interpreted as costs associated to different (subjective) attributes that the DM finds tempting, and hence needs to exert costly self-control to resist. The sets \( C_i \) that comprise the category \( C \) are also interpreted as indexing attributes that are (potential) temptations. The difference with Dekel et al. (2009) is that there is no costly self-control in our model. Ceteris paribus, for a given choice \( x \), if all attributes that could potentially tempt \( x \) are present in the menu, then it is as if \( x \) is not on the menu at all.

The preceding examples show that there is temptation-driven (terminology from Dekel et al. (2009)) choice data that can be captured by the category model, but which cannot be cannot be explained by Gul and Pesendorfer (2001) (GP), or the DLR extension of this model. Now let us ask the reverse question: what choice data can be explained by these models that cannot be captured by the category model? To answer this without trivializing the question, we need to appropriately define “choice data”. Note that the implied ex post choice correspondences for the DLR/GP models all satisfy WARP. Hence, if we take choice data to simply be the choice correspondence, we can trivially represent this choice data with the category model as the category model nests the standard model. Since part of the motivation for the temptation literature is to reconcile observed conflicts between choice and welfare, we will alternatively take the “choice data” to be a pair \((u, C)\), consisting of (i) a normative (i.e. “true welfare”) ranking on consumption choices and (ii) a choice correspondence on menus. The Gul and Pesendorfer (2001) theory shows that, if
we can observe choices between menus, then we can rationalize the data \((u, C)\) as coming from maximization of a net welfare ranking where we net out self-control costs. Now ask the following question: Fixing a dataset \((u, C)\) that admits a Gul and Pesendorfer (2001) (or, more generally, a no-uncertainty Dekel et al. (2009) representation), can we explain the same data with a category model?

The short answer is “no”. For instance, there might be ex post choices \(\{x, y\}\) under a GP utility for which \(u(x) > u(y)\), yet \(u(x) + v(x) = u(y) + v(y)\). Both of these elements will lie in the ex post choice correspondence; yet, a category representation will only select \(x\). This leads us to a refinement of our question: If we can’t exactly fit the data, we can nevertheless seek out a (category) model of “best fit”.

**Definition 16.** Fix choice data \((u, C)\). A category model \((u, C')\) is said to be a best approximation to \((u, C)\) if it has the following properties. Let \(\succeq (\text{resp. } \succeq ')\) denote the menu preference generated by \((u, C)\) (resp. \((u, C')\)) and \(C_{\succeq} (\cdot)\) the induced choice correspondence, i.e. \(C_{\succeq}(A) := \bigcup_i \inf(C_i \cap A)\).

1. \(C_{\succeq}(A) \subseteq C(A), \forall A \in \mathcal{M}\).
2. There is no model \((u, C')\) with \(C_{\succeq}(A) \subseteq C_{\succeq'}(A) \subseteq C(A), \forall A \in \mathcal{M}\).

The first condition says that choice induced by the category model should be consistent with the choice data. The second condition is the best fit property. It says that there is no other category model that explains more of the choice data than the putative best approximation. The result below addresses the no-uncertainty, multi-attribute generalization of the Gul and Pesendorfer (2001) model in Dekel et al. (2009). Denote these representations by pairs \((u, \{v_k\})\). For a technical reason, we restrict attention to observables \((u, C)\) where \(u(\cdot)\) is strict.\(^{34}\)

**Proposition 3.** Assume \(u(\cdot)\) is strict and that the observables \((u, C)\) admit a \((u, \{v_k\})\) representation. Then, there is a unique (sharp) Strotzian category \((u, C_{DLR})\) with the property that (i) \(C_{\succeq}(A) = \arg \max_{z \in C(A)} u(z)\) and (ii) \((u, C_{DLR})\) is a best approximation to \((u, C)\).\(^{35}\)

\(^{34}\)Note that if \(u(x) = u(y)\), it is in principle possible for the \(u + v\) value of \(y\) to exceed that of \(x\) since \(y\) may entail a lower cost of commitment. This means that the GP utility will only select \(y\) from a menu in which both \(x, y\) are feasible. Any category model, on the other hand, will select both – which would violate consistency with the choice data.

\(^{35}\)We choose the condition that \(u(\cdot)\) is strict since it is intuitively clear and sufficient for the result. However, the necessary and sufficient condition for the result is as follows: for observables \((u, C)\) which admit a GP (resp. no-uncertainty DLR) representation \((u, v)\) (resp, \((u, \{v_k\})\) the \(v\) function \(\sum_k v_k\) is constant on \(u\)-indifference classes. A pair of observables with a GP/DLR representation admits a best approximation category model, as described in the proposition, if and only if the pair \((u, \{v_k\})\) in the representation satisfies this condition. Note that when \(u(\cdot)\) is strict, the condition is trivially satisfied.
The uniqueness statement quantifies over the class of all sharp category models, not just Strotzian categories. Nevertheless, the unique best approximation turns out to be a Strotzian category. Note also that the containment, $C_\geq(A) \subseteq C(A)$, is strict if and only if there is a pair $(x, y)$ in the menu $A$ where the utility drop from $x$ to $y$ is exactly offset by savings in commitment costs from $x$ to $y$. In this case, the category model can explain the choice of $x$, but not the choice of $y$. Moreover, by the best approximation property this is endemic to the class of category models. For choice data $(u, C)$ where $u(\cdot)$ is strict, the proposition describes exactly when we require the costly self-control approach as opposed to the subjective feasibility approach to explain the observables $(u, C)$. In particular, insofar as explaining the data $(u, C)$, the GP/DLR model and category model don’t share common ground beyond Strotzian choices, i.e. where the DM is “overwhelmed” by the cost of self-control.

### 4 Menu Choice vs. Choices from Menus

The preceding section considered two sets of observables: (i) the “ex post” observables consisting of a pair $(u, C)$ and (ii) the ex ante observable of the menu preference, where the latter is (for most menu choice models) a richer observable. For our representation theorems, we have taken a preference relation on menus as the observable. However, given that one of our main objectives is to produce a model where choice reveals welfare, the more natural observable might be the pair $(u, C)$. The result in this section shows that, for the category model, these two sets of observables are equivalent, viz. the menu preference induces a unique pair $(u, C)$ and, conversely, any pair $(u, C)$ corresponds to a unique menu preference. We begin with a description of the alternate observables and recast the axioms for the category model using these observables.

- Let $X = \{x_1, \ldots, x_n\}$ be an enumeration of the prize space.
- Let $M$ denote the collection of non-empty subsets of $X$ (menus).
- Let $C : M \rightarrow M$ denote a non-empty choice correspondence, where $C(A) \subseteq A$.
- Let $u : X \rightarrow X$ denote an observable “welfare” ranking on consumption choices.

The first three items on this list are completely standard, but the fourth is not. When we say we can observe a consumption ranking $u(\cdot)$ we do not mean that we

---

36Given that the menu preference has a $(u, \{v_k\})$ representation, second period choice is the maximizer of $u + \sum_k v_k$. Hence, the fact that the $u$-maximal piece of the choice set $C(A)$ can be recovered by the Strotzian menu utility (with second period utility $u + \sum_k v_k$) is obvious. The more subtle part is that the associated category model has the best fit property.

38
can literally measure the consumer’s welfare from choices. Rather, what we have in mind is the idea that choice objects are made up of several attributes. Different DM’s might have diverse preferences over these attributes. However, there might be some attributes over which everyone share the same preference. For example, all else equal, every DM would rather pay less for a given choice $x$ than more. Note that this is exactly what $u$ would denote in the Simonson and Tversky (1993) experiment, viz. $u = \text{negative of price}$. Hence, these observables are exactly the observables of the Simonson and Tversky (1993) experiments.

The representation for the full category model uses axioms that involve both observables. This is needed if we want a model that can capture both the compromise and attraction effects. However, there is a subclass of the category model which can be characterized using the choice correspondence $C(\cdot)$ alone. This subclass is the largest category model for which the DM is constrained by only attraction effects, i.e. no compromise effects or (non-standard) Strotzian choices are admitted. Hence, while both observables are needed to model compromise and attraction, only choices from menus are needed to model attraction. That the utility on menus can be inferred from just choices from menus (and the welfare ranking on singleton choices) is a phenomenon particular to the category model. For instance, if we just observe ex post choices and the normative ranking on singleton menus, we cannot recover the Gul and Pesendorfer (2001) utility which generates these observables. Let us first recast the definition of subjective feasibility in the language of the observables $(u,C)$.

**Definition 17.** An element $x \in A$ is *subjectively feasible* if it satisfies the following property. For any $A' \subseteq A$ with $x \in A'$, $u(z) \geq u(x), \forall z \in C(A')$.

The interpretation of the condition is identical to the one given in Definition 1. However, it might be more transparent now since we earlier invoked the language of choices and welfare to understand the condition – even though choices and welfare were inferred from the menu preference. Here it is an explicit reference to the observables. Let us recapitulate for completeness. The condition that $u(z) \geq u(x)$ says that the DM’s welfare (from his choice) is weakly higher than what it would be had he chosen $x$ for *any* subset $A' \subseteq A$. As discussed before, we interpret this to mean that the DM could always choose $x$ if he wanted to. Hence, comparing welfare from the set of sub-problems $A'$ containing $x$ with welfare from the menu $\{x\}$ reveals whether the option $x$ is feasible. For a given menu $A$, let $A^*$ denote the subset of subjectively feasible elements. Recast the axioms as follows.

---

37In addition, when we observe $u$ we just mean the ordinal ranking on choices that $u$ represents. For brevity’s sake we maintain the $u$ notation, with the caution that we are not taking any cardinal data as observable.
**A1:** \( u(x) = u(y), \forall x, y \in C(A) \).

**A2:** If \( A^* \subseteq A' \subseteq A \), then \( C(A') = C(A) \).

**A3:** If \( A' \subseteq A \) and \( C(A) \cap A' \neq \emptyset \), then \( u(x) \geq u(y), \forall x \in C(A'), \forall y \in C(A) \).

The intuition for A1 is that if choices maximize welfare, then it must be that all elements of \( C(A) \) yield the same welfare value. A2 says two things. First, the value of a menu is determined by its feasible set, \( A^* \). Second, if there is another subset that contains the subjectively feasible set, then its subjectively feasible set agrees with the original feasible set, i.e, \( (A')^* = A^* \) if \( A^* \subseteq A' \subseteq A \). Finally, A3 says that – since choice is constrained by temptation – if we pass to a submenu, but keep at least one of the original choices available, then the DM is weakly better off since this choice is still feasible in the submenu. Moreover, he might be strictly better off on account of having removed some temptations from the menu which rendered options of \( u \)-value greater than \( x \) infeasible in the larger menu. Fixing the category model \((u, C)\), let \( A^*_{(u, C)} := \bigcup_i \arg \min_{x \in C_i \cap A} u(x) \).

**Definition 18.** Given a pair of ex post observables \((u, C)\), we say that the model \((u, C)\) represents the pair \((u, C)\) if

1. \( C(A) = \max x \in A^*_{(u, C)} u(x) \), and
2. \( u(x) = U(A), \forall x \in C(A) \).

The following is a recast of the main representation theorem of the paper.

**Theorem 6** (Main Result, Recast). A pair of ex post observables \((u, C)\) satisfies A1-A3 if and only if it admits a category model representation.

As mentioned earlier, general categories allow for the presence of both compromise and attraction effects, and Strotzian choices. If we strain out the effects of compromise effects and Strotzian choices, then we can make do with just ex post choices as the observable.

**Definition 19.** A category model \((u, C_{sp})\) is called a separated category if for each pair \((x, y)\) with \( u(x) > u(y) \) there is a pair of sets \( C_x, C_y \in C_{sp} \) such that \( x \in C_x, y \in C_y \setminus C_x \).

Notice that separated categories necessarily preclude compromise effects and Strotzian choices. Since the only three sources of cycles in our model are compromise, attraction, and Strotzian choice, once we strain out Strozian choices and

---

38 The \( u \) in the putative category representation need not cardinally agree with the \( u \) in the observable, only the underlying orders need to match. This point is irrelevant for the results so for brevity of notation we denote the (possibly distinct) cardinal representations with a common \( u \).
compromise effects we are left with just attraction effects as a source of deviations from Arrowian choice. Recall the base relation $R_C$ associated to a choice correspondence $C(\cdot)$, i.e. $x R_C y \Leftrightarrow x \in C(\{x,y\})$. Consider the axiom:

**A4:** $R_C$ is transitive.

Since $R_C$ is complete, A4 implies that the base relation $R_C$ is an order on $X$. Let $u(\cdot)$ denote a representation of this order. Now apply axioms A1-A3 to the pair $(u,C)$, where $C(\cdot)$ is the only observable and $u(\cdot)$ is a representation of $R_C$. The following representation is a trivial corollary of the theorem above; however, it is methodologically quite important. The value added is that it yields a *utility* representation of the attraction effect using only ex post choice as the observable. In other words, from observing only choices from the menu we are able to recover (and identify) the category model that determines both the utility of the menu and choices from the menu.

**Corollary 1.** A (non-empty) choice correspondence $C : \mathcal{M} \to \mathcal{M}$ satisfies A1-A4 if and only if it admits a separated category representation $(u,C_{\text{op}})$.

Both the theorem and the corollary follow from Theorem 1 in the text. Given observables $(u,C)$ we show below that there is a (canonical) menu preference that is attached to these observables and, moreover, the menu preference satisfies Axiom 1 (CRW) and Axiom 2 (Strong Reduction), so that Theorem 1 applies. However, more can be said here. The proposition below shows an equivalence between (i) menu preferences which satisfy Axiom 1 (CRW) and Axiom 2 (Strong Reduction) and (ii) pairs of ex post observables $(u,C)$ which satisfy A1-A3. Let $\Sigma_{P(X)}$ denote the subset of menu orders which satisfy Axioms 1 and 2. Let $\Sigma(X)$ denote the set of pairs $(u,C)$ consisting of (i) a utility $u$ on consumption choices and (ii) a choice correspondence $C$, where the pair $(u,C)$ satisfies A1-A3. Define a map $\Theta : \Sigma_{P(X)} \to \Sigma(X)$ as follows. Given $\succ$, let $u$ be a representation of the order on singleton menus. Second, let $C(A) = \arg \max_{x \in A} u(x)$. Put $\Theta(\succ) := (u,C)$.

