A Model of Temptation-Induced Compromise [*]

Madhav Chandrasekher[†]

Current Draft: March 3, 2011

Abstract: This paper models menu choice problems in which the decision maker (DM) faces temptation at the ex ante stage of choice and responds to this temptation by making a compromise. We introduce a new class of utilities, called *the category model*, which captures this phenomenon and provide axiomatic foundations for this model. Our theory provides an approach to modeling temptation that is distinct from the costly self-control approach, initiated by Gul-Pesendorfer (GP) (2001). Moreover, it provides a sharp representation of two behavioral phenomena of current interest in the decision theory literature - the compromise effect and the attraction effect. The innovation of this paper is that we identify temptation as a common cause for both of these effects and nest both of these features within an axiomatic model of compromise.

Keywords: Menu Choice, Preference for Commitment, Strotz Model.

[*]I am deeply grateful to Andrea Cann Chandrasekher, Eddie Dekel, Haluk Ergin, Bart Lipman, and Wolfgang Pesendorfer for their feedback on this project. An earlier draft was titled "A Theory of Temptation without Uncertainty." 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

1 Introduction

Compromise is a central feature of many decision problems. It can arise in the context of bargaining between two agents with non-aligned preferences, e.g. wage bargaining or dispute resolution. However, it can also arise in the context of single-agent decision-making - take the example of career choice or mate selection. In these situations the decision problem has the following structure. There is a menu of 'objective' options from which the decision-maker (DM) must make a choice. However, there is a tension between what the DM wants and what the DM can actually choose. That is, not all of the options on the menu may be feasible choices for the DM. In particular, some of the options that the DM might most prefer may be infeasible. This tension leads the DM to maximize his normative preference not over the full menu of options, but over the subset of attainable options.[1] The utility maximizing outcome from this process is what we refer to as the DM's compromise.

This paper develops and provides axiomatic foundations for a formal model of compromise. Taking the above procedure as a description of the process of compromise, there are two key questions that must be answered by our model if it is to provide insight into this process. First, why does the DM face a tension between what is on the menu and what he considers feasible? In other words, what is the root cause that leads the DM to compromise in the first place? Second, if the root cause is "X", what is the precise mechanism through which the presence of X maps a menu of choices into a sub-menu of (subjective) feasible choices? To answer the first question, this paper singles out temptation as the X factor that is at the root of compromise. As such, it makes a contribution to the axiomatic literature on temptation and self-control, initiated by the seminal piece Gul-Pesendorfer (GP) (2001). However, it is also distinguished from GP (2001) and the subsequent literature in that it is a model of compromise as opposed to a model of 'costly self-control.'

To explain this distinction in more detail, let us first recall that the behavioral primitive in models of temptation is (usually) taken to be an order on menus, or sets of consumption opportunities. The intuition is that the DM wants to, for example, quit smoking. However, he cannot pre-commit to avoid situations where there are opportunities to smoke. This implies a non-trivial ranking on menus, i.e. a ranking that is not induced up from the 'normative' ranking on singleton menus. For example, the DM attempting to quit smoking might exhibit the ranking, $\{\text{coffee}\} \succeq \{\text{coffee, cigarette}\} \succeq \{\text{cigarette}\}$. This means he would rather have the chance to consume something other than a cigarette than be committed to smoking (e.g. $\{\text{coffee, cigarette}\} \succeq \{\text{cigarette}\}$). Moreover, he would rather have himself committed to coffee than allow for the possibility of smoking (e.g. $\{\text{coffee}\} \succeq \{\text{coffee, cigarrete}\}$).

[1]The key point here is, while a given menu of options is 'objective' - i.e. the same for all DM's, the determination of what is feasible is 'subjective' and dependent on the individual DM's preferences.

When the DM exhibits the preference, $\{\text{coffee}\} \succ \{\text{coffee, cigarette}\} \succ \{\text{cigarette}\}$, the interpretation is that he anticipates exerting costly self-control, i.e. he plans on drinking coffee and incurring a utility cost of resisting the temptation to smoke. In contrast, when the preference is weak, i.e. $\{\text{coffee, cigarette}\} \sim \{\text{cigarette}\}$, the interpretation is that the DM caves in to temptation and smokes the cigarette.

Our model requires that the DM compromises in the face of temptation, rather than exerting costly self-control. This is certainly a stark restriction were we to only consider doubleton menus. The force of our theory derives from its explanatory power on more general (non-doubleton) menus. Moreover, the class of preferences analyzed in this paper cannot (necessarily) be built up from rankings on doubleton menus. Let us revisit the previous example, except now add the possibility of consuming nicotine gum (n-gum), which we take to be a milder temptation than a cigarette. How does the DM evaluate the menu {coffee, n-gum, cigarette}? Our model predicts a ranking: $\{\text{coffee, n-gum, cigarette}\} \sim \{\text{n-gum, cigarette}\} \sim \{\text{n-gum}\}$. Following the compromise process outlined in the first paragraph, the DM reduces the tripleton menu {coffee, n-gum, cigarette} to the menu of feasible choices {n-gum, cigarette}. The intuition is as follows. Since cigarettes are available and the DM is tempted to smoke (and since he never exerts costly self-control), he knows he won't be able to choose coffee. Thus, his feasible set of consumption options is {n-gum, cigarette}. Maximizing his normative ranking on this set yields {n-gum}, which is the less harmful temptation.

The reduction procedure that maps the menu {coffee, n-gum, cigarette} to the subset {n-gum, cigarette} relies on both (i) the existence of temptation and (ii) the fact that the DM doesn't exert costly self-control. The fact that temptation exists means that the DM can first sort the menu into a subset of elements that are 'untempted', e.g. {n-gum, cigarette} in the example. Second, the fact that the DM never exerts costly self-control implies that he never chooses elements that are tempted (i.e. coffee), so that he is indifferent between the original menu and the subset of untempted elements. Thus, when coffee and cigarettes are on the same menu together, it is as if coffee was never available in the first place. This reduction procedure explains how temptation reduces the 'objective' set of options to a 'subjective' set of feasible options.

The preceding discussion describes both how and why the presence of temptation can induce compromise, but we have not yet made an argument for why one needs yet another model of temptation. Beyond mere theoretical interest what can a model of temptation-driven compromise explain that existing models cannot? The interest in our model is that it provides a novel representation of two important classes of behavioral phenomena that cannot be explained by standard temptation models.

Respectively label these phenomena as *simple compromise* and *attraction*.[2] canonical example of simple compromise is the one given above. In symbols, we rewrite the example as follows.

Example 1 (Simple Compromise)**.** Put $x \succ y \succ z$.

- $\{x, y\} \sim \{x\}$
- $\{x, z\} \sim \{z\}$
- $\{x, y, z\} \sim \{y\}$.

Note that the preference is formally identical to the previous coffee and cigarettes example. We have called this an example of 'simple' compromise since the compromise of y (from the menu $\{x, y, z\}$) is induced by direct temptation, i.e. $\{x\} \succ \{x, z\} \sim \{z\}$ (for future reference we denote direct temptations with the notation $x \rightarrow_t y$). However, direct temptation is not the only manner in which the DM can be tempted. Consider the following example of 'latent temptation',

Example 2 (Attraction)**.** Put $x \succ y \succ z$.

- $\{x, y\} \sim \{x\}$
- $\{x, z\} \sim \{x\}$
- $\{x, y, z\} \sim \{y\}$.

To fix ideas, take $x =$ coffee, $y =$ n-gum, and $z =$ cigarette. The idea here is that the DM is tempted by nicotine and this temptation is latent (as opposed to direct) so that neither n-gum nor a cigarette can individually induce a compromise. However, when both are available on the menu the temptation for nicotine is triggered and he compromises. The example is an illustration of what is known as an "attraction effect".

In the choice theory literature, the attraction effect typically refers to the following choice behavior. A DM is offered choices between pairs $\{x, y\}, \{y, z\}, \{x, z\}$, in which case he chooses (respectively) x, y, and x. However, when offered a choice from the triple $\{x, y, z\}$ he chooses y. Thus, his choices from menus cannot be recovered from a single preference order over singletons.[3] There have been several recent attempts to model compromise and/or attraction effects, e.g. Ok-Ortoleva-Riella

[2] The phenomenon we refer to as simple compromise is not the same as the "compromise effect" from the behavioral economics literature. In this literature, the compromise effect refers to the tendency of consumers to choose a middle-ground choice when offered, say, a menu of 3 different brands (e.g. ((high reputation, high price) (middle reputation, middle price), (low reputation, low price))). Our example of compromise pertains specifically to temptation preference.

[3] The reason for the name is that the presence of z evidently creates an attraction for y where previously none was present when x was available.

(2007), Cherepanov-Feddersen-Sandroni (2008), Masatlioglu-Nakajima-Ozbay (2009), and de Clippel-Eliaz (2010). Several of these papers model the choice behavior exhibited in the attraction effect as arising from an aggregation amongst multiple (possibly conflicting) preference relations.[4] Our paper is distinguished from these other models in that it explains the attraction effect as an outcome of temptation-induced compromise, and hence provides a novel explanation for why it occurs. We now go into details of the model and conclude with a summary of our main representation results.

1.1 The Category Model

The model in this paper has two parameters:

- Take $\mathcal{C} := \{\mathcal{C}_i\}_i$ to be a collection of subsets of the consumption space X (a category).
- Let $u(\cdot) : X \to \mathbf{R}$ be a representation of the singleton ranking.

Using these two ingredients we construct the following menu utility:

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

Hereafter, in writing the above formula we will suppress the qualification that the maximum is taken only over those i such that $\mathcal{C}_i \cap A \neq \emptyset$. In this paper, we refer to any given pair $(u, \mathcal{C})$ as a category model. To motivate the model, let us match the features of the model with the informal compromise process described in the opening paragraph. Fix a DM whose preference on menus is represented by a category model, $(u, \mathcal{C})$. Break up the decision process into three steps, (i) categorization, (ii) reduction, and (iii) maximization. The last two steps formalize the compromise procedure outlined in the opening paragraph. From a given menu A, the DM makes a subjective determination of what is feasible from the menu (reduction) and then maximizes his normative ranking over the reduced subset (maximization). The first step, categorization, is the mechanism that determines how reduction occurs. Observe that the categories are derived from the DM's order on menus, so that the categorization process is what makes the reduction procedure subjective. Putting these three steps together we can associate the category utility with the following decision procedure: (put $A^* := \cup_i \arg\min_{x\in A\cap\mathcal{C}_i} u(x)$)

$$A \overset{\text{categorize}}{\longmapsto} \{\mathcal{C}_i \cap A\}_i \overset{\text{reduce}}{\longmapsto} A^*(= \cup_i \arg\min_{x\in A\cap\mathcal{C}_i} u(x)) \overset{\text{maximize}}{\longmapsto} \max_{x\in A^*} u(x) = U(A)$$

[4]The "multiple preferences" approach is not homogenous and each of these papers provides a novel perspective. For example, in Cherepanov-Feddersen-Sandroni (2008) the attraction effect arises within a model of multiple rationales, so that choice from a given set of options is determined by maximizing over the subset of rationale-dependent maxima. In contrast, de Clippel-Eliaz (2010) axiomatize a model of two-player bargaining and nest the attraction effect as a solution to this bargaining problem. de Clippel-Eliaz (2010), Masatlioglu, et al. (2009), and Manzini-Mariotti (2010) and our paper have some overlap, which we elaborate on in section 4.

The key feature of our model is the first step. Thus, we now briefly describe how categories are created from the preference order and then illustrate the construction with a representation of the simple compromise and attraction examples (Examples 1 and 2). The reduced menu A^* is then easy to compute once we understand how temptation feeds into the creation of categories.

Categorization

The DM faces temptation (perhaps only weak temptation, in which case $\mathcal{C}$ is just a collection of singletons). Moreover, there is a subjective collection of temptation characteristics that describe the aspects of the DM's temptation problem. The subscripts on the categories $\mathcal{C}_i$ index these attributes. How does the DM categorize objects into the attribute-specific bins $\mathcal{C}_i$? To illustrate, fix any x (e.g. something healthy) and note that, with the process of compromise in mind, the DM faces two kinds to obstructions that could prevent him from choosing x (for any given menu A). First, a menu A might possess a direct temptation for x (leading to simple compromise). Second, there might be a latent temptation (due to attraction effects). Recall that a direct temptation is one where x is tempted by a singleton y (i.e. $x \rightarrow_t y$). A latent temptation is where there is some (non-singleton) menu A where $x \succ A$ and $x \succ \{x, A\}$ and $x \sim \{x, A \backslash y\}, \forall y \in A$. That is, the element x is tempted by the set A and no smaller subset of A tempts x. Call a set $A(x)$ with this property an attraction set (for x). For a given attraction set $A(x)$, the DM considers the attributes of each element $z \in A(x)$ (each of these attributes are latent temptation characteristics). All elements that share, say, attribute i are mentally categorized into the set $\mathcal{C}_i$. Thus, the set $\mathcal{C}_i$ contains all elements y that possess the (subjective) attribute i and that belongs to an attraction set $A(x)$ that tempts x.