**Proposition 4** (Equivalence of Observables). The sets $\Sigma(X)$ and $\Sigma_{P(X)}$ are bijectively equivalent under the map $\Theta$.

The proof of the recast version of Theorem 1 only requires that the map $\Theta$ is a surjection. It is, in principle, possible for two distinct menu preferences to induce the same set of ex post choices. The content of the proposition is that for menu preferences which satisfy Axioms 1-2 this cannot happen. This statement can be proven using the full force of the representation theorem, but we give a direct argument since (i) the statement of the proposition does not invoke the category model and (ii) this makes a more transparent connection between the observables.
5 Conclusion

This paper presents an axiomatic model of choice induced by temptation. One of the achievements of the existing temptation literature has been to provide a class of models that can explain discrepancies (experimental and anecdotal) between choice and welfare. Existing models of temptation connect choice and welfare via the idea of costly self-control. In these models, welfare is equal to the consumption utility from the choice net of the self-control cost required to make that choice (over other possibly tempting options on the menu). Moreover, choice is recovered as the arg max of this net welfare ranking. Our approach lies outside of the costly self-control paradigm. In our model, temptation still distorts choice from a first-best optimum – as in the costly self-control models. However, unlike costly self-control models, choice distortion occurs because temptation renders certain choices infeasible, as opposed to being too costly. Hence, choice can be distorted away from the first-best (unconstrained) optimum because (welfare) utility is maximized only over the subset of the choice problem consisting of subjectively feasible choices.

This approach allows us to explain experimental evidence, viz. the compromise and attraction effect, which seems to be temptation-driven (after Dekel et al. (2009)), but which cannot be explained using the Dekel et al. (2009) model. Or, in the case of the attraction effect, by any model that uses the costly self-control approach. Another difference between this paper and the extant temptation literature is that we can explain these effects with the same observable as the experiments in which these effects were documented (e.g. Simonson and Tversky (1993)) – so that we do not need the full richness of the menu preference. In the case of attraction, for example, the axioms that characterize the model can be entirely recast using choices from menus alone. Moreover, from observing just choices from the menu we elicit and identify the model (within the class of category models) that generates these choices. The model developed in this paper is a minimal departure from Arrowian choice. In addition to compromise and attraction, it admits an additional class of non-standard choice patterns – “Strotzian choice”. Compromise, attraction, and Strotzian choice delineate three distinct types of choice cycles. We show a one-to-one correspondence between each of these cycle classes with subclasses of our model where, for a fixed subclass, deviations from WARP are due to that type of cycle alone. Moreover, these are the only non-standard choice patterns captured by the overall model. If choices exhibit neither compromise/attraction effects and are not Strotzian, then the model collapses to Arrowian choice, i.e. unconstrained utility maximization.
6 Appendix

6.1 Proofs for Section 2

Proof of Lemma 1. Since $x \rightarrow_1 y$ we have $x \notin ({x} \cup \{y\} \cup A(y))^*$, by Strong Reduction. Moreover, $y \rightarrow_1 A(y)$ implies that $y \notin ({x} \cup \{y\} \cup A(y))^*$. Therefore, $({x} \cup \{y\} \cup A(y))^* \subset A(y)$. By Strong Reduction, this implies that $\{x\} \cup \{y\} \cup A(y) \sim \{x\} \cup A(y) \sim A(y)$. In particular, $\{x\} \succ \{x\} \cup A(y)$. \hfill $\Box$

Proof of Theorem 1. Necessity of Axiom 1 (CRW) is straightforward, hence we omit the argument. Necessity of Axiom 2 (Strong Reduction) is also straightforward, but we nevertheless provide the argument since it justifies why Axiom 2 contains the word “reduction” in its title. To this end, let $(u, \mathcal{C})$ be a category model and let $\succeq$ denote the underlying menu preference represented by $(u, \mathcal{C})$. Let $A^* = \{x \in A : A' \supseteq \{x\}, \forall A' \subseteq A, x \in A'\}$ be the subjectively feasible subset of $A$, and let $A^*_{(u, \mathcal{C})} = \bigcup_i \arg \min_{x \in \mathcal{C}_i \cap A} u(x)$, where we take $\mathcal{C} \equiv \{\mathcal{C}_i\}$. Note that Axiom 2 is implied by the equality

$$A^* = A^*_{(u, \mathcal{C})}$$

To check this, consider the left-to-right inclusion. Let $x \in A^*$ and put $\Sigma(x) = \{\mathcal{C}_i \in \mathcal{C} : x \in \mathcal{C}_i\}$. We claim that there must be some $\mathcal{C}_i \in \Sigma(x)$ such that $x \in \inf(\mathcal{C}_i \cap A)$. Else, for each $\mathcal{C}_i \in \Sigma(x)$ choose some $z_i \in \inf(\mathcal{C}_i \cap A)$ with $\{x\} \succ \{z_i\}$ and consider the menu $A' := \{z_i : z_i \in \mathcal{C}_i\} \cup \{x\}$. Note that $\{x\} \succ A'$ contradicting the fact that $x \in A^*$. Thus, there is some $\mathcal{C}_i \in \mathcal{C}$ such that $x \in \inf(\mathcal{C}_i \cap A) \subseteq A^*_{(u, \mathcal{C})}$. For the right-to-left inclusion take $x \in A^*_{(u, \mathcal{C})}$ and take any set $A' \subseteq A$ with $x \in A'$. Find $\mathcal{C}_i \in \mathcal{C}$ such that $x \in \inf(\mathcal{C}_i \cap A)$ and note that this implies $x \in \inf(\mathcal{C}_i \cap A')$. It follows that $A' \supseteq A$, implying that $x \in A^*$. Now take $A^* \subseteq A' \subseteq A$ and substitute $A^* = A^*_{(u, \mathcal{C})}$. Notice that this implies $(A')^*_{(u, \mathcal{C})} = A^*_{(u, \mathcal{C})}$. Hence, $A' \sim (A')^*_{(u, \mathcal{C})} = A^*_{(u, \mathcal{C})} \sim A$.

We now turn our attention to the sufficiency of the axioms. The proof of the representing set of categories uses a recursive construction that we call the “tree category.” To each $x \in X$ we associate an “x-tree”. The terminal nodes in each x-tree will be the elements of the overall category. We present the general construction, then return to Example 3 in the text to show how the construction applies in a concrete case.

Step 1: Constructing x-trees.

Introduce the following terminology. Fix an index set $\{1, 2, \ldots, N\}$. An x-tree is a triplet of data $(\{C^i_j\}_{j=1}^n, \{L^i_j\}_{j=1}^N, \{C^i_j \rightarrow C^i_{k+1}\})$, consisting (resp.) of nodes, levels, and branches, with the following structure:

- A collection of nodes $C^i_1, C^i_2, \ldots, C^i_n$ for each index $i$.
- A collection of levels $L^i(1), \ldots, L^i(N)$, where each $L^i := \{C^i_1, C^i_2, \ldots, C^i_n\}$.
A collection of branches \( \{ C_j^i \rightarrow C_j^{i+1} \} \) connecting nodes on consecutive levels.

Call \( C_j^i \) the root of the branch \( \{ C_j^i \rightarrow C_j^{i+1} \} \).

Every node \( C_k^i \) in level \( i \) (for \( i > 0 \)) has a unique root in level \( i - 1 \).

Every node \( C_k^i \) in level \( i \) is the root of a branch.

Using these objects, we inductively construct an \( x \)-tree as follows. First, make the following simplification. Since representability requires that \( B_t(x) \in \mathcal{C}_i \) whenever \( x \in \mathcal{C}_i \), we make no distinction between the element \( \{ x \} \) and the set \( \{ x \} \cup B_t(x) \). That is, whenever we say \( x \in \mathcal{C}_i \), what we implicitly mean (unless explicitly stated otherwise) is that \( B_t(x) \cup \{ x \} \subseteq \mathcal{C}_i \). Let \( A_1(x), A_2(x), \ldots, A_n(x) \) enumerate the attraction sets of \( x \) (here we distinguish between \( \{ x \} \) and \( B_t(x) \cup \{ x \} \)). The \( x \)-tree construction proceeds by double-induction. The outer induction is on the \( \geq \)-rank of the element \( x \) for which the \( x \)-tree has been constructed. The inner induction is on the tree construction for a fixed element \( x \). For a \( (\geq)|x| \)-minimal element \( x \) (in \( X \)), let the \( x \)-tree be just the singleton node \( \{ x \} \). Taking this as the base step of the outer induction, induct upwards on \( (\geq)|x| \)-rank to construct an \( x \)-tree as follows.

Let \( A_i(x) = \{ x_1^i, \ldots, x_k^i \} \) be the elements of the attraction set and take

\[
C_i^1 = \{ x, x_1^i \}, \mathcal{L}_1 = \{ C_1^1, C_2^1, \ldots, C_k^1 \}
\]

Note that the sizes of the attraction sets \( A_i(x) \) need not be the same. For notational brevity, we suppress this dependence – it will make no difference whatsoever for the ensuing arguments. Inductively, assume we have defined nodes and branches (with unique root restriction) for levels \( \{ 1, 2, \ldots, m \} \) (for \( m \leq n \) – where \( n \) is the total number of attraction sets for \( x \)) and define level \( m + 1 \) as follows.

Let \( \{ C_i^m \}_{i=1}^{N_m} \) be an enumeration of the nodes that form \( \mathcal{L}_m \). For each node \( C_i^m \) create \( |A_{m+1}(x)| \) branches as follows. Let \( A_{m+1}(x) = \{ x_{1}^{m+1}, x_{2}^{m+1}, \ldots, x_{k}^{m+1} \} \) and put

\[
C_i^{m+1} = C_i^{m} \cup \{ x_{1}^{m+1} \}, C_{i+1}^{m+1} = C_i^{m} \cup \{ x_{2}^{m+1} \}, \ldots, C_{k}^{m+1} = C_i^{m} \cup \{ x_{k}^{m+1} \}.
\]

Similarly, put

\[
C_{i-1,k+1}^{m+1} = C_i^{m} \cup \{ x_{1}^{m+1} \}, C_{i-2,k+2}^{m+1} = C_i^{m} \cup \{ x_{2}^{m+1} \}, \ldots, C_{0,k+1}^{m+1} = C_i^{m} \cup \{ x_{k}^{m+1} \}.
\]

Thus, level \( \mathcal{L}(m+1) \) consists of \( N_m |A_{m+1}(x)| \) nodes, \( C_{i}^{m+1} \), and \( N_m |A_{m+1}(x)| \) branches, \( \{ C_{i}^{m} \rightarrow C_{j}^{m+1} \} \) (where \( (i-1) \cdot k + 1 \leq j \leq i \cdot k \)). Inductively proceed until we exhaust all of the attraction sets \( \{ A_1(x), \ldots, A_n(x) \} \). Let \( \mathcal{L}(n) = \{ C(1), C(2), \ldots, C(N) \} \) be an enumeration of the nodes at level \( \mathcal{L}(n) \). For the next step, find the \( (\geq)|x| \)-maximal \( y \) such that \( \{ x \} \succ \{ y \} \) and for each \( C(i) \) with \( y \in C(i) \) attach a \( y \)-tree (which has been constructed by the induction hypothesis). This extends the levels in the original \( x \)-tree by the number of levels in the \( y \)-tree. For each \( C(i) \) in level \( \mathcal{L}(n) \) that does not contain \( y \) we just extend a single branch \( \{ C(i) \rightarrow C_{n+1}(j_i) \}, \{ C_{n+1}(j_i) \rightarrow C_{n+2}(k_i) \} \), and so on, for each subsequent level, where we put \( C(i) = C_{n+1}(j_i) = \cdots = C_{n+M}(l_i) \) (here we take \( M \) to be the number of levels in a \( y \)-tree). Thus, we obtain a tree with \( n + M \) levels. Now continue this procedure. Take a \( (\geq)|x| \)-maximal \( z \) with \( \{ y \} \succ \{ z \} \) and for each \( C(i) \in \mathcal{L}(n+M) \) with \( z \in C(i) \) attach a \( z \)-tree. Iteratively
then we fix a top down labeling of the singleton ranking, the claim is that the pair $(u, C)$ represents $\succeq$. Let $U(\cdot)$ be any cardinal representation of $\succeq$ (which extends $u$) and let $U^C(\cdot)$ denote the utility defined by the category formula for the pair $(u, C)$, where $u \equiv U(\cdot)|_X$.