Note also that if $x \rightarrow_t y$, then y is a microcosm of all the attributes in $A(x)$ that tempt x so that y is in every category $\mathcal{C}_i$ that contains an element of $A(x)$. The DM then evaluates a menu as follows. Fix an $x \in A$ and imagine that the DM asks himself the question, "If I wanted to pick x from A can I do so without compromise?" If all the components of an x-attraction set are present, then a latent temptation is triggered and he compromises. If not all attractions are present, then he (costlessly) commits to x. If any direct temptations are present, since each direct temptation itself contains all attributes of the attraction set, he compromises in this case as well. To illustrate the categorization process, consider the simple compromise and attraction examples. The elements x, y, z below are exactly as described in those examples, e.g. $x =$ coffee, $y =$ n-gum, etc.

Example 1 (Simple Compromise). *Let* $\{x, y\} \sim x, \{x, z\} \sim z$ *and* $\{x, y, z\} \sim y$. *Put* $\mathcal{C}_1 := \{y\}, \mathcal{C}_2 := \{x, z\}$ *and* $\mathcal{C} \equiv \{\mathcal{C}_1, \mathcal{C}_2\}$.

- $U^{\mathcal{C}}(\{x, z\}) = u(z), U^{\mathcal{C}}(\{x, y\}) = u(x)$.
- $U^{\mathcal{C}}(\{x, y, z\}) = u(y)$.

To explain this representation note that the indices on the sets $\mathcal{C}_1$ and $\mathcal{C}_2$ enumerate the temptation characteristics, $\{\text{low nicotine}, \text{high nicotine}\}$. The set $\mathcal{C}_2$ is the "high nicotine" bin and $\mathcal{C}_1$ is the "low nicotine" bin. Since y is a low nicotine item and nothing is (directly) tempted by y, the set $\mathcal{C}_1$ contains only y. Moreover, since z has high nicotine and x is directly tempted by z, the set $\mathcal{C}_2$ contains both x and z.[5] Thus, using the category utility formula, we can see that the DM only makes the switch from x to y when z is present. The category model also provides a natural representation of the attraction example.

Example 2 (Attraction)**.** *Let* $\{x, y\} \sim x \sim \{x, z\}, \{x, y, z\} \sim y$. *Put* $\mathcal{C}_1 := \{x, y\}, \mathcal{C}_2 := \{x, z\}, \mathcal{C} \equiv \{\mathcal{C}_1, \mathcal{C}_2\}$. *Notice that*

- $U^{\mathcal{C}}(\{x, y\}) = u(x) = U^{\mathcal{C}}(\{x, z\})$.
- $U^{\mathcal{C}}(\{x, y, z\}) = u(y)$.

Similar reasoning explains this representation. The indices on the bins $\mathcal{C}_1$ and $\mathcal{C}_2$ index the (latent) temptation characteristics, $\{\text{low nicotine}, \text{high nicotine}\}$. Since y possesses the first attribute and z possesses the second, they are (resp.) placed in bins $\mathcal{C}_1$ and $\mathcal{C}_2$. Furthermore, the set $\{y, z\}$ is an attraction set (or compound temptation) for x. Thus, x is placed in any bin which contains an element of the attraction set for x - as x is implicitly tempted by the attribute (nicotine) possessed by that element.

Neither simple compromise nor attraction can be explained by the benchmark models in the temptation literature. Recall the following utility representation, axiomatized in GP (2001):

$$U^{GP}(A) = \max_{x \in A} (u(x) + v(x)) - \max_{x \in A} v(x)$$

This utility is often re-written in the following form. Put $c(x, A) := \max_{z \in A} v(z) - v(x)$ and note that

$$U^{GP}(A) = \max_{x \in A} (u(x) - c(x, A))$$

This representation was generalized in DLR (2009) who axiomatized the following model:

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

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

[5] We have stated that the categories are subjective. The labels 'low nicotine' and 'high nicotine' are certainly objective - all DM's would probably agree that n-gum has lower nicotine than a typical cigarette. What is subjective is the grouping of elements into bins, e.g. whether the DM is tempted by low nicotine vs. high nicotine.

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

We take this to be the benchmark 'no uncertainty' model of temptation preferences. The simple compromise example violates one of the DLR behavioral postulates (axiom AIT). 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.

Since the DM does not exert costly self-control in either example, perhaps the Strotz model is the right place to look for a representation. Recall that this model is described by a pair of mappings $u, v : X \to \mathbf{R}$ which are used to define the following menu utility: (for any menu A, put $A_v := \arg\max_{x \in A} v(x)$)

$$U^{ST}(A) := \max_{x \in A_v} u(x)$$

It is straightforward to see that the attraction example does not admit a Strotz representation. Moreover, the simple compromise example doesn't admit a Strotz representation either, although the argument is slightly more subtle. The preference is an example of what GP (2001) call a 'cycle', which they show is precluded by the Strotz model. Thus, none of the benchmark models in the temptation literature (neither GP (2001)/DLR (2009) nor Strotz) can rationalize either simple compromise or the attraction effect.

1.2 Summary of Results

The paper is broken into two main sections. Section 2 analyzes category models, i.e. pairs $(u, \mathcal{C})$, which are free of attraction effects. Hence, the category models in this section are models of simple compromise alone. Why bother analyzing a less general model? First, note that the issues of simple compromise and attraction are distinct. Thus, it is important to see whether we can use categories to model this distinction. Second, and somewhat surprisingly, the category model turns out to nest the Strotz model. Moreover, the class of categories that represent Strotz preferences do not admit attraction effects (call these Strotz categories). Thus, in order to understand where the class of Strotz categories is situated in the universe of category models it is important to analyze models of simple compromise. Third, the axioms that characterize the main representation (Theorem 2) in this section translate very easily into the verbal process of compromise described in the opening. Thus, the representation gives an intuitive behavioral description of the category model.

Having analyzed models of simple compromise, section 3 takes up the issue of attraction and contains two primary representation results, Theorems 4 and 5. We also introduce two new axioms in the section, respectively labelled "Compromise" and "Strong Reduction". The Compromise axiom is our formalization of the idea that the DM always compromises in the face of temptation. Theorem 4 shows that a version of the category model, called the *local category* model, is characterized by the Order and Compromise axioms. The import of this result is that, through the Compromise axiom, it ties the functional form of the category model to an axiomatic formulation of compromise. The second axiom introduced in the section, Strong Reduction, formalizes the idea that a menu can be iteratively reduced (by removing elements that are either directly or latently tempted) to an 'untempted' menu, A^* that is indifferent to the original menu. The relevance is that we maintain the link between the behavioral (i.e. axiomatic) description of our model and the informal verbal description of compromise given in this introduction. The main result in the section, and in the paper, is Theorem 5 which shows that the full category model is characterized by the Order, Compromise, and Strong Reduction axioms.

Since all categories are subjective, an important issue is identification. More precisely, is there a one-to-one map between category models and menu preferences representable by category models? In most cases, the answer to this question is 'yes' - however, this answer comes with two caveats. First, with few exceptions, we cannot identify categories when there are ties in the singleton ranking.[6] Second, akin to the refinement in Dekel, et al. (2001) that requires states to be relevant, we need to make a refinement in the class of allowable categories in order to pin down the model. As in DLR (2001), the problem is that we can artificially inflate the model with irrelevant categories even though the utility representation only 'lives on' a smaller sub-collection of categories. Once we make a refinement to eliminate such wasteful representations, the model is essentially pinned down. Finally, omitted proofs from the main text are collected in the appendix.

2 Models of Simple Compromise

We begin with a description of the choice environment.

- Let $X = \{x_1, \ldots, x_n\}$ be an enumeration of the prize space.
- Let 2^X denote the collection of subsets of X (menus).
- Let $\mathbf{P}(X)$ be the set of complete, transitive preference relations defined on 2^X.

Finally, a word on notation. For any menu A we take $\sup(A), \inf(A)$ to be (resp.) the set of $\succeq$-maximal ($\succeq$-minimal) elements, i.e. singleton menus, in the menu A.

[6] We state explicitly in our uniqueness results when we do or do not require the no ties condition.

Definition 1: Fix any $A \in 2^X$. An element $x \in A$ is called a *strong compromise* of A if the following two properties hold:

- $x \sim A$.
- $A' \sim A$ whenever $A' \subseteq A$ and $x \in A'$.

Let $\Sigma_s(A)$ denote the strong compromises of a menu A. To motivate the title, note that the first condition says that every menu possesses a compromises. The second condition says that if a sub-menu contains the original compromise, then the DM is indifferent between the two menus.

A1*: (Strong Compromise) If $A \neq \emptyset$, then $\Sigma_s(A) \neq \emptyset$.

A1*: (Degenerate Set-Betweenness) For any menus A, B either $A \cup B \sim A$ or $A \cup B \sim B$.

We will refer to the latter restriction as 'Degenerate Set-Betweenness,' after GP (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. Refer to Degenerate Set-Betweenness as 'DSB' and let 'SC' denote Strong Compromise.

Lemma 1. *DSB and SC are equivalent.*

Definition 2: Let $u : X \to \mathbf{R}$, $v : X \to \mathbf{R}$ be a pair of mappings. Let $A_v := \arg\max_{x \in A} v(x)$ and put

$$U^{ST}(A) := \max_{x \in A_v} u(x)$$

where the super-script stands for Strotz utility.

In GP (2005), it is shown that the Strotz model is characterized by solely the order and DSB axioms. Note that the (u, v) formulation of the Strotz model given above allows for there to be ties in $v(\cdot)$ ranking. In the menus of lotteries choice domain there are observable restrictions (on the menu preference) implied by the existence of ties in the $v(\cdot)$ ranking. However, in the finite menus case this is no longer true. To see this, given a pair (u, v) we resolve all ties in the $v(\cdot)$ function to obtain an equivalent representation by a pair (u, L) as follows (where the ranking L is free of ties). Let $I_v^1, \ldots, I_v^k$ be a top-down enumeration of the indifference classes of $v(\cdot)$. Within each class resolve ties in favor of the u-ranking. If there is a tie in both the v and the u ranking, then break the tie arbitrarily. Label the resulting function $L(\cdot)$ and note that the pair (u, L) generates the same menu preference (via the Strotz formula) as the pair (u, v). It will be easier to carry out comparisons with the Strotz model when the $v(\cdot)$-ranking is strict, so we shall use the (u, L) formulation when referring to a Strotz preference. The following definition just re-labels the parameters of the Strotz utility.

Definition 3: Let $u : X \to \mathbf{R}$ be a *commitment* mapping and let $L : X \to \{1, \ldots, |X|\}$ denote a *labeling* of the elements of the prize space X. Define the induced relation, $\succeq_L$, on X as follows: $x \succeq_L y$ if and only if $L(x) \leq L(y)$. Put

$$U^L(A) := u(\succeq_L -\max_{x \in A} x)$$

and call this a *rank-based* representation.

The following Theorem characterizes preferences representable by rank-based (Strotz) utilities.

Theorem 1. *(GP (2005)) A preference* $\succeq \in \boldsymbol{P}(X)$ *satisfies* $A1^*$ *if and only if it can be represented by a rank-based utility.*

We now turn our attention to an alternative model of no uncertainty temptation preferences.

Definition 4: Let $\sqcup_i B_i$ be a partition of X and $u : X \to \mathbf{R}$ a commitment mapping. Define a *partition utility*, $U^P(\cdot)$, as follows:

$$U^P(A) := \max_i \{\min_{x \in A \cap B_i} u(x)\}$$

Call a menu A *temptation free* (i.e. there are no direct temptations) if for any pair $x, y \in A$ with $x \succ y$ we have $\{x, y\} \sim \{x\}$. Introduce the following axioms. The first axiom will be referred to as *No Attractions*, or NAT for short. In words, it says that if there are no direct temptations, then there are no latent temptations either.

A1′: If $A \cup B$ is temptation free, then $A \cup B \sim A$ or $A \cup B \sim B$.

A2*: (Reduction) If $x \succ y$ and $\{x, y\} \sim y$, then $A \sim A \cup x$ for any $A \ni y$.

The former axiom clearly weakens $A1$ by requiring that Degenerate Set-Betweenness only holds on temptation free menus. The latter axiom can be seen as a consistency condition on the preference: Think of the first statement ($y \sim \{x, y\}$) as saying that y *tempts* x. The axiom requires that when we enlarge the ambient menu from $\{x, y\}$ to $A \ni y$, the menu A should still tempt x since the original temptation y is still present. Note that $A1^*$ implies Reduction (this is easiest to see using the Strong Compromise formulation of $A1^*$). Despite these relations between the axioms, the $U^L(\cdot)$ model is distinct from the $U^P(\cdot)$ model as a consequence of the following additional axioms. Let $(-\infty, z] = \{y \in X : z \succeq y\}$ denote an order interval in X and say that $x \to_t y$ if $x \succ y$ and $\{x, y\} \sim \{y\}$ (i.e. x is overwhelmed by y). Denote $B_t(x) := \{y \in X : x \to_t y\}$.

A3a: Fix $x \succ y \succ z$. If $x \to_t y$ and $x \to_t z$, then $y \to_t z$.

A3b: Fix $x \succ y \succ z$. If $x \to_t z$ and $y \to_t z$, then $x \to_t y$.

Axiom A3a can be thought of as a "temptation monotonicity" condition: It says that the relation "$\to_t$" places an ordering on the set of all elements that tempt x. Axiom 3b says that no two x and y can share a common temptation without one tempting the other. The following examples verify that the partition and rank-based models are formally disjoint in the sense that neither one nests the other.