We show representability by checking equality $U^C(\cdot) = U(\cdot)$ on all menus. The tree structure of the categories allows for a useful decomposition of the function $U^C(\cdot)$. Let $\{\mathcal{L}(N_x)\}$ denote the set of all terminal levels across all trees. Let $\mathcal{T}_{y_1}, \mathcal{T}_{y_2}, \ldots, \mathcal{T}_{y_n}$ be an enumeration of all trees (where $|X| = n$) and define

$$U^{T_x}(A) := \max_{C_i \in \mathcal{L}(N_y) : C_i \cap A \neq \emptyset} \min_{z \in A \cap C_i} u(z)$$

As before, we suppress the requirement that we maximize only over terminal nodes $C_i$ which intersect $A$. Observe that we have the equality

$$U^C(A) = \max_{x \in X} U^{T_x}(A)$$

Thus, the value the category utility assigns to menu $A$ is the maximum of its value across trees. We analyze $U^C(\cdot)$ by analyzing its behavior on a tree-by-tree basis. In particular, we check that for each $x$ we have $U^{T_x}(A) \leq u_{\min}(x, A)$. If $u_{\min}(x, A) = u(x)$, then this claim is obvious since we clearly have $u_{\min}(x, A) \leq u(x)$ for each terminal node $C_i$ of the $T_x$-tree. Consider the case where $u_{\min}(x, A) < u(x)$. For a given tree $T_x$, let $\mathcal{L}(1), \ldots, \mathcal{L}(N_x)$ denote its levels. For each level $\mathcal{L}(i)$ consider the function

$$U^{\mathcal{L}(i)}(A) := \max_{C_i \in \mathcal{L}(i)} \min_{z \in A \cap C_i} u(z).$$

Note that the menu $A$ contains either a singleton temptation or a non-trivial attraction set $A(x)$ for $x$. Consider the latter case and note that by Strong Reduction and the minimality property of $A(x)$, we may find $z \in \sup(A(x))$ such that $\{x\} \cup A(x) \sim \{z\}$. Let $\mathcal{L}(i)$ denote the level at which this attraction set is introduced into each node of the $T_x$-tree. We then have

$$U^{\mathcal{L}(i)}(A) \geq U^{\mathcal{L}(i+1)}(A) \geq \cdots \geq U^{\mathcal{L}(N_x)}(A) = U^{T_x}(A) \tag{*}$$

Note that $U^{\mathcal{L}(i)}(A) \leq u(z)$, so that $U^{T_x}(A) \leq u(z)$. This holds for all attraction sets $A(x) \subseteq A$.

---

39If there are no ties in the singleton ranking, this construction is canonical. If there are ties, then we fix a top down labeling $\{x_1, \ldots, x_n\}$ at the outset and at each step of the recursion pick a $(\succeq)|_X$-maximal $z$. When there are ties, break the tie according to the pre-selected labeling.
Now we construct a particular attraction set \( A_s(x) \) with \( U(\{x\} \cup A_s(x)) = u_{\min}(x, A) \). Consider the orders \( \succeq_A \) underlying the function \( u_{\min}(-, A) \). Let \( I_\lambda(x) \) denote the \( \succeq_A \)-indifference class of \( x \) and let \( I_\lambda^1, \ldots, I_\lambda^k \) denote a top-down enumeration of those \( \succeq_A \)-classes where \( I_\lambda^1 := I_\lambda(x) \). Put

\[ A'(x) = \cup_{\lambda \geq 1} \inf(I_\lambda^1) \]

Note that \( \{x\} \cup A'(x) \sim \{z\} \) for any \( z \in \inf(I_\lambda^1) \). We would like to claim \( A'(x) \) is the desired attraction set, but it may not be minimal. Consider the collection of all subsets, say \( A''(x) \), of \( A'(x) \) which have the property that \( \{x\} \cup A''(x) \prec \{x\} \). Find a minimal (w.r.t. set inclusion) such set, call it \( A_s(x) \). Since \( u_{\min}(x, A) = u(x) \) for \( x \in \inf(I_\lambda^1) \) and \( \inf(I_\lambda^1) \succ \inf(I_\lambda^2) \succ \cdots \succ \inf(I_\lambda^k) \) (again, see Steps 1 and 2 of the proof of Theorem 2 for the proof of this fact) we must have \( \sup(A_s(x)) \cap \inf(I_\lambda^1) \neq \emptyset \). It follows that \( A_s(x) \) is an attraction set for \( x \) with the property that \( U(\sup(A_s(x))) = u_{\min}(x, A) \). Apply the conclusion of the preceding paragraph to the attraction set \( A_s(x) \) to obtain \( U^{T_x}(A) \leq u_{\min}(x, A) \).

Now consider the case where \( A \) contains a singleton temptation, i.e. \( z \in B_t(x) \) with the property that \( u_{\min}(x, A) = u(z) \). In this case, \( z \) is introduced at the root node of the tree \( T_x \), so that the \( u \)-minimum of each terminal node in this tree is trivially bounded above by \( u(z) \) when \( z \) is on the menu. It follows that, for each \( x \in A \), the value of the category function \( U^C(\cdot) \) on the \( x \)-tree \( T_x \) is bounded above by \( u_{\min}(x, A) \). In other words, for each \( x \in A \) we have \( U^{T_x}(A) \leq u_{\min}(x, A) \). For each \( x \not\in A \) we consider two cases, either (i) \( A \cap C_i = \emptyset, \forall C_i \in L(N_x) \) or (ii) \( C_i \cap A \neq \emptyset \) for some \( C_i \in L(N_x) \). In the former case, the function \( U^{T_x}(A) \) does not enter into the domain of the maximization \( U^C(A) = \max U^{T_x}(A) \). In the latter case, each node \( C_i \) with \( C_i \cap A \neq \emptyset \) contains a node \( C_i' \in L(N_x) \) and \( z \in C_i \cap A \). Note that \( \min_{w \in C_i \cap A} u(w) \leq \min_{w \in C_i \cap A} u(w) \leq U^{T_x}(A) \), where \( C_i' \in L(N_x) \) and \( z \in C_i \cap A \). Hence, we obtain \( U^{T_x}(A) \leq \max_{z \in A} U^{T_x}(A) \leq \max_{z \in A} u_{\min}(z, A) \). Now recall that \( U(A) = \max_{x \in A} u_{\min}(x, A) \) and putting together the bounds on \( U^{T_x}(A) \) for each tree \( T_x \) we obtain

\[ U^C(A) = \max U^{T_x}(A) \leq \max_{z \in A} u_{\min}(x, A) = U(A) \]

**Step 3:** Check the inequality \( U^C(\cdot) \geq U(\cdot) \).

We now check the reverse inequality. First introduce some terminology. For notational brevity (for this step alone) we denote level \( i \) tree nodes as \( x_j(i) \) (the \( j \)-th node in level \( i \)) and branches are denoted \( x_j(i) \to x_k(i+1) \). Fix a menu \( A \) and some \( x \in A \). Consider the \( x \)-tree \( T_x \) and consider the set of all directed paths along branches in the tree, \( \Phi := \{\ell : \ell \in \{x \to x_{i_1}(1) \to x_{i_2}(2) \to \cdots \to x_{i_{N_x}}(N_x)\}\} \).

Let \( \ell = (\ell_1, \ell_2, \ldots, \ell_{N_x}) \) denote the specific nodes which lie along the path \( \ell \).
say that $x$ is unobstructed in $T_x$ by the menu $A$ if there is a path $\ell \in \Phi$ such that $\ell \cap A = \{x\}, \forall \ell$. We claim that $x$ is unobstructed (by a menu $A$) if and only if $x \in A^*$. Note that if we prove this, then it follows that the value of the category function $U^C(A)$ on the tree $T_x$ is $u(x)$, which proves $U^C(A) \geq U(A)$. Hence, we reduce to proving the claim that $x$ is unobstructed in the $x$-tree $T_x$ at the menu $A$ if and only if $x \in A^*$.

For the “if” part of the claim assume that $x$ is unobstructed. Then clearly the value of the category function on the tree $T_x$ is $u(x)$. It follows that $x \in A^*$. Now for the reverse direction. Proceed by induction on the $(\succeq)_X$-rank of $x$. That is, for the base step take $z \in \inf(X)$ and verify that: For any menu $A$ with $z \in A$ and $z \in A^*$, the tree $T_z$ is unobstructed by $A$. Now induct upwards. If $x$ is the lowest ranked singleton in $A^*$, then the unobstruction claim is obvious. Thus, assume $x$ is not the lowest ranked singleton in $A^*$ and wlog that $x \sim A^*$ (since, if $T_x$ is obstructed in $A$, it can only be obstructed by elements with $(\succeq)_X$-rank strictly lower than $x$). Moreover, for the unobstruction claim we actually need only consider menus $A$ of the form $A \equiv \{x\} \cup A'$ where $A' := \cup_{y:x \succ y} y$. Thus, we shall assume that $A$ has this form.

Let $A_1(x), A_2(x), \ldots, A_k(x)$ be an enumeration of the attraction sets of $x$. We claim that for each $A_i(x)$, there is some $z_i \in A_i(x)$ such that (i) the $z_i$-tree is unobstructed in $A$ and (ii) $z_i \notin A$. Check this via contradiction. Assume $A_1(x)$ is such that for every $z \in A_1(x)$ either (i) $z$ is obstructed in $A$, or (ii) $z \in A$. Then, when $z$ is obstructed in $A$ we have (by the induction hypothesis) $z \notin (A \cup \{z\})^*$. Let $\{z_1, \ldots, z_n\}$ be elements of $A_1(x)$ that are obstructed by $A$ and let $\{z_{n+1}, z_{n+2}, \ldots, z_m\}$ be the elements of $A_1(x)$ that are unobstructed by $A$, but for which $z_i \notin A$. Since we are alleging $\{z_1, z_2, \ldots, z_n\} \cup \{z_{n+1}, \ldots, z_m\} = A_1(x)$, we then obtain $(A_1(x) \cup A)^* \subseteq A^*$. It follows (by Strong Reduction) that $A^* \sim A_1(x) \cup A$. OTOH, by Strong Reduction again, $A \sim A^* \sim \{x\}$ - contradiction. Thus, for each $A_i(x)$ find $z_i$ such that (i) $z_i$ is unobstructed in $A$ and (ii) $z_i \notin A$.

For what follows we will need to concatenate paths from different levels (in the $x$-tree $T_x$). Let $\ell(i,j)$ denote a path connecting a node in level $\ell(i)$ to a node in level $\ell(j)$. Recall that the element $x$ had $k$ attraction sets. Let $z_1, \ldots, z_k$ be a list of unobstructed elements chosen respectively from $A_1(x), \ldots, A_k(x)$ (and such that $z_i \notin A$). Let $x(k)$ denote a node in level $\ell(k)$ of the $x$-tree that contains $\{x, z_1, \ldots, z_k\}$, i.e. we sequentially attach $\{z_1\} \cup B_1(z_1), \{z_2\} \cup B_1(z_2), \ldots, \{z_k\} \cup B_1(z_k)$ to the initial node comprised of $\{x\} \cup B_1(x)$. Let $\ell(0,k)$ denote the unobstructed path from level 0 to level $k$ that ends at the node $x(k)$. From the construction of the $x$-tree, we successively attach $y_i$-trees for some $y_1, y_2, \ldots, y_l$. Let $y_{i_1} = y_1 = z_{i_1}, y_{i_2}, \ldots, y_{i_k}$ be the subsequence of $\{y_1, y_2, \ldots, y_l\}$ where we first attach a $z_i$-tree in the $x$-tree construction algorithm. Let $\ell(N_1), \ldots, \ell(N_l)$ denote the terminal levels of the $y_i$-trees.
being attached to the nodes at level $L(k)$. We inductively concatenate the unobstructed path $\ell(0,k)$ with an unobstructed path $\ell(k,k+N_1)$ and in turn with an unobstructed path $\ell(k+N_1,k+N_1+N_2)$, and so on.

If $y_1 \in \{z_1,\ldots,z_k\}$ (call this the “unobstructed set”), there is an unobstructed path $\ell(k,k+N_1)$ from the node $x(k)$ to some node $x(k+N_1)$ in level $L(k+N_1)$ of the partial tree obtained by concatenating the $y_1$-tree to the preceding levels $L(1), L(2), \ldots, L(k)$. Concatenate $\ell(0,k)$ to $\ell(k,k+N_1)$ to obtain an unobstructed path from level $L(0)$ to level $L(k+N_1)$. If $y_1 \in B_t(x)$, then $y_1 \notin A$ and $y_1$ must be unobstructed at $A$. The argument for this mimics the companion case for $y_2$ (resp. $y_3$) and is presented in more detail below. Now consider $y_2$. If $y_2 \notin x(k+N_1)$, then a copy of the node $x(k+N_1)$ is replicated on every level of the attached $y_2$-tree. Thus, take $\ell(k+N_1,k+N_1+N_2)$ to be the path which concatenates the branches $\{x(k+N_1) := C_0 \rightarrow C_1\}, \{C_1 \rightarrow C_2 = x(k+N_1)\}, \ldots, \{C_{N_2−1}(=x(k+N_1)) \rightarrow C_{N_1}(=x(k+N_1))\}$. If $y_2 \in x(k+N_1)$, consider two cases. Either $y_2$ is in the unobstructed set or it is not. In the former case, repeat the argument for the $y_1$ case to extend the unobstructed path. If $y_2$ is not in the unobstructed set, then either (i) $y_2$ is first introduced at some level $L(k+k_y)$ of the $y_1(=z_1)$-tree or (ii) $y_2 \in B_t(x)$ or $B_t(z_1)$. Consider case (i). Let $x(k+k_y)$ be the unique predecessor node at this level for which there is an unobstructed path starting at $x(k+k_y)$ and terminating at $x(k+N_1)$. Since we attach a $y_2$-tree at some level of the construction of the $y_1$-tree, a portion of this path must pass through (unobstructed) an $y_2$-tree. Denote this path segment as $\ell(k_1,k_2)$ (where there are $N_2$ branches that comprise this segment). Note that the $y_2$-tree is embedded inside the $y_1$-tree, so that (by the recursiveness of the tree construction) the image of the same path in any embedded $y_2$-tree is unobstructed (w.r.t the menu $A$). Let $\ell(k+N_1,k+N_1+N_2)$ be a copy of the path $\ell(k_1,k_2)$ with root node $x(k+N_1)$, and that passes through the $y_2$-tree. Consider the concatenation, $(\ell(0,k); \ell(k,k+N_1); \ell(k+N_1,k+N_1+N_2))$ and note that this concatenation is unobstructed. Let the terminal node of this path be denoted $x(k+N_1+N_2)$.