Example 3 ($U^L \not\Rightarrow U^P$)**.** *Let $X = \{x \succ y \succ z\}$ and let $x \to_t y, x \to_t z$, but $y \not\to_t z$. Put $y \succ_L z \succ_L x$ and note that $U^L(A) := u(\succeq_L -\max_{x \in A} x)$ represents $\succeq$.*

Example 4 ($U^P \not\Rightarrow U^L$)**.** *Let $X = \{x \succ y \succ z \succ w\}$. Put $B_1 = \{x, z\}, B_2 = \{y, w\}$ and set $U^P(A) := \max_{i=1,2} \{\min_{r \in A \cap B_i} u(r)\}$. Note that the preference generated by $U^P(\cdot)$ does not satisfy A1*: $\{x, y, z, w\} \sim \{z\}$, yet $\{y, z\} \sim y \succ z$.*

The following is the final axiom required by the partition model.

A4: If $z_1 \sim z_2$ and $x \to_t z_1, x \to_t z_2$, then $\{x : x \to_t z_1\} = \{x : x \to_t z_2\}$.

This final axiom is imposed only on pairs (z_1, z_2) where $z_1 \sim z_2$. In words, it says that the partition model is too blunt to distinguish between elements with the same normative value: Their respective sets of temptations must either be disjoint or identical. Note that when there are no ties in the normative ranking, then Axiom A4 is completely vacuous. Among the axioms that characterize the partition model, A3b, is in our view, the most severe. We next consider what model obtains when we keep $A1', A2^*$, and $A3a$, but drop $A3b$. The resulting category model turns out to be an interesting generalization of the partition model, which we call the *narrow category.* First, we present the representation result for the partition model. The proof requires the following lemma.

Lemma 2. *Assume $\succeq \in \boldsymbol{P}(X)$ satisfies A2*. If $x \to_t y$ and $y \to_t z$, then $x \to_t z$.*

Proposition 1. *A preference $\succeq \in \boldsymbol{P}(X)$ satisfies $A1', A2^*, A3a, A3b$, and $A4$ if and only if it admits a representation by the partition model.*

We have the following straightforward uniqueness claim, so that partitions are identifiable from preference.

Corollary 1. *Let $(\{B_i\}, u(\cdot)), (\{B'_i\}, u'(\cdot))$ be two distinct $U^P(\cdot)$ representations of $\succeq$. Then, $\{B_i\} \equiv \{B'_i\}$ and $u(\cdot)$ and $u'(\cdot)$ are ordinally equivalent.*

As a consequence of the NAT axiom, the partition model only represents preferences where temptations are multi-dimensional, but not compound. However, it goes a little further than this. As a consequence of $A3a$ and $A3b$, there are no 'shared temptations' either. For example, one could imagine that two candidate commitments x_1 and x_2 are both subject to various types of simple temptations. Moreover, some of the objects that tempt x_1 may also tempt x_2. This possibility is excluded by the partition model. We now consider a category model, the *NAT* category, that relaxes this restriction while keeping the NAT axiom intact. Nevertheless, a relative virtue of the partition model over the NAT category is that it is much easier to actually construct a partition model whereas it isn't as transparent how to construct a NAT category (that isn't a partition) from scratch.

Definition 5: A *No Attractions* (NAT) category is comprised of a collection of sets $\mathcal{C} \equiv \{\mathcal{C}_i\}_i$ and a commitment mapping $u : X \to \mathbf{R}$ with the following two properties: (let $(-\infty, x) := \{z \in X : x \succ z\}$)

1. $\cup_i \mathcal{C}_i = X$.

2. Let $\Sigma(x) = \{\mathcal{C}_i \in \mathcal{C} : x \in \mathcal{C}_i\}$. Then, there is some $\mathcal{C}_x \in \Sigma(x)$ s.t. $(-\infty, x) \cap \mathcal{C}_x \subseteq \mathcal{C}_i, \forall \mathcal{C}_i \in \Sigma(x)$.

Theorem 2. *A preference $\succeq \in$ $\boldsymbol{P}(X)$ is representable by a NAT category model if and only if it satisfies $A1'$ and $A2^*$. Moreover, if $(u, \mathcal{C})$ is any category model that represents $\succeq$, then $\mathcal{C}$ is a NAT category.*

Proof. We first check the representation claim, then the minimality claim. We first verify necessity of $A1'$ and $A2^*$. To check Reduction ($A2^*$) take $x \to_t y$ and note that if $(u, \mathcal{C})$ represents $\succeq$, then whenever $x \in \mathcal{C}_i$ we must also have $y \in \mathcal{C}_i$. It follows that $A \backslash x \sim A$. To check $A1'$, it suffices to show that if A is temptation-free, then for each $x \in A$ there is some $\mathcal{C}_x \in \mathcal{C}$ with $((-\infty, x) \cup \{x\}) \cap \mathcal{C}_x = \{x\}$. Since $\mathcal{C}$ is a NAG category, find $\mathcal{C}_x$ such that $(-\infty, x) \cap \mathcal{C}_x \subseteq \mathcal{C}_i, \forall \mathcal{C}_i \in \Sigma(x)$. I claim that if A is temptation-free, then $(-\infty, x) \cap \mathcal{C}_x = \emptyset$. Else, if $x \succ y \in \mathcal{C}_x$, then $y \in \mathcal{C}_i, \forall \mathcal{C}_i \in \Sigma(x)$ implying that $x \to_t y$ - contradiction. Axiom $A1'$ follows. Now we turn to the sufficiency argument. For each $x \in X$ consider the set $\mathcal{C}_x := B_t(x) \cup \{x\}$ and let $\mathcal{C} \equiv \{\mathcal{C}_x : x \in X\}$. Let $U^{\mathcal{C}} : 2^X \to \mathbf{R}$ be the function defined by

$$U^{\mathcal{C}}(A) := \max_{\mathcal{C}_x} \min_{z \in \mathcal{C}_x \cap A} u(z)$$

I claim that (i) $\mathcal{C} \equiv \{\mathcal{C}_x\}_x$ is a NAT category and (ii) $U^{\mathcal{C}}(\cdot)$ represents $\succeq$. For the first claim let $x \in \mathcal{C}_y \cap \mathcal{C}_z$ and note that $(-\infty, x) \cap \mathcal{C}_x = B_t(x)$. Moreover, if $x \in B_t(y) \cap B_t(z)$, then $B_t(x) \subseteq B_t(y) \cap B_t(z)$ (by Lemma 2). Thus, $\mathcal{C}_x \subseteq \mathcal{C}_y \cap \mathcal{C}_z$ so that the collection $\{\mathcal{C}_x\}_x$ is a NAT category. Now we check representability. Given a menu A, let A^* denote its maximal temptation-free subset (which exists by $A1'$ and $A2^*$). Note that $U^{\mathcal{C}}(A) = U^{\mathcal{C}}(A^*)$, so that (as in the proof for the partition model) it suffices to check representability on the set of temptation-free menus. But this is obvious since for any $x \in \sup(A)$ we have $\mathcal{C}_x \cap A = \{x\}$ as A is temptation free. Thus, $U^{\mathcal{C}}(A) = u(x) = U(A)$. Now for the minimality claim.[7] For each $x \in X$ put $\Sigma(x) := \{\mathcal{C}' \in \mathcal{C} : x \in \mathcal{C}'\}$. Let $(u, \mathcal{C})$ denote any category that represents the preference and, towards contradiction, assume there is some x such that for every $\mathcal{C}' \in \Sigma(x)$ we have $(-\infty, x) \cap \mathcal{C}' \nsubseteq \mathcal{C}''$, for some $\mathcal{C}'' \in \Sigma(x)$. For each $\mathcal{C}'$, select some $z' \in \{(-\infty, x) \cap \mathcal{C}'\} \backslash \mathcal{C}''$ and note that, by representability, $z' \notin B_t(x)$. Let $A := \{z' \in \mathcal{C}' : \mathcal{C}' \in \Sigma(x)\} \cup \{x\}$ and note that, by $A1'$ and $A2^*$, $A \sim x$. On the other hand, we have $U^{\mathcal{C}}(A) < u(x)$ - a contradiction. □

We turn now to the issue of identification of the NAT category. First, note that the definition of the NAT category allows categories to be redundant. That is, if $\mathcal{C} \equiv \{\mathcal{C}_i\}_i$ is a category, then we say that the collection is redundant if there is some $\mathcal{C}_i \in \mathcal{C}$ such that $\mathcal{C}_i \subseteq \cup_{j \neq i} \mathcal{C}_j$. Notice that, absent a restriction that rules out redundant categories, there isn't a unique category model $(u, \mathcal{C})$ that represents a preference $\succeq$. To see this most clearly, let x_* be the lowest ranked singleton in X and note that if the model $(u, \mathcal{C})$ represents $\succeq$, then so does the model $(u, \mathcal{C}')$, where $\mathcal{C}' := \mathcal{C} \cup \{x_*\}$. Of course, this begs the question of why we don't just add on a non-redundance restriction in the definition of the NAT category. The reason we don't do this is that there is an important sub-class of NAT preferences, namely - Strotz preferences, that only admit a representation by a redundant category. Thus, in order to represent these preferences with a category model we must move out of the domain of non-redundant categories.

All is not lost if we relax the non-redundance restriction. For example, akin to the refinement on the state space in DLR (2001) that requires all states to be relevant, we can allow categories to be redundant, but not irrelevant. That is, every category is relevant for the comparison of some pair (A, B) so that the reduced model obtained by omitting this category produces an incorrect ranking on the pair (A, B). Formally stated, we have:

Definition 6: A category $(u(\cdot), \mathcal{C})$ is called *sharp* if for any complete sub-category $\mathcal{C}' \subseteq \mathcal{C}$ we have $\succeq_{(u,\mathcal{C})} \neq \succeq_{(u,\mathcal{C}')}$.

The identification result we present is for sharp category representations. That is, if we enrich the class of models $(u, \mathcal{C})$ to include categories that are possibly redundant, but still sharp, then we can pin down the representing category. This claim, however, comes with a caveat - there is no hope of obtaining identification when there are ties in the singleton ranking. To see this, take $X = \{x_1, \ldots, x_n\}$ to be any set

[7] This argument is not the same as checking necessity. The preceding argument just shows the axioms are necessary and sufficient for a NAT category representation. The minimality claim shows that, moreover, there cannot be *any* other category model that represents these preferences. Thus, the *only* way to interpret axioms $A1'$ and $A2^*$ (NAT preferences) with a category model is via a NAT category.

and assume all elements are indifferent under the singleton ranking. Then, the following two categories constitute distinct, sharp representations: $\mathcal{C}_1 \equiv \{x_1, \ldots, x_n\}$, $\mathcal{C}_2 = \{\{x_1\}, \{x_2\}, \ldots, \{x_n\}\}$. A similar multiplicity can be constructed if there is any non-trivial indifference class in the singleton ranking. Thus, for the identification result we will assume that the restriction of the menu preference to singletons is strict.

Say that the set $B_t(x) \cup x$ is *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 transitive relation on the collection of sets $\{B_t(x) \cup x\}$. To see this, note that the association $x \mapsto B_t(x) \cup x$ is a bijection from X to the collection of sets of the form $B_t(x) \cup x$: 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 $x \in B_t(y)$. Thus, we get a well-defined induced "dominance" relation on X which we can equivalently think of as a relation on the sets $B_t(x) \cup x$.

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. Say that a category $\mathcal{C}_2$ is a *prolongation* of $\mathcal{C}_1$ if there is a bijection $\pi : \mathcal{C}_1 \to \mathcal{C}_2$ such that for every $\mathcal{C}'_1 \in \mathcal{C}_1$ we have $\mathcal{C}'_1 \subseteq \pi(\mathcal{C}'_1)$. We then have the following identification claim for the NAT category.

Corollary 2. *Assume $\succeq |_X$ is strict and let $(u_1, \mathcal{C}_1), (u_2, \mathcal{C}_2)$ be any two sharp NAT category representations of $\succeq$. Then, both $\mathcal{C}_1$ and $\mathcal{C}_2$ are prolongations of the category $\mathcal{B}^*$ and u_1 and u_2 are ordinally equivalent.*

Say that a model $(u, \mathcal{C})$ is a *retraction* of $(u, \mathcal{C}')$ if the latter is a prolongation of the former. The corollary says that there is a unique sharp category that represents $\succeq$ with the property that this model is the retraction of all other sharp models $(u, \mathcal{C})$ that represent $\succeq$. We take the NAG axiom along with $A2^*$ (Reduction) as a definition of temptation without aggregation, so that the theorem provides a behavioral identification of compromise without latent temptation. Our final representation result for this section is for a model that we call the *narrow category.* The narrow category relaxes the partition model by removing axiom $A3b$. It is also rich enough to allow different sets of temptations for pairs of elements (z_1, z_2) with $z_1 \sim z_2$, so that it also relaxes axiom $A4$.,

Definition 7: Let $(u(\cdot), \mathcal{C})$ be a category, where $\mathcal{C} = \{\mathcal{C}_i\}$ has the following two properties:

1. $\cup_i \mathcal{C}_i = X$.