Now consider case (ii). If $y_2 \in B_t(x)$, then we claim it must be unobstructed at $A$ (in addition to $y_2 \notin A$). Else, $y_2 \notin (A \cup \{y_2\})^*$ (again, by the induction hypothesis). Hence, $(A \cup \{y_2\})^* \subset A^*$ which implies, by Strong Reduction, that $A^* \sim A \cup \{y_2\}$ and, in turn, $A \sim A \cup \{y_2\}$. OTOH, $\{x\} \sim A$ and $\{x\} > A \cup \{y_2\}$ – contradiction. Similarly, if $y_2 \in B_t(z_1)$ it must be unobstructed in $A$ (in addition to $y_2 \notin A$). Else, an analogous argument shows that $z_1$ is obstructed in $A$. Hence, in either case we can find an unobstructed path going through the $y_2$-tree. As in case (i), we concatenate the initial path $(\ell(0,k); \ell(k,k+N_1))$ with an unobstructed path through the attached $y_2$-tree, call the latter $\ell(k_1+N_1,k+N_1+N_2)$. Note that the concatenation $(\ell(0,k); \ell(k,k+N_1); \ell(k+N_1,k+N_1+N_2))$ is unobstructed. Next consider $y_3$. If $y_3$ is in the unobstructed set we proceed verbatim as above. Else, either (i) $y_3$ is first introduced along an unobstructed path in the $y_t$-tree for
some \( i \leq 2 \) or (ii) \( y_3 \in B_i(x), B_i(z_1), \) or \( B_i(z_2) \) (the latter if \( y_2 = z_2 \)). In case (i), a portion of one of the paths \( \ell(k + N_1, k + N_1 + N_2) \) or \( \ell(k, k + N_1) \) must pass unobstructed through the \( y_3 \)-tree (at the point in the construction of the \( y_1 \) (resp. \( y_2 \)) tree where the \( y_3 \) tree is attached). Let \( \ell(k + N_1 + N_2, k + N_1 + N_2 + N_3) \) denote a replica of this path that goes through the \( y_3 \)-tree. Continue the concatenation \( (\ell(0, k); \ell(k, k + N_1); \ell(k + N_1, k + N_1 + N_2); \ell(k + N_1 + N_2, k + N_1 + N_2 + N_3)) \) to obtain an unobstructed partial path. Case (ii) is dealt with the same way as in the preceding argument (i.e. case (ii) for \( y_2 \)). Inductively proceed to extend the concatenated path to obtain a sequence (put \( k_i = k + \sum_{j=1}^{t} N_j \) \( (\ell(0, k); \ell(k, k_1); \cdots; \ell(k_{t-1}, k_t)) \). Note that this is a complete path in the \( x \)-tree and, moreover, it is unobstructed. It follows that the value of the category function \( U^C(\cdot) \) on the tree \( T_x \) is \( u(x) \), proving that \( U^C(A) \geq U(A) \). This concludes the proof of the theorem.

Illustration of Algorithm:
Let us apply this tree construction to Example 3 in the main text. In particular, we demonstrate the argument in step 3 of the theorem (the “concatenation of paths” argument). Consider the menu \( A = \{x, z\} \). Since the underlying menu preference is generated by the category model \((u, C)\) we know that \( A \sim \{x\} \). Thus, for each of the attraction sets \( A_1(x), \ldots, A_4(x) \) there is some \( z_i \in A_i(x) \) such that \( z_i \not\in A \) and such that \( z_i \) is unobstructed in \( A \). Note that only \( x \) has any non-trivial attraction sets, so that for any \( z_i \in A_i(x) \) the statement that \( z_i \) is unobstructed in \( A \) just means that there are no direct temptations for \( z_i \) in the menu \( A \). Pick

- \( z_1 = y \in A_1(x) = \{y, z\} \)
- \( z_2 = p \in A_2(x) = \{p, q\} \)
- \( z_3 = y \in A_3(x) = \{y, q\} \)
- \( z_4 = p \in A_4(x) = \{p, z\} \).

The choice of the quadruple \((z_1, z_2, z_3, z_4)\) is not canonical. For example, we could equally well have picked \( z_2 = q, z_4 = z \) – among several other alternatives. This is of no consequence, however, since all we need to do is produce a single unobstructed path that passes through the \( x \)-tree \( T_x \). To this end, we just do this for the chosen quadruple \((z_1, z_2, z_3, z_4)\). Consider the following path (labelings of nodes follow the notation given in example 3):

\[
C_1^0 \leftrightarrow C_1^1 \leftrightarrow C_1^2 \leftrightarrow C_1^3 \leftrightarrow C_1^4
\]

Clearly this path is unobstructed, with \( A \cap C_i^1 = \{x\}, \forall i = 1, 2, 3, 4 \). For the general argument given in the proof of the theorem, the concatenation of paths need not be so straightforward, although the idea is the same. In the above example the choice of concatenations is particularly simple since only \( x \) has non-trivial attraction sets.
Proof of Theorem 2. Necessity: Fix a local category model \((u, \{C_A\}_A)\) and let \(C^1_A, \ldots, C^n_A\) denote the categories in the collection \(C_A\) and let \(X_1, \ldots, X_n\) denote the sets arg \(\min_{y \in C_A^n} u(y)\). Let \(\succeq\) denote the menu preference generated by the local category model and let \(A \sim X_1\). By downwards rigidity, there is some \(x_1 \in X_1\) such that \(x_1 \notin \cup_{j \neq 1} C^j_A\). Put \(A' \subseteq A\). We claim that if \(x_1 \in A'\), then \(A' \succeq A\). Let \(C^1_A, \ldots, C^n_A\) be the categories in the collection \(C_A\) and label so that \(x_1 \in C^1_A\). By coherence, there is a category \(C^1_A \in C_A\) such that \(C^1_A \subseteq C_i^1\). Since \(x_1 \in C^1_A\), downwards rigidity implies that \(C^1_A = C^1_A\). Thus, \(\min_{z \in C^1_A} u(z) = u(x_1) = U(A)\). It follows that \(A' \succeq A\).

Sufficiency: Introduce an auxiliary function, \(u_{\min}(x, A) := \min\{U(A') : A' \subseteq A, x \in A'\}\). Here we take \(U(\cdot)\) to be any cardinal representation of the menu order \(\succeq\). Note that the function \(u_{\min}(\cdot, \cdot)\) yields a family of rankings \(\{\succeq_A\}_{A \in M}\) that is independent of the choice of \(U(\cdot)\). Recall the following axiom, which is “one-half” of the Set-Betweenness axiom introduced in Gul and Pesendorfer (2001): \(A \succeq B \Rightarrow A \succeq A \cup B\). Note that this property, referred to as “Positive Set-Betweenness” (PSB) in the literature, is implied by CRW. Introduce two pieces of notation. First, let \(A_x\) denote the maximal menu (w.r.t. set inclusion) such that \(\succeq\) and let \(\Sigma(\cdot)\). Note that the function \(\theta_x(A) := \{y : x \succeq_A y\}\). Construct a local category representation in a sequence of steps.

Step 1: Check that \(A_x = \theta_x(A)\).

Note that if \(x \succeq_A y\), then \(A_x \succeq A_y\). This implies, by PSB, that \(A_x \succeq A_x \cup A_y\). Maximality then implies \(A_x = A_x \cup A_y\). Thus, \(\theta_x(A) \subseteq A_x\). To check the reverse containment, let \(y \in A_x\) with \(y \succeq_A x\). Then, since \(y \in A_x\), we obtain \(u_{\min}(y, A) = U(A_y) \leq U(A_x) = u_{\min}(x, A)\). On the other hand, \(y \succeq_A x\) implies, by definition, \(u_{\min}(y, A) \geq u_{\min}(x, A)\) so that \(y \succeq_A x\).

Step 2: Let \(I^1_A, \ldots, I^n_A\) denote a top-down enumeration of the \(\succeq_A\)-indifference classes and let \(\Sigma(A)\) denote the set of implied choices in \(A\), i.e. \(\Sigma(A) := \{x \in A : (i) \{x\} \sim_A A, (ii) x \in A' \subseteq A \Rightarrow A' \succeq A\}\). We check that \(\Sigma(A) = \inf\{y : y \in I^1_A\}\). This connects the set of choices with the revealed relation \(\succeq_A\). First observe that \(\Sigma(A) \subseteq I^1_A\).

Otherwise, if \(x \in \Sigma(A) \cap I^1_A\) for some \(j > 1\), then by CRW we obtain \(\cup_{j \geq 2} I^j_A \succeq A\). On the other hand, by Step 1, \(A = \cup_{j=1}^k I^j_A \supset \cup_{j \geq 2} I^j_A\) – contradiction. Thus, \(\Sigma(A) \subseteq I^1_A\). Now for any \(y \in I^1_A\) consider the menu \(A' := \{y, I^2_A, \ldots, I^n_A\}\) and note that \(A' \succeq A\), again by definition of \(\succeq_A\) and Step 1. Take any \(x \in \Sigma(A)\) so that we obtain: \(A' \succeq A \sim \{x\}\). On the other hand, we claim that \(u_{\min}(y, A') = U(A')\).

To see this, observe that \(u_{\min}(y, A') = u_{\min}(x, A), \forall x \in A'\}\) by Step 1. Since \(u_{\min}(y, A) > u_{\min}(x, A), \forall x \in A'\}\) it follows that

\[
    u_{\min}(y, A') \geq u_{\min}(y, A) > u_{\min}(x, A) = u_{\min}(x, A'), \text{ for all } x \in A'\}\).
Therefore, by Step 1, we must have \( u_{\min}(y, A') = U(A') \). Thus, we obtain
\[
U(\{y\}) = u_{\min}(y, \{y\}) \geq u_{\min}(y, A') = U(A') \geq U(A) = U(\{x\}).
\]
It follows that \( \{y\} \succeq \{x\}, \forall y \in I_1^A \), implying that \( \Sigma(A) \subseteq \inf\{y : y \in I_1^A\} \). To show the reverse containment, simply note that if \( y \in \inf\{y : y \in I_1^A\} \) and \( y \in A' \subseteq A \), then by definition of \( \succeq_A \) and Step 1 we have \( A' \succeq A \).

**Step 3**: We now construct the local categories. Note that the above argument implies that, \( \inf I_1^A \succ \inf I_2^A \succ \cdots \succ \inf I_k^A \). We first present the construction under the assumption that there are no ties in the singleton ranking, \((\succeq)_X \). Having carried out the construction without ties, we then introduce ties. The reason for breaking the argument up like this is that we need to introduce several pieces of notation to deal with ties and it is easier to absorb the construction when it is freed from this notation.

**Step 3a**: Construction when \((\succeq)_X \) is strict.

Consider the following candidate. Fix a menu \( A \) and let \( I_1^A, \ldots, I_k^A \) be a top-down enumeration of the \( \succeq_A \)-indifference classes. For any menu \( A \), let \( I_A(x) \) denote the \( \succeq_A \)-indifference class of \( x \in A \) and let (abusing notation)
\[
I_A(x) := \bigcup_{\Phi_x(A)} \{y \in A' : \{y\} \succeq \{x\}, y \in I_A'(x)\}
\]
where the union is taken over all sets \( A' \) such that (i) \( A' \subseteq A \), and (ii) \( x \in \Sigma(A') \). Denote this collection of sets as \( \Phi_x(A) \). Inductively define a sequence of sets as follows. Put
\[
D_A^i(1) := \bigcup_{x \in I_A^i} I_A(x)
\]
and let
\[
D_A^i(k) := \bigcup_{x \in D_A^i(k-1)} I_A(x)
\]
Note that \( D_A^i(k) \subseteq D_A^i(k+1) \) so that the sequence terminates. Let \( D_A^i \) denote the terminal element for each \( i \) and let \( C_A \equiv \{D_A^1\} \) be the candidate category. We claim that the pair \((u, \{C_A\})\) gives a local category representation of \( \succeq \). This requires us to check (i) representability and (ii) that \( \{C_A\} \) satisfies coherence and downwards rigidity.

**Coherence**: Let \( A' \subseteq A \). We wish to show that for any \( D_{A'}^i \), there is some \( j \) such that \( D_{A'}^j \subseteq D_A^i \). Note that once we obtain \( D_{A'}^i(n) \subseteq D_A^i(m) \) for a fixed pair of integers \( m \geq n \), then it follows that \( D_{A'}^i(n+1) \subseteq D_A^i(m+1) \), and so on; so that \( D_{A'}^i \subseteq D_A^i \). To this end, let \( x \in \inf(I_{A'}^i) \) and note that \( I_{A'}^i(x) \subseteq D_A^i(1) \) for some \( j \).
with $x \in I^j_A$. Thus, $D^j_{A'}(1) \subseteq D^j_A(2)$.

**Downwards Rigidity:** We claim that if $x \in \inf(I^k_A)$, then $x \notin D^j_A$ for any $j \neq k$. Towards contradiction, let $x \in \inf(I^k_A) \cap D^j_A$. Thus, $x \in I_{A'}(y)$ for some $y \in D^j_A(n)$ and $A' \subseteq A$ such that $y \in \Sigma(A')$. Note that since $x \in \inf(I^k_A)$ and the indifference classes $I^k_A$ are disjoint, if $x \in D^j_A$ then there is a minimal $n$ such that $x \in D^j_A(n+1), x \notin D^j_A(n)$ (where we take $D^j_A(0) := I^k_A$). Thus, we can assume $x \neq y$. Since $x \in D^j_A(n+1)$, this means that $x \in I^j_A(y)$, where $y \in D^j_A(n)$. This, in turn, implies that $\{x\} \succ \{y\}$ (since $(\succeq)_X$ is strict). On the other hand, Step 2 implies that $u_{\min}(x, A) = u(x) -$ contradicting the requirement that $x \in I_{A'}(y)$.

**Representability:** Step 2 implies that $\inf(I_A(x)) \sim x$ whenever $x \in \Sigma(A)$. Thus,

$$\min_{x \in D^j_A(1)} u(x) = \min_{x \in I^j_A} u(x)$$

Applying the same reasoning to each $D^j_A(n)$ we find:

$$\min_{x \in D^j_A(n)} u(x) = \min_{x \in D^j_A(n-1)} u(x) = \cdots = \min_{x \in I^j_A} u(x).$$

It follows that, for each $i$, $\inf(D^j_A) = \inf(I^j_A)$. Fix some $x^i \in \inf(I^j_A)$. Since $\inf(I^j_A) \succ \cdots \succ \inf(I^j_A)$, we obtain $u(x^i) = \max_i \{u(x^1), \ldots, u(x^i), \ldots, u(x^k)\} = U(A)$.

**Step 3b:** Construction when we allow ties in $(\succeq)_X$.

We will just indicate the where we need to make changes in the argument in Step 3a. Fix a menu $A$ and recall the sets $D^j_A(k)$. Recall the notation

$$I^j_A(x) = \bigcup_{\Phi_x(A)} \{z : \{z\} \succeq \{x\}, z \in I_{A'}(x)\}$$

where the union is over the collection of subsets $\Phi_x(A)$ consisting of all $A'$ with $A' \subseteq A$ and $x \in \Sigma(A')$. Using these sets we recursively put (with $D^j_A(0) := I^j_A$)

$$D^j_A(k) = \bigcup_{z \in D^j_A(k-1)} I^j_A(z)$$

This summarizes the construction without ties. When there are ties among singletons we amend this definition as follows. First, let

$$I^{\triangleright}_A(x) := \{x\} \bigcup_{\Phi_x(A)} \{z : \{z\} \succ \{x\}, z \in I_{A'}(x)\}$$

and put $\{x^1_1, x^2_1, \ldots, x^i_n\} = \inf(I^j_A)$. For each $x^i_j$ recursively define (put $D^j_A^{(j,i)}(0) := \{x^i_j\} \cup \{z \in I_A(x^i_j) : \{z\} \succ \{x^i_j\}\}$)

$$D^j_A^{(j,i)}(k) = \bigcup_{z \in D^j_A^{(j,i)}(k-1)} I^j_A(z)$$

52
As before the sequence of these sets is increasing, i.e. \( D_{A}^{(j,i)}(k) \subseteq D_{A}^{(j,i)}(k+1) \) – hence, terminates at some \( D_{A}^{(j,i)} \). Also note that \( \cup_{(j,i)} D_{A}^{(j,i)} = A \). Representability follows as before, so that the only possible issues are coherence and downwards rigidity.

Coherence: Fix \( A' \subseteq A \) and take \( D_{A'}^{(j,i)} \). Note that, by construction, \( \inf(D_{A'}^{(j,i)}) = x_{j}^{i} \in \inf(I_{A'}) \). Find \( I_{A}^{l} \) such that \( x_{j}^{i} \in I_{A}^{l} \). Let \( \{z_{1}^{l}, \ldots, z_{n_{l}}^{l}\} = \inf(I_{A}^{l}) \) and note that \( D_{A'}^{(j,i)}(0) \subseteq D_{A}^{(q,l)}(1) \) for some \( 1 \leq q \leq n_{l} \). As in the proof in Step 3a this implies that \( D_{A'}^{(j,i)}(n) \subseteq D_{A}^{(q,l)}(n+1), \forall n \geq 0 \). Thus, since both sets stabilize we obtain \( D_{A'}^{(j,i)} \subseteq D_{A}^{(q,l)} \) – proving coherence.

Downwards Rigidity: Note that we can enumerate the sets in the local category

\[
C_{A} = \{D_{A}^{(j,i)}\}_{j \leq n_{i}, i \leq k}
\]

where \( I_{A}^{1}, I_{A}^{2}, \ldots, I_{A}^{k} \) enumerates the \( \succeq_{A} \)-indifference classes. Moreover, by construction, \( x_{j}^{i} = \inf(D_{A}^{(j,i)}) \). Thus, \( x_{j}^{i} \notin D_{A}^{(j,i)} \) for any other \( n \neq j \) and \( n \leq n_{i} \). The only way downwards rigidity could fail is if \( x_{j}^{i} \in D_{A}^{(q,l)} \) for some \( l > i \), i.e. \( \{x_{j}^{i}\} \succ \{x_{j}^{q}\} = \inf(D_{A}^{(q,l)}) \). Thus, towards contradiction, let us allege that there is such a pair \( (q,l) \).

Since \( x_{j}^{i} \notin D_{A}^{(q,l)}(0) \), there is a first time \( n \) such that \( x_{j}^{i} \in D_{A}^{(q,l)}(n) \setminus D_{A}^{(q,l)}(n-1) \). Find \( z \in D_{A}^{(q,l)}(n-1) \) such that \( x_{j}^{i} \in I_{A}(z) \). Note that \( \{x_{j}^{i}\} \succ \{z\} \) and, moreover, \( x_{j}^{i} \in I_{A'}(z) \) for some \( A' \subseteq A \). This contradicts the fact that \( u(x_{j}^{i}) = u_{\min}(x_{j}^{i}, A) \). Hence, downwards rigidity also holds.

Reflexivity: We check that the local category model constructed above, \( (u, \{C_{A}\}) \), is reflexive. This means that the set \( \Theta(A) := \cup_{C_{A} \in C_{A}} \inf(C_{A}) \) equals the set of subjectively feasible elements under the menu preference, \( \succeq_{(u, \{C_{A}\})} \), induced by the local category model. From the construction we see that \( \Theta(A) = \{x \in A : u_{\min}(x, A) = u_{\min}(x, \{x\})\} \), where \( u_{\min}(x, A) := \min\{U(A') : A' \subseteq A, x \in A'\} \). Hence, being subjectively feasible is equivalent to the statement that \( u_{\min}(x, A) = u_{\min}(x, \{x\}) \) – which implies reflexivity.

\[\Box\]

Proof of Theorem 3. In the forthcoming proof we will commit an abuse of notation and denote nodes of the tree category, i.e. menus, and elements of menus both with lower case Roman letters, e.g. \( x, y \), etc. Menus (or, equivalently, nodes) will be bolded to denote the difference between an element and a set of elements. The reason for this is that we will also need to introduce notation for sets of nodes and wanted to avoid introducing a third hierarchy of notation. The argument follows two steps, resp. labeled (i) embedding and (ii) pruning. The first step shows that every category model \( (u, C) \) that is sharp is a prolongation of some sharp sub-category of the tree category – hence, it can be embedded in the tree category. Moreover, it
shows that the prolongation is of the form prescribed in the theorem. Each node in the category $\mathcal{C}$ corresponds one-to-one with a terminal node in the tree category and this terminal node is a lower bound order interval of the corresponding node in the category $\mathcal{C}$. The second step shows that there is a unique sharp subcategory of the tree category.

**Step 1: Embedding.**
For each $x \in X$ consider $\Sigma'(x) = \{ C' \in C_1 : x \in C', x \in \sup(C') \}$. If $\Sigma'(x) \neq \emptyset$, with labeling determined by the $x$-tree construction, let $A_1(x), \ldots, A_k(x)$ be the attraction sets associated with $x$. We reconstruct a set of paths through the $x$-tree whose associated set of terminal nodes is contained in nodes in the set $\Sigma'(x)$. Introduce some notation. Let $\Phi$ be the set of all paths in the $x$-tree and for each $\ell \in \Phi$, let $\ell^{-1}(x_1(1))$ denote the set of all terminal nodes which have $\{ x, x_1(1) \}$ as a predecessor node.\footnote{Recall that when we put $x$ in the node, this is shorthand for $\{ x \} \cup B_1(x)$.}

Similarly, for any (partial) path $\ell(1, n)$ from level 1 to level $n$, let $\ell^{-1}(\ell(1, n))$ denote the set of all terminal nodes whose associated paths through the tree all share the initial segment $\ell(1, n)$. Consider the attraction set $A_1(x)$ and an element $x_1(1) \in A_1(x)$. We do not know yet that the sets in $\Sigma'(x)$ correspond to terminal nodes in the $x$-tree – this will be our conclusion. Nevertheless, we will apply the notation $\ell^{-1}(\ell(1, n))$ to the sets in $\Sigma'(x)$. The meaning is the following: For $x_1(1) \in A_1(x)$ we take $\ell^{-1}(x_1(1))$ to be the set of all elements (terminal nodes) in $\Sigma'(x)$ that contain the element $x_1(1)$. Similarly, for any $x_2(i) \in A_2(x)$ let $\ell^{-1}(x_1(1), x_2(i))$ denote all sets in $\Sigma'(x)$ which contain both $x_1(1)$ and $x_2(i)$. Note that, by representability, every node in $\Sigma'(x)$ is in $\ell^{-1}(x_1(i))$ for some $x_1(i) \in A_1(x)$ (there may be some nodes contained in $\ell^{-1}(x_1(i))$ for more than one $i$). Thus, we have a correspondence $\gamma_1 : A_1(x) \Rightarrow \Sigma'(x)$ given by $x_1(j) \mapsto \ell^{-1}(x_1(j))$. Now iterate this process. For each $x_2(i)$ consider the set $\ell^{-1}(x_1(j), x_2(i))$ and note that, by representability, every node in $\Sigma'(x)$ lies in $\ell^{-1}(x_1(j), x_2(i))$ for some pair $x_1(j), x_2(i)$. Consider the correspondence $\gamma_2 : A_1(x) \times A_2(x) \Rightarrow \Sigma^{-1}(x)$ given by $(x_1(i), x_2(j)) \mapsto \ell^{-1}(x_1(i), x_2(j))$. Inductively construct correspondences, $\gamma_n : \prod_{j=1}^{n} A_j(x) \Rightarrow \ell^{-1}(\ell(1, n))$. Note that (after $k$ levels) for the next step of the $x$-tree construction we attach an $x_i(j)$-tree for some $x_i(j) \in A_i(x)$. For any partial path $\ell(1, k)$ such that $\ell^{-1}(\ell(1, k)) \neq \emptyset$ we can extend via the same procedure as above to obtain an extension of the path through the $x_i(j)$-tree. Observe that each path that passes through a node that contains $x_i(j)$ extends by representability – there must be an element of any attraction set associated to $x_i(j)$ contained in the node. Also note that if $x_i(j) \not\in x(k)$ for some $x(k) \in \ell^{-1}(\ell(1, k))$ then this path is extended without branching until the terminal level of the $x_i(j)$-tree. Inductively proceeding we obtain a collection of paths in the $x$-tree.

Note that every node $x' \in \Sigma'(x)$ has the property that $x' \in \ell^{-1}(1, k), \forall k \leq k_\ell$
for some path in the \( x \)-tree of, say, length \( k \). Fix a node \( x' \in \Sigma'(x) \) and let \( \{\ell_1, \ell_2, \ldots, \ell_n\} \) be an enumeration of paths in the \( x \)-tree (let \( k_i \) denote the length of path \( \ell_i \)) such that \( x' \in \ell_i^{-1}((\ell(1), k)), \forall k \leq k_i \) and for all paths \( \ell_i \). Also let \( x(i) \) denote the terminal node (in the \( x \)-tree) of path \( \ell_i \). It follows that any node \( x' \in \Sigma'(x) \) contains some terminal nodes \( x(i) \) in the \( x \)-tree. We verify that (by retracting if necessary) any \( x' \) in a sharp model \((u, C)\) contains precisely one such \( x(i) \). Since \((u, C)\) is a sharp representation, for each set \( C_i \in C \) one of the following must be true. Either

i. There is a menu \( A \) for which \( \arg \max_{C_j \in C} \min_{z \in C_j \cap A} u(z) = C_i \), or

ii. \( C_i \not\subseteq \cup_{j \neq i} C_j \).

In either case, we say that “the maximum occurs at \( x \)” on \( C_i \) if under case (i) the maximum of the function \( U^C(A) \) occurs at \( x \in C_i \) for some menu \( A \), or if \( x \not\in \cup_{x' \not\subseteq C} x'' \) (in which case we can take the menu that supports this as a maximum to be \( A = \{x\} \)). Assume now that the maximum occurs at \( x = \sup(x') \). We claim that there is only one \( x(i) \subseteq x' \). Else, if there is a distinct pair \((x(i), x(j))\) with \( x(i), x(j) \subseteq x' \) find \( z \in x(j) \setminus x(i) \) (or \( z \in x(i) \setminus x(j) \) if \( x(j) \subseteq x(i) \)). Consider the menu \( A \cup \{z\} \). Since \( z \not\in x(i) \) and \( A \sim \{x\} \) (and \( x(i) \) is a terminal node in the tree category \((u, T)\) – which represents \( \succeq \)) we must have \( A \cup \{z\} \sim \{x\} \). OTOH, \( U^C(A \cup \{z\}) < u(x) \) – contradicting the hypothesis that \((u, C)\) represents \( \succeq \) as well. Hence, there is no such pair \((x(i), x(j))\).

Notice that exactly the same argument shows that the node \( x(i) \) coincides with \( x' \). Now assume the maximum occurs not at \( x \), but for some \( z_x \in x' \) where \( \{x\} \succ \{z_x\} \) (i.e. in case (i) holds, or \( z_x \not\in \cup_{x'' \subseteq C} x'' \) in case (ii) holds). Take \( z_x \) to be the \((\succeq)\)-\( x \)-maximal such element. Replace \( x' \) in the category \( C \) with the lower bound order interval, \( (\neg \infty, z_x] \cap x' : = x_* \). Note that the category \( C' := C \setminus x' \cup x_* \) (i.e. \( C' \) is obtained from \( C \) by deleting node \( x' \) and replacing it with \( x_* \), all other nodes are left intact) also represents \( \succeq \). Now apply the preceding argument with \( z_x \) replacing \( x \) to find a unique terminal node \( z_x(i) \) in the \( z_x \)-tree contained in \( x_* \subseteq x' \). The preceding argument shows that \( z_x(i) = (\neg \infty, z_x] \cap x' \). Hence, \( z_x(i) \) is a lower bound order interval in \( x' \). \(^{42}\) This shows that every node \( x' \in C \) contains a terminal node of the tree category \( T \) as a lower bound order interval. To complete the argument we need that no two nodes \( x', x'' \in C \) map to the same terminal node (under the mapping \( x \mapsto (\neg \infty, z_x] \cap x \) described above) of the tree category as a lower bound order interval. However, this is a consequence of sharpness. If two nodes map to the same \((\neg \infty, z_x] \cap x \), then both cannot be relevant for the representation since for any menu where the maximum \( U^C(A) \) occurs on the node \( x \) (resp. \( x' \)), it must occur on the common lower bound order interval \((\neg \infty, z_x] \cap x \). Thus, not both

---

\(^{42}\)The same argument for \( x' \) shows that the node \( x_* \) must contain some terminal node of the \( z_x \)-tree since the model \((u, C')\) represents \( \succeq \).
of $x, x'$ can be relevant for the representation.