2. If $x \in \mathcal{C}_i \cap \mathcal{C}_j$, then $\{y \in \mathcal{C}_i : x \succ y\} = \{y \in \mathcal{C}_j : x \succ y\}$.

Call a pair $(u(\cdot), \mathcal{C})$ where $\mathcal{C}$ satisfies these properties a *narrow category*. Note that the narrow category model nests the partition model. We have the following characterization:

Theorem 3. *A preference $\succeq \in \boldsymbol{P}(X)$ satisfies $A1', A2^*$, and A3a if and only if it admits a narrow category representation.*

Since the narrow category sits inside the NAT category, we know that the identification result from corollary 2 applies. However, for this particular model we can provide an even stronger form of identification.

Corollary 3. *(Uniqueness) Assume that $(u_1(\cdot), \mathcal{C}_1), (u_2(\cdot), \mathcal{C}_2)$ are two narrow category representations of a preference $\succeq$ (where $\succeq|_X$ is strict) and assume $\mathcal{C}_1$ and $\mathcal{C}_2$ are sharp. Then, $u_1(\cdot)$ and $u_2(\cdot)$ are ordinally equivalent and $\mathcal{C}_1 \equiv \mathcal{C}_2$.*

Given a rank-based Strotz pair $(u(\cdot), L(\cdot))$, we now back out an explicit category $\mathcal{B}_L$ where the pair $(u(\cdot), \mathcal{B}_L)$ is such that $U^{\mathcal{B}_L}(\cdot) = U^L(\cdot)$. Let $X = \{x^1, \ldots, x^n\}$ be the L-labeling of X. For each x^i put $D_i := B_t(x^i) \cup \{x^i\}$ where $B_t(x) := \{y : x \rightarrow_t y\}$. Set $\mathcal{B} := \{D_i\}$ and note that if $\succeq$ admits a rank-based representation, then the collection $\mathcal{B}$ is typically *not* non-redundant.

Proposition 2 (Strotz Model $\hookrightarrow$ Category Model). *Let $D_i := B_t(x^i) \cup \{x^i\}$ and let $(u(\cdot), \mathcal{B})$ be the corresponding category-based pair. Then, $U^{\mathcal{B}}(\cdot) \equiv U^L(\cdot)$.*

Proof. Note that if $x_A = \arg\min_{z \in A} L(z)$, then $B_t(x_A) \cap A = \emptyset$. Thus, $U^{\mathcal{B}}(A) \geq u(x_A) \geq U^L(A)$. To show the reverse inequality note that for any $x \succ x_A$ with $x \in A$ we must have $L(x) > L(x_A)$. Thus, $x \rightarrow_t x_A$ so that $\min_{z \in (B_t(x) \cup \{x\}) \cap A} u(z) \leq u(x_A)$. It follows that $U^{\mathcal{B}}(A) \leq u(x_A) = U^L(A)$. □

Note that the category used to represent the rank-based model is possibly non-redundant. This begs the question of whether we can actually do better in general and represent the rank-based model with a non-redundant system of categories. The following example shows that the answer to this question is negative.

Example 5. *Let $X = \{x_1 \succ x_2 \succ x_3\}$ and define $x_2 \succ_L x_3 \succ_L x_1$. Note that if $(u(\cdot), \mathcal{B})$ is a category-based model that represents the preference $\succeq$ generated by the rank-based pair $(u(\cdot), L(\cdot))$ then let $\mathcal{C}_{x_1} \in \mathcal{B}$ be a set that contains x_1. Since $x_1 \rightarrow_t x_2, x_1 \rightarrow_t x_3$ we must have $x_2, x_3 \in \mathcal{C}_{x_1}$. But non-redundance then implies that $\mathcal{B} = \{\mathcal{C}_{x_1}\}$. This implies $U^{\mathcal{B}}(\{x_2, x_3\}) = u(x_3) \neq u(x_2) = U^L(\{x_2, x_3\})$ - contradicting representability.*

Thus, the refinement to sharp categories is necessary in order to represent all NAT preferences. We now move on to consider models with attraction effects.

3 Models with Attraction Effects

The following axiom is the main behavioral restriction in the paper, and is implied by all versions of the category model. The axiom constitutes a definition of the idea of compromise.

A1: (Compromise) Every non-empty menu A possesses an $x \in A$ such that:

- $x \sim A$
- If $A' \subseteq A$ and $x \in A'$, then $A' \succeq A$

Note that a compromise is defined by two characteristics. First, the DM is indifferent between being committed to the compromise or being given the option set A. Second, any subset of the original option set which contains the original compromise is weakly preferred. This second feature captures the idea that compromise is relative. When we shrink the menu of options to A', then we have possibly removed some of the temptations that led the DM to compromise to x in the first place. Moreover, the original compromise is still available. Hence, the weak preference for A' over A. Let $\Sigma_w(A)$ denote the set of compromises in the menu A (the "w" subscript is for weak compromise).

3.1 Local Categories

First consider the following definitions.

Definition 8: For each menu A let $\mathcal{B}_A$ be a collection of categories, one for each menu A. Say that the system $\{\mathcal{B}_A\}$ is *coherent* if the following additional property holds: If $A' \subseteq A$, then for each $D_{A'} \in \mathcal{B}_{A'}, \exists D_A \in \mathcal{B}_A$ with $D_{A'} \subseteq D_A$.

To understand the coherence condition, fix a menu of options A' and think of the collection $\mathcal{B}_{A'}$ as a grouping of elements of A' by shared temptation characteristics. The indices of the sets $\mathcal{C}_i \in \mathcal{B}_A$ label the (subjective) set of characteristics that the DM finds tempting when the option set is A'. Now enlarge the option set to A. This could have two conceivable effects, called respectively consistency and coarsening. First, since there may be new temptations added to the menu, the set of tempting characteristics is possibly larger. Moreover, the items that the DM found tempting in the smaller option set are still present. Thus, when he groups objects by shared characteristics, there should be still be a bin, $\mathcal{C}_i \in \mathcal{B}_A$, assigned to every characteristic that the DM found tempting 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. In symbols, imagine that when the menu is $A' = \{x, y, z\}$, the collection of subjective characteristics is $\{\text{red}, \text{white}\}$. Let's say that this leads to the following grouping

$$\mathcal{C}_{\text{red}} = \{x, y\}, \mathcal{C}_{\text{white}} = \{z\}$$

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

$$\mathcal{C}_{\text{red}} = \{x, y, w\}, \mathcal{C}_{\text{white}} = \{z\}, \mathcal{C}_{\text{blue}} = \{z, w\}$$

Imagine that the colors are types of temptations, so that the presence of blue temptation, w, creates a temptation for z where previously none was present. Note that the coherence restriction is silent on the composition of bins that correspond to new temptations. It only requires that bundles in the smaller menu are carried over to the larger menu, where possibly more elements are added. This is the consistency requirement.

A second possible effect of enlarging the option set is that the DM might make coarser comparisons between objects with similar characteristics when there are more comparisons to make. To fix ideas, imagine A is the set of options in a chocolate artisan shop. Patrons have a chance to consume something (relatively) healthy (e.g. biscotti), along with several varieties of tempting chocolates. The bins in the subjective category a patron associates to this menu index the variety of chocolates. Now enlarge the option set to A', where A' is a large candy store. While there are new temptation bins added to the category (due to the first effect), there are also many more objects to categorize. For example, truffles and caramels are in separate bins for the category associated to chocolate artisan shop, but they all go in one bin for the category associated to the large store. This is the coarsening effect. The objective of the coherence restriction is to capture both effects - consistency and coarsening.

Definition 9: A collection $\{\mathcal{B}_A\}_A$ is *downwards rigid* if the following condition holds: For each $D_i \in \mathcal{B}_A,\ \inf(D_i) \nsubseteq \cup_{j \neq i} D_j$.

This requirement maintains that there is some separability in the collection of tempting characteristics. Namely, for each characteristic there is an object in the menu that possesses that characteristic only. The separability condition is formalized as follows:

$$(\textbf{Separability}) \quad D_i \nsubseteq \cup_{j \neq i} D_j,\ \forall D_i \in \mathcal{B}_A$$

That is, each set D_i in the collection $\mathcal{B}_A$ contains a distinguished element that is not in any of the other sets in the collection. The downwards rigidity condition strengthens this notion of separability. The distinguished element can be taken to be the normatively worst ranked element in the bin D_i. The interpretation is as follows. Given a menu A, the DM enumerates a list of the temptation characteristics possessed by the feasible elements of A (e.g. red, white, blue in the preceding discussion). Next, he categorizes elements of the menu into bins labelled by these characteristics. With this procedure in mind, note that normatively minimal elements of the sets D_i are always feasible since they are free of direct or latent temptation relative to the menu A. Downwards rigidity requires that for each subjective temptation characteristic on the list there is a *unique* feasible element in A that possesses this characteristic.

This completes the description of the parametric restrictions on the collection $\mathcal{B}_A$. A pair $(u, \{\mathcal{B}_A\}_A)$ consisting of (i) a commitment mapping $u : X \to \mathbf{R}$ and (ii) a coherent and downwards rigid collection of categories is called a *local category* model. The forthcoming theorem is the representation result for the local category model. Since the local category is characterized by the order and compromise axioms, the value of the result is that it ties a behavioral notion of compromise as embodied in $A1$) with the functional form of the category model. For a clearer exposition, I assume in the forthcoming proof that the singleton ranking $\succeq|_X$ is strict. This clears up some of the notation required to define the local categories and is wlog. A modified construction shows the same result when ties on singletons are allowed. The proof of the required modification is in the appendix.

Theorem 4. *A preference* $\succeq \in \boldsymbol{P}(X)$ *satisfies* $A1$ *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 $A1$. Without either one, it is possible to construct an example of a local category that does not satisfy the compromise axiom. We now turn to the proof of this theorem. The key to the sufficiency construction is to study the following 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 $\{\succeq_A\}_{A \in 2^X}$ that is independent of the choice of $U(\cdot)$. Recall the following axiom, introduced in DLR (2009):

Positive Set-Betweenness: $A \succeq B \Rightarrow A \succeq A \cup B$.

The Positive Set-Betweenness axiom yields the following important structural result on the relation $\succeq_A$. Let A_x denote the maximal menu such that $U(A_x) = u_{\min}(x, A)$ and put $\theta_x(A) := \{y : x \succeq_A y\}$. The following result is taken from Chandrasekher [2009]. We reproduce the proof here for completeness.

Proposition 3. $A_x = \theta_x(A)$.

Proof. Note that if $x \succeq_A y$, then $A_x \succeq A_y$. This implies, by Positive Set-Betweenness, 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 $y \sim_A x$. □

Proof of Theorem 4. Necessity of *A*1 follows easily from the fact that the system of local categories, $\{\mathcal{B}_A\}$, is coherent and downwards rigid.[8] Note also that both of these conditions are necessary to imply *A*1. It is straightforward to construct counterexamples if either condition is omitted. Thus, we require the full structure of the local category system. For the sufficiency argument we first make the following claim. Let $I_A(x)$ denote the $\succeq_A$ equivalence class of an element x and let $I_A^1, \ldots, I_A^k$ be a top-down enumeration of the $\succeq_A$ equivalence classes.

Claim 1. *Let $\succeq \in \boldsymbol{P}(X)$ satisfy A1. Then, $\Sigma_w(A) = \inf\{y : y \in I_A^1\}$.*

Proof of Claim 1. First observe that $\Sigma_w(A) \subseteq I_A^1$: Otherwise, if $x \in \Sigma_w(A) \cap I_A^j$ for some $j > 1$, then by *A*1 we obtain $\cup_{j\geq 2} I_A^j \succeq A$. On the other hand, by Proposition 3, $A = \cup_{j=1}^k I_A^j \succ \cup_{j\geq 2} I_A^j$ - contradiction. Thus, $\Sigma_w(A) \subseteq I_A^1$. Now for any $y \in I_A^1$ consider the menu $A' := \{y, I_A^2, \ldots, I_A^k\}$ and note that $A' \succeq A$, again by definition of $\succeq_A$. Take any $x \in \Sigma_w(A)$ 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 $u_{\min}(x, A') = u_{\min}(x, A), \forall x \in A' \backslash y$ by Proposition 3. Since $u_{\min}(y, A) > u_{\min}(x, A), \forall x \in A' \backslash y$ it follows that

$$u_{\min}(y, A') \geq u_{\min}(y, A) > u_{\min}(x, A) = u_{\min}(x, A'), \text{ for all } x \in A' \backslash y$$

Therefore, by Proposition 3, 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 $y \succeq x, \forall y \in I_A^1$, implying that $\Sigma_w(A) \subseteq \inf\{y : y \in I_A^1\}$. To show the reverse containment, simply note that if $y \in \inf\{y : y \in I_A^1\}$ and $y \in B \subseteq A$, then by definition of $\succeq_A$ we have $B \succeq A$. □

Similar arguments can be utilized to show the following claim.

Claim 2. *Let $I_A^1, \ldots, I_A^k$ be a top-down enumeration of the $\succeq_A$ indifference classes. Then, $\inf I_A^1 \succ \inf I_A^2 \succ \cdots \succ \inf I_A^k$.*

Now we construct a system of local categories. Consider the following candidate. Fix a menu A and let $I_A^1, \ldots, I_A^k$ be a top-down enumeration of the $\succeq_A$ indifference classes. For any menu B, let $I_B(x)$ denote the $\succeq_B$ indifference class of $x \in B$ and

[8] Necessity of *A*1: Fix a local category model $(u, \{\mathcal{B}_A\}_A)$ and let $\mathcal{C}_1, \ldots, \mathcal{C}_n$ denote the categories in the collection $\mathcal{B}_A$ and let $X_1, \ldots, X_n$ denote the sets $\arg\min_{y \in \mathcal{C}_i} u(y)$. 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 $x_1 \notin \cup_{j \neq 1} \mathcal{C}_j$. Put $A' \subseteq A$. I claim that if $x_1 \in A'$, then $A' \succeq A$. Let $\mathcal{C}_1', \ldots, \mathcal{C}_m'$ be the categories in the collection $\mathcal{B}_{A'}$ and label so that $x_1 \in \mathcal{C}_1'$. By coherence, there is a category $\mathcal{C}_i \in \mathcal{B}_A$ such that $\mathcal{C}_1' \subseteq \mathcal{C}_i$. Since $x_1 \in \mathcal{C}_i$ this implies (by downwards rigidity) that $\mathcal{C}_i = \mathcal{C}_1$. Thus, $\min_{z \in \mathcal{C}_1'} u(z) = u(x_1) = U(A)$. It follows that $A' \succeq A$, implying that $\Sigma_w(A) \neq \emptyset$ - and, hence, *A*1.

let (abusing notation) $I^A(x) := \bigcup \{y \in A' : y \succeq x, y \in I_{A'}(x)\}$ where the union is taken over all sets A' such that (i) $A' \subseteq A$, and (ii) $x \in \Sigma_w(A')$. Inductively define a sequence of sets as follows. Put

$$D^i_A(1) := \bigcup_{x \in I^i_A} I^A(x)$$

and let

$$D^i_A(k) := \bigcup_{x \in D^i_A(k-1)} I^A(x)$$

Note that $D^i_A(k) \subseteq D^i_A(k+1)$ so that the sequence terminates. Let D^i_A denote the terminal element for each i and let $\mathcal{B}_A \equiv \{D^i_A\}$ be the candidate category. We claim that the system $\mathcal{B}_A$ is a coherent category such that $(u(\cdot), \mathcal{B}_A)$ represents $\succeq$. Completeness is obvious from definition, so that it remains to check coherence, representability, and downwards rigidity.

Coherence

Let $A' \subseteq A$. We wish to show that for any $D^i_{A'}$, there is some j such that $D^i_{A'} \subseteq D^j_A$. Note that once we obtain $D^i_{A'}(n) \subseteq D^j_A(m)$ for any integers $m \geq n$, then it follows that $D^i_{A'}(n+1) \subseteq D^j_A(m+1)$, and so on; so that $D^i_{A'} \subseteq D^j_A$. Now let $x \in \inf(I^i_{A'})$ and note that $I^{A'}(x) \subseteq D^j_A(1)$ for some j with $x \in I^j_A$ by completeness. Thus, $D^i_{A'}(1) \subseteq D^j_A(2)$.

Representability

Note that, from the preceding proposition, $\inf(I^x_A) \sim x$ whenever $x \in \Sigma_w(A)$. Thus, $\min_{x \in D^i_A(1)} u(x) = \min_{x \in I^i_A} u(x)$. Applying the same reasoning to each $D^i_A(n)$ we find: $\min_{x \in D^i_A(n)} u(x) = \min_{x \in D^i_A(n-1)} u(x) = \cdots = \min_{x \in I^i_A} u(x)$. It follows that, for each i, $\inf(D^i_A) = \inf(I^i_A)$. Fix some $x^i \in \inf(I^i_A)$. Since $I^1_A \succ I^2_A \succ \cdots \succ I^k_A$ (by the preceding claim), we obtain $U^{\mathcal{B}_A}(A) = \max_i \{u(x^1), \ldots, u(x^k)\} = u(x^1) = U(A)$.

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 $y \in \Sigma_w(A')$. Note that since $x \in \inf(I^k_A)$ and the indifference classes I^i_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 $D^j_A(0) := I^j_A$). Thus, we can assume $x \neq y$. Since $x \in D^j_A(n+1)$, this means that $x \in I^A(y)$, where $y \in D^j_A(n)$ - which, in turn, implies that $x \succ y$ (since $\succeq |_X$ is strict). On the other hand, we know (by Claim 1) that $u_{\min}(x, A) = u(x)$ - contradicting the requirement that $x \in I_{A'}(y)$. □

The importance of this result is that it ties the functional form of the category model to the process of compromise. Recall that our description of compromise involves two stages: (i) a reduction from an objective menu A to a subjectively determined feasible set A^* and (ii) standard maximization over the feasible set. The key to this process is understanding how a DM extracts a feasible subset A^* from a given menu A. To this end, Theorem 4 provides the answer - categorization. Thus, if we believe that $A1$ is a behavioral description of compromise, then we are naturally led to study the category model. We also note that the system of local categories is not uniquely pinned down by the menu preference $\succeq$. However, the construction given in the above proof is not *ad hoc* either. We demonstrate in the following two examples that it actually gives a concrete algorithm to produce a system of local categories (whenever the preference satisfies $A1$).

Example 6. *Denote weak compromises of menus with an underscore. Define a preference as follows:*

- $X = \{x_1 \succ \underline{x_2} \succ x_3\}$
- $\{\underline{x_1}, x_2\}, \{\underline{x_2}, x_3\}, \{x_1, \underline{x_3}\}$.

Put $A_1 = \{x_1, x_2\}, A_2 = \{x_2, x_3\}, A_3 = \{x_1, x_3\}$ *and note that* $\mathcal{B}_{A_1} = \{\{x_1\}, \{x_2\}\}$ *and* $\mathcal{B}_{A_2} = \{\{x_2\}, \{x_3\}\}, \mathcal{B}_{A_3} = A_3$. *Moreover, we have* $\mathcal{B}_X = \{\{x_1, x_2\}, \{x_1, x_3\}\}$. *Note that there two* $\succeq_X$ *indifference classes with* $\inf_X^1 = \{x_2\}, \inf_X^2 = \{x_3\}$.

For the second example we compute the categories $\mathcal{B}_A$ generated by a rank-based (Strotz) preference. Note that since $A1^*$ (DSB) implies $A1$ we know that all Strotz preferences admit representations by a category model.

Proposition 4. *Let* (u, L) *be a Strotz model. Then, for a given menu* A*, the category* $\{\mathcal{B}_A^i\}_i$ *(in the proof of Theorem 4) is defined via the following formula: (put* $x^i \in \inf(I_A^i)$ *and let* $I_A^1, \ldots, I_A^k$ *denote a top-down enumeration of the* $\succeq_A$ *indifference classes)*

$$\mathcal{B}_A^i = \{y \in A \backslash (\cup_{j \leq i-1} I_A^j) : y \succeq x^i\}$$

Moreover, the local category model $(u(\cdot), \{D_A\})$ *is such that* $U^{\mathcal{B}_A}(A) \equiv U^L(A)$.

3.2 Main Result

We now come to the main result of the paper, a representation result for the category model which generalizes the NAT category. The representation shows that the general category model is characterized by 3 axioms: Order, Compromise, and Strong Reduction. We now introduce the second main axiom of the section, the Strong Reduction axiom.

For a given menu A, I want to define a subset A^* of "subjectively feasible" elements. Intuitively, I consider $x \in A$ a (subjectively) feasible element if for all subsets $D \subseteq A$ such that $x \in D$ we have $D \succeq x$. That is, whenever x is available, then I can do no worse than x. One can check that (assuming order and $A1$) this is exactly the statement that there are no direct temptations or attraction sets for the element x in the menu A. Thus, following the notion of compromise outlined in the introduction, the idea is that the element x should survive the 'reduction' process that maps a menu A to the sub-menu A^* of (subjective) feasible consumption elements. Deliberately carrying over the A^* notation we put

$$A^* = \{x \in A : D \succeq x, \forall D \subseteq A \text{ s.t. } x \in D\}$$

Summarizing the preceding discussion, x is considered feasible relative to the menu A if whenever x is available, then the DM never has to compromise to something worse than x (since he can always choose x if he wants to). The following axiom, a strengthening of Reduction, makes two restrictions. First, the set of subjectively feasible elements in a menu determines the class of a menu - this is the analogue of Reduction when there are attractions as well as direct temptations. Second, if a subset of a menu contains all feasible, i.e. untempted, elements of the original menu, then the subset is in the same indifference class. We interpret this second point as a consistency restriction on the notion of feasibility. As will become clear from Lemma 4 (below), the effect of the restriction is that it makes the categories 'global', i.e. independent of the menu.

A2: (Strong Reduction) If $A^* \subseteq A' \subseteq A$, then $A' \sim A$.

The axiom says that if A' is contained in A but contains all of the feasible elements from A, then A' is equivalent to A. Note the link with the Reduction axiom ($A2^*$) which said that if all the simple temptations in A are present in A', then A' is indifferent to A. Call a menu A *simple* if $A = A^*$ - that is, the menu, and all of its subsets, are compromise-free. The following lemma, which shows that $A1 + A2$ implies a form of Set-Betweenness (after GP (2001)). An implication of the lemma (and the ensuing representation result) is that the category model allows non-trivial Set-Betweenness on menus, despite the fact that the DM is always compromising on any given menu.

Lemma 3. *Assume* $\succeq \in \boldsymbol{P}(X)$ *satisfies* $A1 + A2$ *and let* A *be simple. Then, for any* $y \in X$, $\{y\} \succeq A \cup \{y\} \succeq A$, *when* $y \succeq A$ *and* $A \succeq A \cup \{y\} \succeq \{y\}$ *when* $A \succ y$.

Recall Lemma 2 from section 2 which said that Reduction implied transitivity of simple temptations, i.e. $x \rightarrow_t y$ and $x \rightarrow_t z$ implies $x \rightarrow_t z$. In words, all simple temptation of y are simple temptations of x if y is, itself, a simple temptation of x. Our first observation is that Strong Reduction ($A2$) yields a generalization of this property when temptations can be latent. We refer to the generalized property as *substitution*. Fix $x \in X$ and say that a menu $A(x)$ is an attraction set (of x) if we have (i) $x \succ y, \forall y \in A(x)$, (ii) $x \succ A(x) \cup x$, and (iii) $x \sim (A(x) \backslash y), \forall y \in A(x)$. That is, the menu $A(x)$ tempts x, but only does so when all components of the attraction set $A(x)$ are present. Summarize (i), (ii), and (iii) with the common notation $x \rightarrow_t A(x)$. We then have the following generalization of Lemma 2.

Lemma 4. *Assume* $\succeq$ $\boldsymbol{P}(X)$ *satisfies* A1 *and* A2. *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 $x \succ x \cup A(y) \cup (A(x)\backslash y)$.[9] Similarly, we can inductively extend this result to arbitrary chains of direct/latent temptations.

Say that a pair $(u, \mathcal{C})$ is a *flexible* category if $\cup_i \mathcal{C}_i = X$. Thus, the *only* restriction on the sets $\mathcal{C} \equiv \{\mathcal{C}_i\}$ is that they cover X. The following is the representation theorem for the flexible category model and provides a generalization of the NAG category representation.

Theorem 5 (Main Result). *A preference* $\succeq \in \boldsymbol{P}(X)$ *satisfies* A1 *(Compromise) and* A2 *(Strong Reduction) if and only if it admits a flexible category representation.*

The proof of this theorem is a little long, and the construction of the categories required to model attraction (aggregation) does not resemble the earlier constructions - where we just took collections of simple temptations. Also note that if $(u, \mathcal{C})$ is a flexible category representation, then it is straightforward to use this to obtain a system $\mathcal{B}_A$ of local categories. For example, enumerate $\mathcal{C} \equiv \{\mathcal{C}_1, \ldots, \mathcal{C}_n\}$ and note that the collection $\mathcal{B}_A \equiv \{\mathcal{C}_1 \cap A, \mathcal{C}_2 \cap A, \ldots, \mathcal{C}_n \cap A\}$ is already coherent and represents $\succeq$. We omit the straightforward modification of this system that also delivers downwards rigidity.

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 is that "a (latent/direct) temptation of a (latent/direct) tempation of $\cdots$ of a (latent/direct) temptation of x is itself a (latent/direct) temptation of x." This gives us a hint as to how to account for the temptations of x. 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 the root node denotes the singleton $\{x\}$, each of the new nodes denotes a doubleton $\{x, x_i^1\}$ for some x_i^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 $\{x, x_i^1\}$ create $|A_2(x)|$ branches. The new nodes at the ends of these branches denote triples of elements, $\{x, x_i^1, x_j^2\}$, where x_j^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 in the node and repeat this procedure.

[9]Proof: Note that $x, y \notin (x \cup y \cup A(y) \cup A(x))^*$, so that $(x \cup y \cup A(y) \cup A(x))^* \subseteq A(y) \cup (A(x)\backslash y)$. Strong Reduction then implies $x \cup (A(y) \cup (A(x)\backslash y)) \sim A(y) \cup (A(x)\backslash y)$, so that $x \succ x \cup A(y) \cup (A(x)\backslash y)$.