This concludes the argument for the embedding step. Let us summarize what has been shown so far. Starting with any sharp model $(u, C)$ we have shown that this model is a prolongation of a sharp model $(u, C')$, where the latter model is a sub-category of the tree category $\mathcal{T}$. Moreover, each node, $C'_i$, in $C'$ is a lower bound order interval for a unique $C_j$ in $C$. We can further retract $(u, C')$ to a (sharp) sub-model of the tree category with the property that for each node $C'_i \in C'$ the $(\succeq)_{|\mathcal{X}}$-maximal element for which the maximum of the function $U_{C'}(\cdot)$ is attained on $C'_i$ (for some menu $A$) is $\sup(C'_i)$.\footnote{This follows from identical reasoning as given in the preceding paragraph for the embedding argument.} Formally, we take

$$\sup\{x_A \in C'_i : (i) \; A \sim \{x_A\}, \; (ii) \; C'_i = \arg \max_{C' \in C'} \min_{x \in C' \cap A} u(z)\}$$

and reduce (by the preceding argument) to nodes $C'_i$ such that $\sup(C'_i) = \sup\{x_A : (i) \; A \sim \{x_A\}, \; (ii) \; C'_i = \arg \max_{C' \in C'} \min_{x \in C' \cap A} u(z)\}$. Put $x_C$ equal to this common element and note that $C'_i$ must be a minimal (w.r.t. set inclusion) node in the $x_C$-tree. Now turn to the next step.

**Step 2: Pruning.**

Fix the tree category representation of $(u, \mathcal{T})$ of $\succeq$. Take any two sharp sub-categories $(u, C_1), (u, C_2)$ which also represent $\succeq$ and have the property that for each $C'_i \in C_1$ (resp. $C'_i \in C_2$) the maximum of $U^C(\cdot)$ (resp. $U^{C_2}(\cdot)$) for some menu $A$ is attained on $\sup(C'_i)$ (resp. $\sup(C'_i)$). By the preceding remarks, this implies that all nodes in $C_1, C_2$ are minimal nodes of the respective $x$-trees to which the nodes belong. We show that $C_1 = C_2$ by downwards induction on $\mathcal{T}_x$-trees. That is, we inductively show that all minimal nodes in the $\mathcal{T}_x$-tree that are present in $C_1$ must also be in $C_2$ and vice-versa. Start with the $(\succeq)_{|\mathcal{X}}$-maximal $x$-tree. We claim that for maximal $\mathcal{T}_x$-trees all minimal terminal nodes are present in both $C_1$ and $C_2$. Towards contradiction, assume there is a minimal terminal node, say $x_*$ that is not in the sub-category $(u, C_1)$. Let $\{x_1, \ldots, x_N\}$ be an enumeration of the minimal nodes in the $\mathcal{T}_x$-tree with $x_1 \equiv x_*$. For each $x_i (\neq x_1)$ pick some element $z_i \in x_i \setminus x_1$ (by minimality of $x_*$). Consider the menu $A' := \{z_i : z_i \in x_i\} \cup \{x\}$. Since the $\mathcal{T}_x$-tree is unobstructed at $A'$ we must have $A' \sim \{x\}$. OTOH, consider the value of the function $U^{C_1}(A')$ (where $U^{C_1}(\cdot)$ denotes the menu utility generated by $(u, C_1)$). Note that the only nodes in the category that contain $x$ are the terminal nodes from the $x$-tree (by $(\succeq)_{|\mathcal{X}}$-maximality). Since $x_* \not\in C_1$ and $\{x\} \succ \{z_i\}$, where $z_i \in x_i$, every node in $C_1$ that contains the singleton $x$ is obstructed by some $z_i$. It follows that $U^{C_1}(A') < u(x)$, which contradicts representability since the $\mathcal{T}_x$-tree is unobstructed at the menu $A'$. It follows that all minimal terminal nodes for a $(\succeq)_{|\mathcal{X}}$-maximal
x-tree are present in the model \((u,C_1)\) (and resp. \((u,C_2)\), by the identical argument).

Similar reasoning shows that all minimal terminal nodes are present for any non-embedded \(T_x\)-tree.\(^{44}\) Accordingly break up the \(x\)-trees into two groups: group I is the set of non-embedded \(x\)-trees and group II is the set of embedded \(x\)-trees. Copying the preceding argument we check that \((u,C_1)\) and \((u,C_2)\) agree on all trees in group I, i.e. all nodes in \(C_1\) from group I trees are in \(C_2\) and vice-versa. Next, we induct downwards on \(\geq \) \(|x|\)-rank to show that all nodes in \((u,C_1)\) that come from group II trees are also in \((u,C_2)\) and vice-versa. Enumerate group II trees via \(\{x_1\} \succ \cdots \succ \{x_n\}\) and assume we have shown agreement of nodes in \(C_1\) and \(C_2\) on all \(T_x\)-trees for \(1 \leq i \leq k\). Consider the \(T_{x_{k+1}}\)-tree and, towards contradiction, say that \(x_*\) is a node in \((u,C_1)\) that is not in \((u,C_2)\). Consider each \(T_x\)-tree in which the \(T_{x_{k+1}}\)-tree is embedded. Consider the nodes in this tree which contain the node \(x_*\) and which are present in \(C_2\). Formally denote this as:

\[
\kappa_i(x_*) := \{y_i \in C_2 : x_* \subseteq y_i\}.
\]

We claim that \(\kappa_i(x_*)\) is non-empty for some \(x_i\) in which the \(T_{x_{k+1}}\)-tree embeds. Proceed via contradiction. Consider all the minimal nodes of the \(x_{k+1}\)-tree and for each such node \(x_i \neq x_*\) find \(z_i \in x_i \setminus x_*\). Now note that, for any \(x_i\)-tree in which the \(x_{k+1}\)-tree embeds, any terminal node of this tree which contains \(x_{k+1}\) contains a node from the \(x_{k+1}\)-tree. If \(\kappa_i(x_*) = \emptyset\) for all \(x_i\)-trees in which the \(x_{k+1}\)-tree embeds, then each of these terminal nodes contains one of the nodes \(x_i \neq x_*\) from the \(x_{k+1}\)-tree. Now consider the menu \(A' := \{z_i \in x_i : x_i \neq x_*\} \cup \{x\}\). Note that \(x_{k+1}\) is unobstructed at \(A'\). However, the category \(C_1\) is obstructed since we are alleging that \(x_* \notin C_1\) and that \(\kappa_i(x_*) = \emptyset, \forall i\) - contradiction. It follows that \(\kappa_i(x_*)\) is non-empty for some \(x_i\)-tree in which the \(x\)-tree embeds.

Now pick any \(y_i \in \kappa_i(x_*)\). By the induction hypothesis, since \(y_i \in C_2\) we also have \(y_i \in C_1\). Since \(C_1\) is sharp, \(x_*\) cannot be a lower bound interval in \(y_i\) (as deleting \(x_*\) from \(C_1\) would then yield a representation of the same menu preference, implying that \(x_*\) is not relevant for the representation – contradicting sharpness). Hence, for each such \(y_i\) we can find \(z^i \in y_i \setminus x_*\) with \(\{x_{k+1}\} \succ \{z^i\}\) and do this for every \(y_i \in \kappa_i(x_*)\) and for every \(i\) with \(\kappa_i(x_*) \neq \emptyset\). Also find \(z_i \in x_i \setminus x_*\) for each minimal node other than \(x_*\) of the \(x_{k+1}\)-tree. Let \(A' := \{z_i : z_i \in x_i \setminus x_*\} \cup \{x_{k+1}\}\). Also put

\[
A_i = \{z^i : \{x_{k+1}\} \succ \{z^i\}, z^i \in y_i \setminus x_*, y_i \in \kappa_i(x_*)\}.
\]

Put these all together to construct the menu:

\[
\hat{A} := A' \cup_i A_i.
\]

\(^{44}\)By a “non-embedded” \(T_x\)-tree we mean any \(T_x\)-tree such that \(x\) does not appear in an attraction set \(A(x')\) for some \(x' \succ \{x\}\).
Note that the $x_{k+1}$-tree is unobstructed at $\hat{A}$, so that $\hat{A} \sim \{x_{k+1}\}$. OTOH, computing the value of $U^{C_2}(\hat{A})$ note that the embedded image of the node $x_*$ in any $x_{k}$-tree is obstructed by $\hat{A}$ and, by construction, the nodes in $C_2$ from the $x_{k+1}$-tree are also obstructed by $A$. It follows that $U^{C_2}(\hat{A}) < u(x_{k+1})$—contradicting representability. Thus, $x_* \in C_2$, so that all nodes from the tree $T_{x_{k+1}}$ that are in $C_1$ are also in $C_2$. By symmetry, all nodes from this tree that are in $C_2$ are also in $C_1$—completing the inductive step. It follows that $C_1 \equiv C_2$. \hfill $\Box$

\textit{N.B:} The argument shows that there is a unique, sharp sub-category of the tree category with the property that (i) all nodes are minimal and (ii) the highest value for which a strict maximum of the function $U^C(\cdot)$ occurs on each of these nodes is the $(\geq)|_X$-maximal singleton in the node. Note that it is easy enough to construct such a category from scratch. Namely, start with the tree category. Pass to the sub-category of all minimal nodes—which also yields a representation. Next, pass to any sharp sub-category. Note that the above argument already shows that all such sub-categories are prolongation of some $(u,C_*)$. To find $C_*$, go node by node and take $x_{C_*}$ to be the maximum singleton for which the maximum is attained on node $C_*$. If $x_{C_*} \neq \sup(C_*)$, then replace this node with $(-\infty,x_{C_*}] \cap C_*$. The same arguments as presented above show that this must be a minimal terminal node of the $x_{C_*}$-tree. Inductively proceed until we obtain a sharp sub-category consisting entirely of minimal nodes with $x_{C_i} = \sup(C_i)$ for all sets $C_i$ in the sub-category.

### 6.2 Proofs for Section 3

\textit{Proof of Lemma 2.} That Strong CRW implies DSB is obvious. We check the converse. Consider the set $\Sigma(A) = \{x \in A : \{x\} \sim A\}$. We claim that there is some $x_A \in A$ such that $\forall A' \subseteq A$ with $x_A \in A'$ we have $A' \sim A$. Otherwise, for each $x$ such that $\{x\} \sim A$ there is a subset $A(x) \subseteq A$ with $x \in A(x)$ and $A \neq A(x)$. Put $\hat{A} := A \setminus \cup_{x \in \Sigma(A)} A(x)$ and note that $A = \cup_{x \in \Sigma(A)} A(x) \cup \hat{A}$. By iterative application of DSB, $A \sim \hat{A}$ which implies (by iterative application of DSB again) that $\exists x \in \hat{A}$ such that $\{x\} \sim A$. On the other hand, $x \in A(x)$ and $A(x) \cap \hat{A} = \emptyset$—contradiction. \hfill $\Box$

\textit{Proof of Observation 1.} Fix $(u,v)$ and for each $x$ consider the set $C_x := \{x\} \cup \{y : v(y) > v(x), u(x) \geq u(y)\}$. Let $C := \{C_x\}_{x \in X}$. Let $U^C$ denote the menu utility generated by $(u,C)$ and let $U$ denote the menu utility generated by $(u,v)$ via the Strotz formula. Fix any menu $A$ and let $x \in A_u$ be a $u$-maximal element. Note that $U^C(A) = u(x)$. Moreover, $C_x \cap A = \{x\}$ so that $U^C(A) \geq u(x)$. For the reverse inequality, take any $z \in A$ with $u(z) > u(x)$. Then, $z \notin A_v$ so that there is some $y \in A_v$ with $v(y) > v(z)$. Since $u(z) > u(x)$ and $x$ is $u$-maximal in $A_v$ we have $u(z) > u(x) \geq u(y)$, so that $y \in C_z$. Hence, $U^C(A) \leq u(x)$. It follows that $U^C(A) = U(A)$. \hfill $\Box$
Proof of Theorem 4. We first check the representation claim, then the minimality claim. We first verify necessity of Axiom 6 (NAT) and Axiom 2* (Reduction). To check Reduction, take \( x \to_t y \) and note that if \((u, C)\) represents \( \succeq \), then whenever \( x \in C \), we must also have \( y \in C \). It follows that \( A \setminus x \sim A \). To check Axiom 6 (NAT), it suffices to show that if \( A \) is a pairwise feasible menu, then for each \( x \in A \) there is some \( C_x \in C(x) \) with \( ((-\infty, x) \cap C_x = \emptyset) \). Since \( C \) is a NAT category, find \( C_x \) such that \((-\infty, x) \cap C_x = \emptyset \). Thus, \( (A \setminus x) \sim \emptyset \). Axiom 6 follows. Now we turn to the sufficiency argument. For each \( x \in X \) consider the set \( C_x := B_t(x) \cup \{x\} \) and let \( C \equiv \{C_x : x \in X\} \). Define \( U(A) := \max_{C \in C} \min_{z \in C \cap A} u(z) \). We claim that (i) \( C \equiv \{C_x\}_x \) is a NAT category and (ii) \( U(\cdot) \) represents \( \succeq \). For the first claim let \( x \in C_y \cap C_z \) and note that \((-\infty, x) \cap C_y = B_t(x) \). Moreover, if \( x \in B_t(y) \cap B_t(z) \), then \( B_t(x) \subseteq B_t(y) \cap B_t(z) \) (by the argument given in the text). Thus, \( C_x \subseteq C_y \cap C_z \) so that the collection \( \{C_x\}_x \) is a NAT category. Now we check representability. Given a menu \( A \), let \( A^* \) denote the pairwise feasible subset of \( A \), viz. \( A^* := \{x \in A : x \not\sim y, \forall y \in A\} \), and note that \( A \sim A^* \). Note that \( U(A) = U(A^*) \), so that it suffices to check representability on the set of pairwise feasible menus. But this is obvious since for any \( x \in \operatorname{sup}(A) \) we have \( C_x \cap A = \{x\} \). Thus, \( U(A) = u(x) \), where \( \{x\} \sim A \). Now for the minimality claim. For each \( x \in X \) put \( \Sigma(x) := \{C' : x \in C' \} \). Let \((u, C)\) denote any category that represents the preference and, towards contradiction, assume there is some \( x \) such that for every \( C' \in \Sigma(x) \) we have \((-\infty, x) \cap C' \not\subseteq \emptyset \), for some \( C'' \in \Sigma(x) \). For each \( C' \), select some \( z' \in \{-\infty, x\} \cap C' \setminus C'' \) and note that, by representability, \( z' \not\in B_t(x) \). Let \( A := \{z' \in C' : C' \in \Sigma(x)\} \cup \{x\} \) and note that \( A \sim \{x\} \). On the other hand, we have \( U(A) < u(x) \) - a contradiction. \( \square \)