Formally, this is where the recursivity 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 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. Elements in nodes of z-trees, where z is a latent x-temptation, are also latent x-temptations - hence, should be contained in any category that contain x. This is the construction in a nutshell. Note that we are being quite wasteful in this construction. That is, many of the categories created in this tree-like construction will never be pivotal for the comparison for any pair of menus (A, B). Hence, the representation will not generally be sharp. The work of the identification corollary which follows the Theorem is to see whether this issue can be suitably repaired.

Proof. Necessity of $A1$ is straightforward, hence we omit the argument. Necessity of $A2$ is also straightforward, but we nevertheless provide the argument since it justifies why $A2$ contains the word 'reduction' in its title. To this end, let $(u, \mathcal{C})$ be a (flexible) category and let $\succeq$ denote the underlying menu preference represented by $(u, \mathcal{C})$. Let $A^* = \{x \in A : D \succeq A, \forall D \subseteq A, x \in D\}$ be the "no compromise" subset of A and let $A^*_{(u,\mathcal{C})} = \cup_i \arg\min_{x \in \mathcal{C}_i \cap A} u(x)$, where we take $\mathcal{C} \equiv \{\mathcal{C}_i\}$. Note that $A2$ 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 $\Sigma(x) = \{\mathcal{C}_i \in \mathcal{C} : x \in \mathcal{C}_i\}$. I 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 $A' \succeq A$, 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 I 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.

Step 1: Constructing x-trees.
For $x \in \Sigma_1$ we define an object called an x-tree. Introduce the following terminology. Fix an index set $\{1, 2, \ldots, N\}$. An x-tree is a triplet of data $(\{\mathcal{C}^i_j\}_{j=1}^{n_i}, \{\mathcal{L}_i\}_{i=1}^N, \{\mathcal{C}^j_i \to \mathcal{C}^k_{i+1}\})$, consisting (resp.) of nodes, levels, and branches, with the following structure:

- A collection of *nodes* $\mathcal{C}^i_1, \mathcal{C}^i_2, \ldots, \mathcal{C}^i_{n_i}$ for each index i.
- A collection of *levels* $\mathcal{L}(1), \ldots, \mathcal{L}(N)$, where each $\mathcal{L}_i := \{\mathcal{C}^i_1, \mathcal{C}^i_2, \ldots, \mathcal{C}^i_{n_i}\}$

- A collection of *branches* $\{\mathcal{C}_i^j \to \mathcal{C}_{i+1}^k\}$ connecting nodes on consecutive levels. Call $\mathcal{C}_i^j$ the *root* of the branch $\{\mathcal{C}_i^j \to \mathcal{C}_{i+1}^k\}$.
- Every node $\mathcal{C}_k^i$ in level i (for $i > 0$) has a unique root in level $i-1$.
- Every node $\mathcal{C}_k^i$ in level i is the root of a branch.

For each $x \in \Sigma_1$ I inductively construct an x-tree as follows. First make the following simplification. Since representability requires that $B_t(x) \in \mathcal{C}_i$ whenever $x \in \mathcal{C}_i$, I make no distinction between the element x and the set $\{x\} \cup B_t(x)$. That is, whenever I say $x \in \mathcal{C}_i$ I implicitly mean (unless explicitly stated otherwise) that $B_t(x) \cup \{x\} \subseteq \mathcal{C}_i$. Let $A_1(x), A_2(x), \ldots, A_n(x)$ be the compound temptations of x (here I distinguish between x and $B_t(x) \cup \{x\}$). The x-tree construction is as follows. Proceed by double-induction. The outer induction is on the $\succeq$-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 $\succeq$-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 $\succeq$-rank to construct an x-tree as follows. Let $A_i(x) = \{x_1^i, \ldots, x_k^i\}$ be the elements of the compound temptation set and take

$$\mathcal{C}_i^1 = \{x, x_i^1\}, \mathcal{L}_1 = \{\mathcal{C}_1^1, \mathcal{C}_2^1, \ldots, \mathcal{C}_k^1\}$$

Note that the sizes of the compound temptation 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 compound temptations for x) and define level $m+1$ as follows. Let $\{\mathcal{C}_m^i\}_{i=1}^{N_m}$ be an enumeration of the nodes that form $\mathcal{L}_m$. For each node $\mathcal{C}_m^i$ 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 $\mathcal{C}_{m+1}^1 = \mathcal{C}_m^1 \cup \{x_1^{m+1}\}, \mathcal{C}_{m+1}^2 = \mathcal{C}_m^1 \cup \{x_2^{m+1}\}, \ldots, \mathcal{C}_{m+1}^k = \mathcal{C}_m^1 \cup \{x_k^{m+1}\}$. Similarly, put $\mathcal{C}_{m+1}^{(i-1)k+1} = \mathcal{C}_m^i \cup \{x_1^{m+1}\}, \mathcal{C}_{m+1}^{(i-1)k+2} = \mathcal{C}_m^i \cup \{x_2^{m+1}\}, \ldots, \mathcal{C}_{m+1}^{ik} = \mathcal{C}_m^i \cup \{x_k^{m+1}\}$. Thus, level $\mathcal{L}(m+1)$ consists of $N_m \cdot |A_{m+1}(x)|$ nodes, $\mathcal{C}_{m+1}^j$, and $N_m \cdot |A_{m+1}(x)|$ branches, $\{\mathcal{C}_m^i \to \mathcal{C}_{m+1}^j\}$ (where $(i-1) \cdot k + 1 \leq j \leq i \cdot k$). Inductively proceed until we exhaust all compound temptation sets $\{A_1(x), \ldots, A_n(x)\}$. Let $\mathcal{L}(n) = \{\mathcal{C}(1), \mathcal{C}(2), \ldots, \mathcal{C}(N)\}$ be an enumeration of the nodes at level $\mathcal{L}(n)$. Take any y with $x \succ y$ and for each $\mathcal{C}(i)$ with $y \in \mathcal{C}(i)$ 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. Note that for each $\mathcal{C}(i)$ in level $\mathcal{L}(n)$ that does not contain y we just extend a single branch $\{\mathcal{C}(i) \to \mathcal{C}_{n+1}(i)\}, \{\mathcal{C}_{n+1}(i) \to \mathcal{C}_{n+2}(i)\}$, and so on, for each subsequent level, where we put $\mathcal{C}(i) = \mathcal{C}_{n+1}(i) = \cdots = \mathcal{C}_{n+M}(i)$ (here I 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 any $z \neq y$ with $x \succ y$ and for each $\mathcal{C}(i) \in \mathcal{L}(n+M)$ with $z \in \mathcal{C}(i)$ attach a z-tree. Iteratively proceed as above. Since X is finite, this process terminates at some level $\mathcal{L}(N_x)$. This concludes the construction of the x-tree.

Step 2: Check the inequality $U^{\mathcal{C}}(\cdot) \leq U(\cdot)$.
Having defined the x-tree for each $x \in X$ we take the category, $\mathcal{C}$, to be the set of all terminal nodes in the level $\mathcal{L}(N_x)$ for every $x \in X$. Taking u to be a representation of the singleton ranking, the claim is that the pair $(u, \mathcal{C})$ represents $\succeq$. Let $U(\cdot)$ be any cardinal representation of $\succeq$ (which extends $u(\cdot)$). As in the other proofs we show representability by checking equality $U^{\mathcal{C}}(\cdot) = U(\cdot)$ on all menus. The tree structure of the categories allows for a useful decomposition of the function $U^{\mathcal{C}}(\cdot)$. Let $\mathcal{L}(N_y)$ 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

$$U^{\mathcal{T}_y}(A) := \max_{\mathcal{C}_i \in \mathcal{L}(N_y)} \min_{z \in A \cap \mathcal{C}_i} u(z)$$

Observe that we have the equality

$$U^{\mathcal{C}}(A) = \max_{\mathcal{T}_{y_i}} U^{\mathcal{T}_y}(A)$$

Thus, the value the category utility assigns to menu A is the maximum of its value across trees. Thus, we may analyze $U^{\mathcal{C}}(\cdot)$ by analyzing its behavior on a tree-by-tree basis. For a given tree $\mathcal{T}_y$ let $\mathcal{L}(1), \ldots, \mathcal{L}(N_y)$ denote its levels. For each level $\mathcal{L}(i)$ consider the function

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

and note that if, say, $A \cap \mathcal{C}_j \neq \emptyset, \forall \mathcal{C}_j \in \mathcal{L}_i$ then we have

$$U^{\mathcal{L}(i)}(A) \geq U^{\mathcal{L}(i+1)}(A) \geq \cdots \geq U^{\mathcal{L}(N_y)}(A) = U^{\mathcal{T}_y}(A)$$

Let x_A denote a compromise in A and consider any $x \in A$ with $x \succ x_A$. Take any $z \in \inf(I_A(x))$ and note that $x_A \succeq z$. Note that either $z \in B_t(x)$ or the menu A contains one of the x-compound temptations, say $A(x)$. Consider any tree $\mathcal{T}_y$ and note that the value $u_{\min}(x, A)$ is attained on $\mathcal{T}_x$. Take $A(x) \subseteq A$ with $A(x) \sim z$ (we are implicitly applying Strong Reduction here) and consider the level $\mathcal{L}(i)$ of the tree $\mathcal{T}_x$ at which we place elements of the temptation set $A(x)$ in each node from the preceding level. By construction, $\min_{y \in \mathcal{C}_j \in \mathcal{L}(i)} u(y) \leq u(z)$. Applying the preceding monotonicity condition, we then obtain that $\min_{y \in \mathcal{C}_j \in \mathcal{L}(N_x)} u(y) \leq u(z)$. It follows that, for each $x \in A$, the value of the category function $U^{\mathcal{C}}(\cdot)$ on the x-tree $\mathcal{T}_x$ is bounded above by $u_{\min}(x, A)$. This implies that $U^{\mathcal{C}}(A) \leq U(A)$.

Step 3: Check the inequality $U^{\mathcal{C}}(\cdot) \geq U(\cdot)$.
We now check the reverse inequality. First introduce some terminology. For notational brevity I denote level i tree nodes as $x_j(i)$ (the j-th node in level i) and branches are denoted $x_j(i) \rightarrow x_k(i+1)$. Fix a menu A and some $x \in A$. Consider

the x-tree $\mathcal{T}_x$ and consider the set of all directed paths along branches in the tree, $\Phi := \{\ell : \ell = \{x \to x_{i_1}(1) \to x_{i_2}(2) \to \cdots \to x_{i_{N_x}}(N_x)\}\}$. Let $\ell = (\ell_1, \ell_2, \ldots, \ell_{N_x})$ denote the nodes along the path ℓ. I say that x is *unobstructed* in $\mathcal{T}_x$ by the menu A if there is a path $\ell \in \Phi$ such that $\ell_i \cap A = \{x\}, \forall \ell_i$. I claim that x is unobstructed (by a 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 $\mathcal{T}_x$ is $u(x)$. It follows that $x \in A^*$. Now for the reverse direction. Proceed by induction on the $\succeq$-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 $\mathcal{T}_z$ 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 $\mathcal{T}_x$ is obstructed in A, it can only be obstructed by elements with $\succeq$-rank 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. I 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 $z_i \in A$. 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 I will need to concatenate paths from different levels (in the x-tree $\mathcal{T}_x$). Let $\ell(i,j)$ denote a path connecting a node in level $\mathcal{L}(i)$ to a node in level $\mathcal{L}(j)$. 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 $\mathcal{L}(k)$ of the x-tree that contains $\{x, z_1, \ldots, z_k\}$. Let $\ell(0,k)$ denote the unobstructed path from level 0 (with the singleton node $\{x\}$) 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 $y_1, y_2, \ldots, y_l$ where $y_1 = z_{i_1}$ (for some $z_{i_1} \in \{z_1, \ldots, z_k\}$). 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 $\mathcal{L}(N_1), \ldots, \mathcal{L}(N_l)$ denote the terminal levels of the y_i-trees being attached to the nodes at level $\mathcal{L}(k)$. I 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.

Since $y_1 \in \{z_1, \ldots, z_k\}$ (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 $\mathcal{L}(k + N_1)$ of the partial tree obtained by concatenating the y_1-tree to the preceding levels $\mathcal{L}(1), \mathcal{L}(2), \ldots, \mathcal{L}(k)$. Concatenate $\ell(0, k)$ to $\ell(k, k + N_1)$ to obtain an unobstructed path from level $\mathcal{L}(0)$ to level $\mathcal{L}(k + N_1)$. 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 $\{x(k+N_1) := \mathcal{C}_0 \to \mathcal{C}_1\}, \{\mathcal{C}_1 \to \mathcal{C}_2 = x(k + N_1)\}, \ldots, \{\mathcal{C}_{N_1-1}(= x(k + N_1)) \to \mathcal{C}_{N_1}(= x(k + N_1))\}$. 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, replicate the argument for the y_1 case to extend the unobstructed path. If y_2 is not in the unobstructed set, then y_2 is first introduced at some level $\mathcal{L}(k + k_y)$ of the $y_1(= z_{i_1})$-tree. 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 recursivity 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 $x(k + N_1 + N_2)$. Inductively extend the path to obtain a sequence (put $k_i = k + \sum_{j=1}^{i} N_j$) $(\ell(0, k); \ell(k, k_1); \cdots; \ell(k_{l-1}, k_l))$. 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 $U^{\mathcal{C}}(\cdot)$ on the tree $\mathcal{T}_x$ is $u(x)$, proving that $U^{\mathcal{C}}(A) \geq U(A)$. □

It is possible to obtain an identification result for the flexible category. First note that the flexible category model does not have a non-redundance restriction. When we relax the redundance restriction on the category model, it is not possible to point identify the representing category. As a trivial example, let $(u, \mathcal{C})$ be a non-redundant category representation of some preference $\succeq$. Let x_* be a $\succeq$-minimal singleton and consider the following category. Let $\mathcal{C}' := \{\mathcal{C}, \{x_*\}\}$. That is, we have simply added the singleton set $\{x_*\}$ to the original collection $\mathcal{C}$. It is easy to see that if the original pair $(u, \mathcal{C})$ represents $\succeq$, then the pair $(u, \mathcal{C}')$ also represents $\succeq$. Moreover, the category $\mathcal{C}'$ is obviously redundant.

As we pointed out in the section on simple compromise, this issue can be circumvented when restrict attention to category models $(u, \mathcal{C})$ that are possibly redundant, but still *sharp* - i.e. every set $\mathcal{C}_i \in \mathcal{C}$ is necessary for the comparison of some pair of menus (A, B). Readers familiar with DLR (2001) will note that this is analogous to the notion of a "relevant state," which is a necessary refinement in their identification result as well. The following corollary shows that when we restrict attention to sharp category models $(u, \mathcal{C})$, then flexible category representations can be weakly identified in the following sense. 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 $\pi : \mathcal{C}_1 \to \mathcal{C}_2$ where for every set $\mathcal{C}_1(i) \in \mathcal{C}_1$ we have $\pi(\mathcal{C}_1(i)) \supseteq \mathcal{C}_1(i)$. This refinement was also used in the identification result for the NAT category.

The identification result for the flexible category is as follows. Let $(u, \mathcal{C})$ be the "tree-category" constructed in the sufficiency proof of the theorem. The corollary below shows that there is a unique sharp sub-category $(u, \mathcal{C}^*)$ of the tree category with the property that *every* sharp category model $(u, \mathcal{C}')$ that represents the preference is a prolongation of the model $(u, \mathcal{C}_*)$. We illustrate with an example.

Let $X = \{x, y, z, p, q\}$ where x is attracted to $\{y, z\}, \{p, q\}, \{y, q\}$, and $\{p, z\}$. Also put $y \to_t p$ and $z \to_t q$, and assume there are no other temptations (direct or latent). Consider the following category,

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

where $\mathcal{C}_* \equiv \{\mathcal{C}_1, \mathcal{C}_2\}$. Note 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:

1. $\mathcal{L}(1) \equiv \{\mathcal{C}^1_1 = \{x, y, p\}, \mathcal{C}^1_2 = \{x, z, q\}\}$.
2. $\mathcal{L}(2) \equiv \{\mathcal{C}^2_1 = \{x, y, p\}, \mathcal{C}^2_2 = \{x, y, p, q\}, \mathcal{C}^2_3 = \{x, z, p, q\}, \mathcal{C}^2_4 = \{x, z, q\}\}$.
3. $\mathcal{L}(3) \equiv \{\mathcal{C}^3_1 = \{x, y, p\}, \mathcal{C}^3_2 = \{x, y, p, q\}, \mathcal{C}^3_3 = \{x, y, p, q\}, \mathcal{C}^3_4 = \{x, y, p, q\}, \mathcal{C}^3_5 = \{x, y, z, p, q\}, \mathcal{C}^3_6 = \{x, z, p, q\}, , \mathcal{C}^3_7 = \{x, z, y, p, q\}, \mathcal{C}^3_8 = \{x, z, q\}\}$.
4. $\mathcal{L}(4) \equiv \{\mathcal{C}^4_1 = \{x, y, p\}, \mathcal{C}^4_2 = \{x, y, p, z, q\}, \mathcal{C}^4_3 = \{x, y, p, q\}, \mathcal{C}^4_4 = \{x, y, p, z, q\}, \mathcal{C}^4_5 = \{x, y, p, q\}, \mathcal{C}^4_6 = \{x, y, p, z, q\}, \mathcal{C}^4_7 = \{x, y, p, q\}, \mathcal{C}^4_8 = \{x, y, p, z, q\}, \mathcal{C}^4_9 = \{x, y, z, p, q\}, \mathcal{C}^4_{10} = \{x, y, z, p, q\}, \mathcal{C}^4_{11} = \{x, z, p, q\}, \mathcal{C}^4_{12} = \{x, z, p, q\}, \mathcal{C}^4_{13} = \{x, z, y, p, q\}, \mathcal{C}^4_{14} = \{x, z, y, p, q\}, \mathcal{C}^4_{15} = \{x, z, p, q\}, \mathcal{C}^4_{16} = \{x, z, q\}\}$

Let us make three observations about this example. First, the construction is very wasteful. The tree construction, while producing a representing category, also produces many categories which are either redundant or otherwise irrelevant for the comparison of any two menus. Second, note that in the tree construction we still consider nodes as formally distinct even though the underlying sets are identical (e.g. we think of $\mathcal{C}^4_6$ and $\mathcal{C}^4_8$ as distinct since they represent different paths in the tree even though, as subsets of X, they are identical). This may seem like an artificial distinction, but it turns out to be important in our proof since we will select paths in the tree with specific properties. In order to prove the existence of such paths, I need to leave the entire structure of the tree intact. Of course, after I show representability then we can delete redundant sets since we then know which ones we need for the representation. The key point is that this can only be done *after* we prove the tree construction yields a representing collection of categories.

Third, and most importantly, notice that the category $\mathcal{C}_*$ is a sub-category of the terminal nodes of 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 $(u, \mathcal{C})$. The first step is to use the concepts of prolongation and sharpness to realize this model as a prolongation of some sub-category $(u, \mathcal{C}_*)$ of the tree category. This shows that there is a minimal sharp retraction of the model $(u, \mathcal{C})$ that sits inside the tree category, we call this the "embedding" step. The second and more difficult step, called pruning, shows that there is a unique sharp sub-category of the tree category. The proof is in the appendix.

Corollary 4 (Uniqueness). *Let $(u_1, \mathcal{C}_1)$ and $(u_2, \mathcal{C}_2)$ be any two sharp representations of $\succeq$ (where $\succeq|_X$ is strict). Then, u_1 and u_2 are ordinally equivalent and $\mathcal{C}_1$ and $\mathcal{C}_2$ are prolongations of the unique sharp sub-category, $(u, \mathcal{C}_*)$, of the tree category.*

We have suggested that the notion of sharpness is akin to the Dekel, et al. (2001) notion of a relevant state. However, our identification results (for the NAT and flexible category models) use an additional refinement, the idea of prolongation. 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 'no' - so that the identification we provide is the strongest form of identification that is available.

Let $X = \{x \succ y \succ z \succ p \succ q\}$. Assume $A_1(x) = \{z, p\}, A_2(x) = \{z, q\}$, and $p \to_t q, y \to_t p$. Put $\mathcal{C}_1 = \{y, p, q\}, \mathcal{C}_2 = \{x, z\}, \mathcal{C}_3 = \{x, p, q\}$ and consider the category model $(u, \mathcal{C})$ given by $\mathcal{C} \equiv \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\}$. Note that $(u, \mathcal{C})$ is a sharp representation of $\succeq$. Now consider the following prolongation of $\mathcal{C}$. Put $\mathcal{C}'_1 = \{x, y, p, q\}, \mathcal{C}'_2 = \mathcal{C}_2, \mathcal{C}'_3 = \mathcal{C}_3$ and let $\mathcal{C}' \equiv \{\mathcal{C}'_1, \mathcal{C}'_2, \mathcal{C}'_3\}$. Note that the model $(u, \mathcal{C}')$ also represents $\succeq$ and, importantly, is also sharp - as y is not a latent temptation for x we cannot omit $\mathcal{C}'_3$ from the representation. Thus, we need the prolongation refinement in order to pin down the model.

4 Discussion

This paper presents an axiomatic model of compromise that contributes to the decision theory literature on temptation and self-control. We introduce a class of utilities, referred to as the category model, that can capture several features that are germane to the issue of temptation, e.g. multiple sources of temptation, aggregation, attraction effects. However, in contrast with most other papers in the temptation literature (GP (2005) notwithstanding) we model these issues within a framework of compromise, so that the DM never exercises costly self-control as in GP (2001).

The max-min structure of the category model might look somewhat strange to readers familiar with Gul-Pesendorfer style temptation models. For example, it might seem *ad hoc* that the DM evaluate within categories using the min function, but then take the maximum across categories. We now clarify the connection between the category model and more standard temptation models, which requires going into a little detail. Recall the Positive Set-Betweenness axiom, mentioned in the proof of local category representation. This axiom is implied by the no-uncertainty version of the DLR (2009) utility. We now recall a utility functional (used in the proof of Theorem 4) that is characterized by the Positive Set-Betweenness axiom. Fix a function $U : 2^X \to \mathbf{R}$ 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. Notice that in a Gul-Pesendorfer representation we have[10]

$$u_{\min}(x, A) = u(x) + v(x) - \max_{z\in A} v(z) = u(x) - c(x, A)$$

so that the GP utility can be expressed as $U^{GP}(A) = \max_{x\in A} u_{\min}(x, A)$. Similarly, consider a category pair $(u, \mathcal{C})$ and let $U(\cdot)$ be the utility generated by the formula $U(A) = \max_{\mathcal{C}_i} \min_{z\in A\cap\mathcal{C}_i} u(z)$. For each $x \in X$, let $\Sigma(x) := \{\mathcal{C}_x \in \mathcal{C} : x \in \mathcal{C}_x\}$. That is, $\Sigma(x)$ is the sub-collection of sets in $\mathcal{C}$ that contain x. Notice that

$$u_{\min}(x, A) = \max_{\mathcal{C}_x\in\Sigma(x)} \min_{z\in\mathcal{C}_x\cap A} u(z)$$

Putting $\mathcal{C} \equiv \{\mathcal{C}_i\}_i$ note that we can simply re-organize the sets in $\mathcal{C}$ as $\cup_{x\in X} \Sigma(x)$. It follows that

$$\begin{aligned} U(A) &= \max_{\mathcal{C}_i} \min_{z\in\mathcal{C}_i\cap A} u(z) \\ &= \max_{x\in X} \left[\max_{\mathcal{C}_x\in\Sigma(x)} \min_{z\in\mathcal{C}_x\cap A} u(z)\right] \\ &= \max_{x\in X} u_{\min}(x, A). \end{aligned}$$

Thus, the category model and the GP (2001) model (and the no-uncertainty version of DLR (2009)) are both specializations of a more general functional which is character-

[10]When x is "overwhelmed" by a temptation in A this equality does not hold, but the equality between the max-min utility on menus and the GP utility does always hold - its only the case that the maximands aren't necessarily pointwise equal.

ized by the Positive Set-Betweenness axiom.[11] Consequently, the max-min structure is a thread that runs through *all* of the standard (no-uncertainty) models of temptation.

We now discuss some (possible) alternative approaches to modeling the issues addressed by this paper. First, note that the category model is not a canonical generalization of Strotz. For example, one could take the Strotz model as a baseline model of compromise and try to more directly generalize the elements of the Strotz representation. Recall from GP (2005) that the temptation utility $v(\cdot)$ is a representation of the following binary relation on elements of X (assume no ties on singletons for this discussion),

$$x\mathcal{R}y \Leftrightarrow [x \not\to_t y \vee y \to_t x]$$

Since this relation is transitive (and complete), simple compromise cannot be represented - else it would create a cycle in the v-ranking of a putative (u, v) representation. Thus, one approach to modeling simple compromise is to keep the dual self idea intact but to relax the requirement that $\mathcal{R}$ be transitive. Since $A1$ allows for simple compromise, and hence cycles, it follows that the induced binary relation $\mathcal{R}$ is intransitive. This suggests the following alternative representation. Let $v_1, \ldots, v_k$ be (completions of) the temptation representations on the transitive components of the relation $\mathcal{R}$ (viewing the relation $\mathcal{R}$ as a subset of $X \times X$ these are just the connected components of the corresponding directed graph) and consider the following Bewley-style "multi-self" representation: (let $u \circ v$ denote the Strotz utility on menus)

$$A \succeq B \Leftrightarrow u \circ v_i(A) \geq u \circ v_i(B), \forall i = 1, 2, \ldots, k$$

This is an interesting representation and we have not investigated the connections between this utility and the category model. Certainly they aren't equivalent as this model is incomplete, but there may, nevertheless, be some interesting connections between the observables of the two models. We also note, however, that our goal in this paper is to provide a theory of compromise that tightly models the compromise and attraction effects, not to generalize the Strotz utility. That the category model nests the Strotz model is, in our view, an accidental benefit but not the main feature of the theory.