Proof of Proposition 1. Necessity of Axiom 5: Let \((u, C)\) be a representing narrow category and say that \( x \succ \{y\} \succ \{z\} \) with \( x \to_t z \) and \( y \not\sim_t z \). Towards contradiction, allege that \( x \to_t y \). Note that this implies that whenever \( x \in C_i \), then we must have \( y \in C_i \) and \( z \in C_i \). We now show that \( z \in C_j \) whenever \( y \in C_j \). To this end, let \( C_j \) be a set that contains \( y \). Let \( C_i \) contain \( x \) and note that we must have \( y, z \in C_i \). Since the category is narrow we must then have, \( \{z' \in C_i : y \succ z'\} = \{z' \in C_j : y \succ z'\} \). Thus, \( z \in C_j \) so that \( y \to_t z \) - a contradiction.

Sufficiency: Let \( C \equiv \{B_t(x) \cup \{x\}\}_{x \in X} \) denote the NAT category constructed in the proof of Theorem 4. We know that the corresponding model \((u, C)\) represents \( \succeq \). We check that the NAT category is, in fact, narrow. To see this, put \( C_x = B_t(x) \cup \{x\} \) and let \( z \in C_x \cap C_y \). Either \( \inf(x, y) \succ \{z\} \) or \( z = \inf(x, y) \). In the former case (the

\[\text{To see this, successively delete (from the top-down) infeasible elements from the menu} \ A \ \text{until we reach the menu} \ A^* \). Applying Reduction at each step we obtain a string of indifferences which together yield \( A \sim A^* \).
other one is similar, hence we omit the proof), say that \( \{x\} \succ \{z\} \) and consider the set \( \{z' \in C_x : \{z\} \succ \{z'\}\} \). Thus, for each such \( z' \) we have \( \{x\} \succ \{z\} \succ \{z'\} \) and \( x \rightarrow_t z, x \rightarrow_t z' \). By Axiom 5 (Cycles) we obtain \( z \rightarrow_t z' \). Since \( z \in C_y \) and \( z \neq y \) we obtain that \( y \rightarrow_t z' \). It follows that \( z' \in \{z' \in C_y : \{z\} \succ \{z'\}\} \). Similar reasoning shows the inclusion \( \{z' \in C_y : \{z\} \succ \{z'\}\} \subseteq \{z' \in C_x : \{z\} \succ \{z'\}\} \). Thus, the NAT category is narrow.

**Proof of Proposition 2. Necessity:** First, we verify necessity of the No Cycles axiom. Fix \( \{x\} \succ \{y\} \succ \{z\} \) and put \( x \rightarrow_t z, y \not\rightarrow_t z \). We claim that \( x \rightarrow_t y \). Proceed via contradiction. Let \( C_i \) be any set in \( C \) that contains \( x \) and which omits \( y \) (note that at least one such set exists if \( x \not\rightarrow_t y \)) and let \( C_j \) be any other set that contains \( y \). Since \( C \) is Strotzian this implies that \( (-\infty, y) \cap C_i \subseteq C_j \). Since \( C_j \) was any other category that contains \( y \) it follows that \( (-\infty, y) \cap C_i \subseteq C_j, \forall C_j \) s.t. \( y \in C_j \). Hence, \( y \rightarrow_t z \) – a contradiction. We next need to show that Strotzian preferences are characterized by Axiom 2\(^*\) (Reduction), Axiom 4 (No Cycles), and Axiom 6 (NAT). Since Axiom 1\(^*\) clearly implies Reduction, No Cycles, and NAT, we check the converse. Let \( x \in A \) be such that \( \{x\} \sim A \) (note: the existence of such an \( x \) is guaranteed by NAT and Reduction). We claim that any \( x \in A \) with this property is automatically a Strotzian choice. Towards contradiction, find a subset \( A(x) \subseteq A \) such that \( x \in A(x) \) and \( A(x) \succ \{x\} \). \(^{46}\) Let \( z_x \in A(x) \) be such that \( \{z_x\} \sim A(x) \). Note that \( z_x \) is not pairwise feasible in \( A \). Let \( z \in A \) be such that \( \{x\} \succ \{z\} \) and \( \{z\} \succ \{z_x, z\} \sim \{z\} \). \(^{47}\) Consider the menu \( \{x, z_x, z\} \). Note that, by Reduction and NAT, we must have \( \{x, z_x, z\} \sim \{x\} \). By Axiom 4 (No Cycles), this implies \( z_x \rightarrow_t x \) – contradicting the hypothesis that \( z_x \) is pairwise feasible in \( A(x) \). It follows that Reduction, No Cycles, and NAT are equivalent to Strotzian choice, completing the proof of necessity.

**Sufficiency:** Let \( (u, C) \) denote the NAT category constructed in the sufficiency proof of the representation. We check that this category is Strotzian. Let \( C_z = \{z\} \cup B_t(z) \) denote the sets in the category and take \( z_1, z_2 \) with \( \{z_1\} \succ \{z_2\} \). Assume \( z_2 \notin C_z \). If \( (-\infty, z_2) \cap C_z = \emptyset \) there is nothing to show. Hence, let \( z' \in (-\infty, z_2) \cap C_z \). If \( z' \notin C_z \), then \( z_2 \not\rightarrow_t z' \). Hence, by No Cycles, \( z_1 \rightarrow_t z_2 \rightarrow_t z' \) – contradiction. It follows that \( (-\infty, z_2) \cap C_{z_1} \subseteq C_{z_2} \). This shows the Strotz condition on the categories when comparing pairs \( (C_{z_1}, C_{z_2}) \) with \( \{z_1\} \succ \{z_2\} \). Now consider an arbitrary \( (x, y) \) with \( \{x\} \succ \{y\} \). Find \( C_{z_x}, C_{z_y} \) with \( x \in C_{z_x}, y \in C_{z_y} \setminus C_{z_x} \). If \( (-\infty, y) \cap C_{z_x} = \emptyset \), there is nothing to show. Hence, consider \( z' \in (-\infty, y) \cap C_{z_x} \) and note that we have \( z_x \rightarrow_t z' \). The No Cycles axiom implies that \( y \rightarrow_t z' \), so that \( z' \in C_{z_y} \) (note: since \( \{x\} \succ \{y\} \) and \( x \in C_{z_x} \), we have \( \{z_x\} \succ \{y\} \)). It follows that \( (-\infty, y) \cap C_{z_x} \subseteq C_{z_y} \).

---

\(^{46}\) Since NAT and Reduction imply a category representation, we know that these axioms imply CRW.

\(^{47}\) Note that there exists a \( z \) with these two properties, else we would have \( A \succ \{x\} \).
Redundance of Property (3): We check that condition (2) implies condition (3) when the singleton ranking is strict. Assume that \((u, C_{ST})\) is a category model which satisfies: \(x \in C_i, y \in C_j \setminus C_i \Rightarrow (\infty, y) \cap C_i \subseteq C_j\). We check that this implies the category is a NAT category. Let \(\Sigma(x) := \{C_i \in C_{ST} : x \in C_i\}\). We claim that the collection \(\{C_i \cap (\infty, x) : C_i \in \Sigma(x)\}\) has a unique minimal element (w.r.t. set inclusion). Otherwise, let \(C_1, C_2\) be a pair such that \((\infty, x) \cap C_1 \not\subseteq C_2\) and \((\infty, x) \cap C_2 \not\subseteq C_1\). Let \(z_1 \in C_1 \cap (\infty, x) \cap C_2\) and similarly define \(z_2\). Say that wlog \(\{z_2\} \succ \{z_1\}\). Then, by property (2) we obtain \((\infty, z_2) \cap C_1 \subseteq C_2\), which implies \(z_1 \in C_2\). Hence, \((\infty, x) \cap C_1 \subseteq C_2\), implying that there is a set \(C_x \in \Sigma(x)\) such that \((\infty, x) \cap C_x \subseteq C_1, \forall C_i \in \Sigma(x)\). It follows that \((u, C_{ST})\) is a NAT category model. \(\square\)

Proof of Theorem 5. Let \((u, C_1)\) be a NAT category representation of \(\succeq\). We check that \(C_1 \equiv B^*\). Proceed in three steps. First, we show that \(C_1\) is a prolongation of \(B^*\). Second, we show that it is a trivial prolongation, i.e. \(B^* = C_1\). Third, check that \(B^*\) yields a sharp representation of \(\succeq\).

**Step 1:** Put \(C_x := B_t(x) \cup \{x\}\) and let \(C^* \equiv \{C_x\}_{x \in X}\). Let \(\Sigma_1(x) := \{C' \in C_1 : x \in C'\}\). Note that, by representability, \(C_x \subseteq C^*, \forall C' \in \Sigma_1(x)\). We claim that there is some \(C' \in \Sigma_1(x)\) such that \(C_x = (\infty, x] \cap C'\). Otherwise, for each \(C' \in \Sigma_1(x)\) find some \(z' \in \{(\infty, x] \cap C') \cap C_x\) and put \(A' := \{z' : z' \in (\infty, x] \cap C' \cap C_x\} \cup \{x\}\) and note that \(U^{C_1}(A') < u(x)\) - a contradiction, since representability and Axiom 2* and Axio6 imply that \(A' \sim \{x\}\). Thus, for each \(C_x\), we can find some \(C_1(x_1) \in C_1\) such that \(C_1(x_1) \cap (\infty, x_1] = C_x\). We first claim that \(C_1 \equiv \{C_1(x_1), \ldots, C_1(x_n)\}\). For each \(x_i\), we have \(C_{x_i} \subseteq C_1(x_i)\). Let \(C_x \equiv \{C_{x_i}\}\) and put \(C^1 \equiv \{C_1(x_1), \ldots, C_1(x_n)\}\) (for pairs \(x_i, x_j\) such that \(C_1(x_i) = C_1(x_j)\) we include just one copy in the category \(C^1\). Observe that \((u, C^1)\) and \((u, C_x)\) generate the same menu utility. Since \((u, C^1)\) represents \(\succeq\) it follows that \((u, C^1)\) represents \(\succeq\), implying \(C^1 \equiv C_1\) by sharpness. We now claim that the sets \(C_1(x_i)\) are in one-to-one correspondence with the undominated sets of the form \(B_t(x_i) \cup \{x_i\}\). Via the map \(B_t(x_i) \cup \{x_i\} \mapsto C_1(x_i)\), where \(C_1(x_i) \cap (\infty, x_i] = B_t(x_i) \cup \{x_i\}\). Towards contradiction, assume there is a pair \(\{x_i, x_j\}\) and a common \(C_1(x)\) with both \(C_1(x) \cap (\infty, x_i] = B_t(x_i) \cup \{x_i\}\) and \(C_1(x) \cap (\infty, x_j] = B_t(x_j) \cup \{x_j\}\) (say \(\{x_i\} \succ \{x_j\}\)). This implies:

\[
B_t(x_j) \cup \{x_j\} = C_1(x) \cap (\infty, x_j] \\
= (C_1(x) \cap (\infty, x_i]) \cap (\infty, x_j] \\
= (B_t(x_i) \cup \{x_i\}) \cap (\infty, x_j].
\]

This contradicts the assumption that \(B_t(x_j) \cup \{x_j\}\) is undominated. It follows that the category \(C_1\) is a prolongation of the category \(B^*\).

**Step 2:** We now check that \(C_1 = B^*\). By the preceding argument, \(C_1\) is a prolongation of the category \(B^*\). Moreover, observe that we can write \(C_1 = \{C_1(z_1), \ldots, C_1(z_n)\}\),
where $C_1(z_i) \cap (-\infty, z_i] = \{z_i\} \cup B_i(z_i)$ and each set $\{z_i\} \cup B_i(z_i)$ is undominated. Label the sets $\{C_1(z_1), C_1(z_2), \ldots, C_1(z_n)\}$ such that $\{z_1\} \succ \{z_2\} \succ \cdots \succ \{z_n\}$. This implies that $C_1(z_1) = \{z_1\} \cup B_i(z_1)$. Now consider the set $C_1(z_2)$. We claim that $C_1(z_2) = \{z_2\} \cup B_i(z_2)$, so that $\sup(C_1(z_2)) = z_2$. If $\sup(C_1(z_2)) \succ z_2$, then (putting for brevity, $\hat{z} := \sup(C_1(z_2))$) we must have $\hat{z}$ dominated by $z_1$. Hence, $(-\infty, \hat{z}] \cap C_1(z_1) = \{\hat{z}\} \cup B(\hat{z})$. OTOH, $\{\hat{z}\} \succ \{z_2\}$ and $z_2$ is undominated. Hence, $\exists \tau' \in (-\infty, z_2] \cap C(z_1)$ with $\tau' \not\in B(z_2)$. Since $C_1(z_2) \cap (-\infty, z_2] = \{z_2\} \cup B_i(z_2)$ this implies that $\tau' \not\in C_1(z_2)$ — contradiction, since $\tau' \in B(\hat{z})$. Inductively proceeding we obtain $C_1(z_i) = \{z_i\} \cup B_i(z_i)$.

**Step 3:** It remains to check that $(u, B^*)$ is a sharp representability of $\succeq$. Representability is trivial given the sufficiency proof of Theorem 4, so that the only issue is sharpness. Towards contradiction, say that there is some undominated set $B_i(x_i) \cup \{x_i\}$ which is not relevant. Then, we must have $X = \cup_{j \neq i} (B_i(x_i) \cup \{x_i\})$. Since $B_i(x_i) \cup \{x_i\}$ is not dominated by any of these sets, for each $j \neq i$ with $x_i \in B_i(x_i) \cup \{x_i\}$ find $b_j \in (B_i(x_i) \cap (-\infty, x_i)) \setminus (B_i(x_i) \cup \{x_i\})$ and consider the menu $A = \{b_j\}_j \cup \{x_i\}$. Note that $A \sim \{x_i\}$, by Axiom 2* and Axiom 6. OTOH, evaluating the function $U^C(A)$, where $C := B^* \setminus (B_i(x_i) \cup \{x_i\})$, we get $U^C(A) < u(x_i)$ — contradiction.

**Proof of Observation 2.** We first check that

$$U(A_x) \geq \min \{ \min_{z \in A_1(x)} u(z), u(x) - (\max_{z \in A_2(x)} v(z) - v(x)) \}.$$

Consider the GP kernel, $u(z) - (\max_{w \in A_x} v(w) - v(z))$, and break into two cases, (i) the maximum occurs on $x$ and (ii) the maximum occurs away from $x$, say at some $y \neq x$ (so that $u(y) + v(y) > u(x) + v(x)$). Consider this latter case first and note that $U(A_x) \leq u(y)$. If $U(A_x) = u(y)$, then we must have $u(y) \leq u(x)$ (by definition of $A_x$) so that $y \in A_1(x)$. It follows that $U(A_x) \geq \min_{z \in A_1(x)} u(z)$. Next, note that $U(A_x) < u(y)$ implies that $\exists y' \in A_x$ with $u(y) + v(y) \geq u(y') + v(y')$ and $v(y') = \max_{z \in A_x} v(z) > v(y)$. Since $u(y) + v(y) > u(x) + v(x)$, we obtain that $u(y') - (v(y') - v(y)) > u(x) - (v(y') - v(x))$. Note that $u(y) \leq u(x)$, else $U(A_x) = u(y) - (v(y') - v(y)) \geq u(y') > u(x)$ (the inequality since $y$ is chosen from $A_x$). But $U(A_x) \leq u(x)$ by definition of $A_x$, a contradiction. If $y \in A_1(x)$, then $u(y') + v(y') > u(x) + v(x)$ so that $U(\{x, y\}) = u(y')$. Moreover, $u(y') - (v(y') - v(y)) \geq u(y')$. Hence, $\{y\} \sim \{y, z\} \not\leq A_x$. It follows that, in this case as well, $U(A_x) \geq \min_{z \in A_1(x)} u(z)$. To finish this case, we now claim that $y \not\in A_2(x)$. Else, $x$ is chosen from $\{x, y\}$ so that $U(\{x, y\}) = u(x) - (v(y') - v(x)) < u(y) - (v(y') - v(y)) = U(A_x)$. On the other hand, $\{x, z\} \not\leq A_x$ by definition of the set $A_x$ — a contradiction. Now consider the case where the maximum occurs on $x$, i.e. $x$ is chosen in $A_x$. Note that $U(A_x) \leq u(x)$. If $u(x) = U(A_x)$, then we must have $v(x) = \max_{z \in A_2(x)} v(z)$. Hence, $U(A_x) = u(x) - (\max_{z \in A_2(x)} v(z) - v(x)) \geq \min_{z \in A_1(x)} u(z), u(x) - (\max_{z \in A_2(x)} v(z) - v(x))$. If $U(A_x) < u(x)$, then find
\[ z_x \text{ with } u(x) + v(x) \geq u(z_x) + v(z_x) \text{ and } v(z_x) = \max_{z \in A_x} v(z) > v(x). \] Note that \( z_x \in A_2(x) \) so that \( U(A_x) = u(x) - (v(z_x) - v(x)) \geq u(x) - (\max_{z \in A_2(x)} v(z) - v(x)). \) Hence, \( U(A_x) \geq \min\{\min_{z \in A_1(x)} u(z), u(x) - (\max_{z \in A_2(x)} v(z) - v(x))\} \) in this case as well. We now check the reverse inequality:

\[
U(A_x) \leq \min\{\min_{z \in A_1(x)} u(z), u(x) - (\max_{z \in A_2(x)} v(z) - v(x))\}.
\]

For each \( z \in A_1(x) \) note that \( \{x, z\} \sim \{z\} \). Hence, \( \{z\} \sim \{x, z\} \supseteq A_x \), as \( U(A_x) = \min\{U(A') : A' \subseteq A, x \in A'\} \). Similarly, \( U(A_x) \leq U(A_2(x)) \). Moreover, note that \( U(A_2(x)) = u(x) - (\max_{z \in A_2(x)} v(z) - v(x)) \) as \( x \) is chosen from \( A_2(x) \) (by definition of \( A_2(x) \)). Hence, \( U(A_x) \leq \min\{\min_{z \in A_1(x)} u(z), u(x) - (\max_{z \in A_2(x)} v(z) - v(x))\} \.

Proof of Proposition 3. Let \((u, C)\) admit a \((u, \{v_k\})\) representation. Define the following category \( C_{\text{DLR}} \). For each \( x \in X \) put

\[
C_x := \{x\} \cup \{y \in X : (i) u(x) > u(y), (ii) \sum_k v_k(y) - \sum_k v_k(x) > u(x) - u(y)\}
\]

and put \( C_{\text{DLR}} \equiv \{C_x\}_{x \in X} \) (pass to a sharp sub-category if necessary). We claim that the model \((u, C_{\text{DLR}})\) is (i) Strotzian, (ii) has the property that \( C_{\geq}(A) = \arg\max_{z \in A} u(z) \), and (iii) is the best approximation to \((u, C)\) (within the class of category models). For (i), assume that \( x \in C_{x_i}, y \in C_{x_i} \), where \( u(x) > u(y) \) and \( y \notin C_{x_i} \). Notice that this implies \( u(y) + \sum_k v_k(y) \leq u(x_i) + \sum_k v_k(x_i) \) (note that \( u(x_i) > u(y) \) as \( x \in C_{x_i} \) and \( u(x) > u(y) \)). Hence, for any \( z \in (-\infty, y) \cap C_{x_i} \) we have \( u(z) + \sum_k v_k(z) > u(x_i) + \sum_k v_k(x_i) \geq u(y) + \sum_k v_k(y) \) since \( u(y) + \sum_k v_k(y) > u(x_i) + \sum_k v_k(x_i) \) we obtain \( z \in C_{x_i} \). Hence, the model \((u, C_{\text{DLR}})\) is Strotzian. For (ii), let \( x \in C_{\geq}(A) \) and note that this implies that \( A \cap C_x = \{x\} \).

If \( y \in C(A) = \arg\max_{z \in A} u(z) + \sum_k v_k(z) \), then \( A \cap C_y = \{y\} \). Since \( x \in C_{\geq}(A) \) we obtain \( u(x) \geq u(y) \). It follows that \( x \in \arg\max_{z \in C(A)} u(z) \). The reverse inclusion is trivial. For (iii), note that if \((u, C')\) is any other (sharp) category model with \( C_{\geq}(A) \subseteq C_{\geq}(A) \subseteq C(A) \), then (by (iii)) we must have \( C_{\geq}(A) = C_{\geq}(A) \). Hence, the menu preference induced by the two models \((u, C_{\text{DLR}})\) and \((u, C')\) agree. Since \( C_{\text{DLR}} \) is Strotzian, it is a NAT category. Now use the (strong converse) part of Theorem 4 which states that any category model \((u, C)\) that represents a menu preference that satisfies Axiom 2* (Reduction) and Axiom 6 (No Attractions) must itself be a NAT category. Since both \((u, C_{\text{DLR}})\), \((u, C')\) are sharp models, the identification result for NAT categories (Theorem 5) shows that \( C_{\text{DLR}} = C' \).

6.3 Proofs for Section 4

Proof of Proposition 4. Put \( \Theta(\geq) := (u, C) \). Let us check that \((u, C)\) satisfies A1-3. A1 and A3 are clearly satisfied from construction of the pair \((u, C)\). For A2, let us first check the equality: \( A^* = A'_{(u, C)} \). That is, the subjective feasible set
computed relative to $\succeq$ agrees with the feasible set derived from the ex post observables. The left-to-right containment is obvious from definition. Now check right-to-left. Let $x \in A^*_{(u,C)}$ and note that, by definition, if $A' \subseteq A$ s.t. $x \in A'$, then $u(y) \geq u(x), \forall y \in C(A')$ where $C(A') := \arg \max_{x \in (A')} u(x)$. If, towards contradiction, $x \notin A^*$, then there is some subset $A' \subseteq A$ where $x \in A'$ yet $\{x\} \succ A'$. This implies that $x \notin (A')^*$. Moreover, by Axiom 1 we know that $A' \sim \{x_{A'}\}$ for some $x_{A'} \in (A')^*$. Since, by definition, $A' \supseteq \{y\}, \forall y \in (A')^* \text{ we obtain: } \{x\} \succ A' \sim \{x_{A'}\} \succeq y, \forall y \in (A')^*$. Hence, $u(x) > u(y), \forall y \in (A')^*$, implying that $x \notin A^*_{(u,C)}$ – a contradiction. This shows that $\Theta(\succeq)$ satisfies axioms A1-A3.

For the converse, let $(u, C) \in \Sigma(X)$. We want to find a order on menus $\succeq \in \Sigma_{P(X)}$ such that $\Theta(\succeq) = (u, C)$. For each menu $A$, pick any $x_A \in C(A)$ and define a utility on menus via the formula $U(A) := u(x_A)$. By A1, this is a well-defined function on menus. Let us check that it satisfies Axiom 1 (CRW) and Axiom 2 (Strong Reduction). Consider $A' \subseteq A$ with $x_A \in A'$. Since $x_A \in C(A)$, this implies $C(A) \cap A' \neq \emptyset$. By A3, this implies that $u(x) \geq u(y), \forall x \in C(A'), \forall y \in C(A)$. Hence, $U(A') = u(x_{A'}) \geq u(x_A) = U(A)$, implying Axiom 1 (CRW). To check Axiom 2 (Strong Reduction) consider two sets, $A^*_{(u,C)}$ and $A^*$. The latter is the subjective feasible set computed w.r.t. the induced menu preference $\succeq$ (defined via $U(A) = u(x_A)$). The former is the subjective feasible set defined using the observables $(u, C)$. Notice that if $(u, C)$ satisfies A1 – A3, then these two sets are identical. Axiom 2 (Strong Reduction) then follows immediately. Hence, $\succeq \in \Sigma_{P(X)}$. To finish, we verify that $\Theta(\cdot)$ is a bijection. Since we have already checked surjectivity it suffices to verify that $\Theta$ is injective. Towards contradiction, say that $\succeq, \succeq' \in \Sigma(X)$ map to the same $(u, C)$. This implies $(\succeq)|_X = (\succeq')|_X$, i.e. the singleton ranking is the same. Note that if $\succeq \neq \succeq'$, then there must be some menu $A$ such that $A^*_\succeq \neq A^*_\succeq'$. This implies that there is some $x$ with – wlog – $x \in A^*_\succeq \setminus A^*_\succeq'$. Let $I_A(x)$ denote the $\succeq_A$-indifference class of $x$ and consider the menu

\[
D := \{z \in A : (i) \{x\} \succ \{z\}, (ii) x \succeq_A z\} \cup \{x\}
\]

By Axiom 2 (Strong Reduction) we know that $x \notin D^*_\succeq'$. OTOH, since $x \in A^*_\succeq$, we must have $x \in D^*_\succeq$. Since $x = \sup(D^*_\succeq)$ it follows that $\bar{x} \in C(D^*_\succeq)$, contradicting the fact that $C(D^*_\succeq) = C(D^*_\succeq')$. Hence, there is no such pair $(A^*_\succeq, A^*_\succeq')$, implying that $\succeq = \succeq'$.

\[\text{48}A^*_\succeq' \text{ denotes the feasible set computed under } \succeq \text{ and } A^*_\succeq \text{ denotes the feasible set computed under } \succeq'.\]
References

*Theoretical Economics*, forthcoming.

Rationale for the Attraction and Compromise Effects,” *Theoretical Economics*, 7,
125–162.


937–971.

Behavioral Optimization,” *Mimeo*.


More is Less: Choice with Limited Consideration,” *Mimeo*.

MANZINI, P. AND M. MARIOTTI (2012): “Categorize then Choose: Boundedly
Rational Choice and Welfare,” *Journal of the European Economic Association*,
10, 1141–1165.


*Mimeo*.

SIMONSON, I. AND A. TVERSKY (1993): “Context-Dependent Preferences,” *Man-