A second approach to modeling the issues we have raised is to take a different behavioral primitive. In particular, we could assume that the modeler's observable is ex post choice (as opposed to ex ante choice - which is what we have done here). That is, the primitive is a choice correspondence which maps a menu A into a subset of itself. One could then carry out the exercise of the paper on the ex post choice domain. The

[11] The observation that Positive Set-Betweenness characterizes functions with this general structure is a result of K. Nehring, who provides a related (but formally distinct) characterization of Positive Set-Betweenness. The exact statement and proof of the max-min characterization of Positive Set-Betweenness we have stated is a (mild) generalization of a lemma in GP (2001) and may be found in Chandrasekher (2009).

goal being to characterize when this choice correspondence may be recovered as the arg max of a category model. In order to disentangle normative choice from choice induced by temptation, this approach requires two observables: (i) a choice correspondence and (ii) a ranking on singleton menus (choices). With these two in hand, the axioms and analysis can be transferred to the ex post choice domain. Moreover, one can go in the other direction as well. That is, given a choice correspondence and a normative ranking on singletons, say $(\succeq^*, C(\cdot))$ one can (i) translate the axioms on these observables to axioms on menus (which are, of course, just Compromise and Strong Reduction) and (ii) induce an order, $\succeq$, on menus such this order satisfies the corresponding axioms. In sum, there is a one to one map between menu orders that satisfy Compromise and Strong Reduction and pairs, $(\succeq^*, C)$, of ex post observables given via the obvious map.[12] Hence, the two approaches are *equivalent*. Note that this is in stark contrast to all GP-style models of temptations - since the ex post choice map and normative ranking cannot, by themselves, pin down the order on menus. This is not surprising in view of the fact that the GP model and its successors are models of costly self-control. In order to reveal self-control, one necessarily has to use the richer primitive of an order on menus.

Finally, we conclude with a discussion of some recent and related papers which follow, at least in spirit, the ex post choice approach described above. Let us begin with Manzini-Mariotti (2010) ("Categorize then Choose" - CTC). The observable in this paper is an ex post choice function (mapping menus into choices) and the goal of the authors is to explain observable choices via the following two-stage "categorize then choose" procedure. First, the DM sorts (with a possibly incomplete sorting procedure) menus into categories. Then, for a given menu he eliminates elements that lie in dominated categories (under the sorting procedure). Second, he maximizes an (unobservable) preference relation over the undominated elements in the menu. The resulting element is his choice. Manzini-Mariotti (2010) show that this procedure characterizes choice functions that satisfy a weakening of Arrow's well-known WARP axiom.[13] The CTC procedure is an apt description of the decision procedure axiomatized in this paper. Indeed, from a menu preference induced by a category model $(u, \mathcal{C})$ we obtain an implied ex post choice function (when there are no ties on singletons). Our model in this paper is more narrow than the CTC model. The conceptual goal of the CTC model is to provide a general model of bounded rationality, whereas our goal is to provide a temptation-driven model that rationalizes the compromise

[12]One of the axioms on the ex post observables that we require to make this correspondence well-defined is that, for any menu A we have $x \sim^* y, \forall x, y \in C(A)$. That is, all elements in the choice correspondence have the same normative value. Now from a pair $(\succeq^*, C(\cdot))$ we define an induced order $\succeq$ via: $A \succeq B \Leftrightarrow x \succeq^* y$ where $x \in C(A), y \in C(B)$. Going the other way, given $\succeq$ we put $C(A) = \Sigma_w(A)$ (the set of weak compromises) and $\succeq^* := \succeq|_X$.

[13]More precisely, from observable choices they construct two unobservables, (i) a sorting procedure and (ii) a preference on singletons, such that choices from menus are generated by a CTC procedure using these two unobservables.

and attraction effects alone.

A second related paper is Masatlioglu-Nakajima-Ozbay (2009). This paper also takes ex post choice as the observable and characterizes choice functions generated by the following "limited consideration" procedure. First, from each menu the DM extracts a subset, e.g. the only options that catch his attention. Second, he maximizes a (unobservable) preference relation over this subset of options. The resulting choice function is called a "choice by limited consideration" (CLC) function. The paper also shows that this procedure characterizes choice functions that satisfy a weakening of the WARP axiom.[14] Choice functions generated by category models can be realized as CLC choice functions. To see this, take the consideration map to be the function that maps a menu A to the set A^* of no-compromise elements and take the (hidden) preference relation to be the restriction of the menu order to singleton menus. Using the CLC procedure with this pair recovers the choice function implied by the category model. Also akin to Manzini-Mariotti (2010), the goal of Masatlioglu, et al. (2009) is to provide a general model of bounded rationality. Consequently, the CLC class is much larger than the set of choice functions generated by category models.

Lastly, we discuss de Clippel-Eliaz (2010). The similarity between our paper and the preceding two examples was in the means (of the analysis), but not in the ends. In contrast, the similarity with de Clippel-Eliaz (2010) is in the ends, whereas the divergence is in the means. As in this paper, the goal is to provide a concise model that explains the compromise and attraction effects.[15] The behavioral primitive is a pair of (strict) orders on the set of consumption options. For example, imagine competing selves in the DM's mind who may have non-aligned rankings over consumption options. Taking choices on menus to be generated by a bargaining procedure between the two selves, the authors axiomatize a choice function, called the *fallback* solution, that generates the compromise and attraction effects. Interestingly, the fallback solution picks the choice from a menu that is "least worst" among the two selves. Choices generated by the category model also have a "least worst" feature, although there does not seem (to us) to be an obvious way to embed these functions into the class of fallback choice functions (or vice versa). These are the formal differences in the approaches. The main conceptual difference is in *how* we explain the compromise and attraction effects. In contrast to the multi-self approach, our paper points to temptation as a common source for both effects and suggests a mechanism (the category model) that shows explicitly how the presence of temptation generates these effects.

[14]More precisely, from observable choice the authors extract (i) an "attention filter" and (ii) a preference relation on singletons such that this pair generates the choice data via the CLC procedure.

[15]The attraction effect as described in de Clippel-Eliaz (2010) is formally distinct from, but similar in spirit to, our description of this phenomenon. Their description of the attraction effect is more faithful to the experimental literature in which the attraction phenomenon was documented.

5 Appendix

5.1 Omitted Proofs for Section 2

Proof of Lemma 1. Consider the set $\Sigma(A) = \{x \in A : x \sim A\}$. We claim that there is some $x_A \in A$ such that $\forall B \subseteq A$ with $x_A \in B$ we have $B \sim A$. Otherwise, for each x such that $x \sim A$ there is a subset $B(x) \subseteq A$ with $x \in B(x)$ and $A \not\sim B(x)$. Put $A' := A \backslash \cup_{x\in\Sigma(A)} B(x)$ and note that $A = \cup_{x\in\Sigma(A)} B(x) \cup A'$. By DSB, $A \sim A'$ which implies (by iterative application of DSB) that $\exists x \in A'$ such that $x \sim A$. On the other hand, $x \in B(x)$ and $B(x) \cap A' = \emptyset$ - contradiction. □

Proof of Lemma 2. By $A2^*$, $\{y, z\} \sim z$ implies that $\{x, y, z\} \sim \{x, z\}$. Similarly, $\{x, y\} \sim y$ implies that $\{x, y, z\} \sim \{y, z\}$. Thus, $\{x, z\} \sim \{x, y, z\} \sim \{y, z\} \sim z$ so that $x \to_t z$. □

Proof of Proposition 1. We only check sufficiency as necessity is obvious. Let $\Sigma_1, \ldots, \Sigma_k$ be a top-down enumeration of the $\succeq|_X$-indifference classes. Fix a class Σ_i and for $z_1, z_2 \in \Sigma_i$ say that $z_1 \mathcal{R} z_2$ if $\{x : x \to_t z_1\} \cap \{x : x \to_t z_2\} \neq \emptyset$. Note that, by $A4$, $\mathcal{R}$ is an equivalence relation on Σ_i. For $x \in \Sigma_i$ let $[x]$ denote its $\mathcal{R}$-equivalence class in Σ_i. Elicit a subjective partition as follows. Let $x_n \in \inf(X)$ and put $B(1) := \{x \in X : x \to_t x_n\} \cup [x_n]$. Set $X_i = X \backslash \cup_{j=1}^{i-1} B(j)$ and inductively define

$$B(i) := \{x \in X_i : x \to_t x_i\} \cup [x_i]$$

where $x_i \in \inf(X_i)$. Clearly, the sets $\{B(i)\}$ are disjoint and partition X. Choose any $x^i \in \inf(B(i))$ and let $B(x^i)$ denote the partition that contains x^i. Defining

$$U^P(A) := \max_i \{\min_{x\in A\cap B(x^i)} u(x)\}$$

we claim that $U^P(\cdot)$ represents $\succeq$. From the menu A, extract out the set of partition-dependent minima in A. That is, put $X^i(A) = \arg\min_{A\cap B(x^i)} u(x)$ and consider the set

$$A^* := \{X^0(A), \ldots, X^N(A)\}$$

By successive application of $A2^*$ and $A3b$ we know that $A \sim A^*$. Thus, it suffices to verify that $U^P(\cdot)$ represents $\succeq$ restricted to the set of menus of the form A^* (call such menus *non-reducible*). We claim that $A^* \sim \sup(A^*)$. Note that this observation yields the result since $U^P(\cdot)$ trivially represents the singleton ranking. As a first step, we verify that A^* is temptation-free. Let $x, y \in A^*$ with $x \in B(x^i), y \in B(x^j)$ where $i > j$. If $x \to_t y$, then $x \to_t x^j$ by Lemma 2. But, by construction, this means that $x \notin X(i) = X \backslash \cup_{j\le i-1} B(j)$ - contradiction. Thus, consider the possibility that $y \to_t x$. Note that we have $x^i \succeq x^j$ whenever $i > j$. If $x^i \succ x^j$, then (by Lemma 2 if $x \neq x^i$) $y \to_t x^i$ and $y \to_t x^j$. Thus, by $A3a$, $x^i \to_t x^j$ - contradiction. If $x^i \sim x^j$, then $x^i \notin [x^j]$ so that $\{x : x \to_t x^i\} \cap \{x : x \to_t x^j\} = \emptyset$. OTOH, we assumed $y \to_t x$, so that $y \to_t x^i$ (and $y \to_t x^j$) - contradiction. It follows that A^* is temptation-free. Iterative application of $A1'$ implies that $A^* \sim x$ for any $x \in \sup(A^*)$. □

Proof of Corollary 1. Let $(u(\cdot), \{B_i\})$ be a given partition-based representation. Denote the partition constructed in the proof of the Theorem by $\{B(x^i)\}$. That is, consider the underlying preference $\succeq$ represented by the given pair $(u(\cdot), \{B_i\})$ and construct the partitions $\{B(x^i)\}$ corresponding to this preference $\succeq$. Let $B_t(x) := \{y \in X : x \rightarrow_t y\}$ and note that for any $z \in B_i$ we must have, by representability, $B_t(z) \subseteq B_i$. Thus, take any $x \in B_{i'}$ and find x^i with $x \in B(x^i)$. Then, $x^i \in B_t(x) \subseteq B_{i'}$. Since $\{B_i\}$ partitions X it then follows that $B(x^i) \subseteq B_{i'}$. Thus, the partition $\{B_i\}$ is a coarsening of $\{B(x^i)\}$. If there are two distinct $B(x^i), B(x^j)$ with, say, $x^i \succ x^j$ contained in a common cell B_l, then by representability we obtain $x^i \rightarrow_t x^j$ - a contradiction. If $x^i \sim x^j$ and $B(x^i), B(x^j)$ lie in a common cell, then $x^i \in [x^j]$ - again, a contradiction. It follows that $\{B_i\} \equiv \{B(x^i)\}$. □

Proof of Corollary 2. Put $\mathcal{C}_x := B_t(x) \cup \{x\}$ and let $\mathcal{C}^* \equiv \{\mathcal{C}_x\}_{x \in X}$. Let $\Sigma_1(x) := \{\mathcal{C}' \in \mathcal{C}_1 : x \in \mathcal{C}'\}$. Note that, by representability, $\mathcal{C}_x \subseteq \mathcal{C}', \forall \mathcal{C}' \in \mathcal{C}_1$. We claim that there is some $\mathcal{C}' \in \mathcal{C}_1$ such that $\mathcal{C}_x = (-\infty, x] \cap \mathcal{C}'$. Otherwise, for each $\mathcal{C}' \in \Sigma_1(x)$ find some $z' \in \{(-\infty, x) \cap \mathcal{C}'\} \backslash \mathcal{C}_x$ and put $A' := \{z' : z' \in (-\infty, x) \cap \mathcal{C}' \backslash \mathcal{C}_x\} \cup \{x\}$ and note that $U^{\mathcal{C}_1}(A') < u(x)$ - a contradiction, since representability and $A1', A2^*$ imply that $A' \sim x$. Thus, for each $\mathcal{C}_{x_i}$ we can find some $\mathcal{C}_1(x_i) \in \mathcal{C}_1$ such that $\mathcal{C}_1(x_i) \cap (-\infty, x_i] = \mathcal{C}_{x_i}$. I first claim that $\mathcal{C}_1 \equiv \{\mathcal{C}_1(x_1), \ldots, \mathcal{C}_1(x_n)\}$. For each x_i we have $\mathcal{C}_{x_i} \subseteq \mathcal{C}_1(x_i)$. Let $\mathcal{C}_* \equiv \{\mathcal{C}_{x_i}\}$ and put $\mathcal{C}^1 \equiv \{\mathcal{C}_1(x_1), \ldots, \mathcal{C}_1(x_n)\}$. Observe that $U^{\mathcal{C}^1}(\cdot) = U^{\mathcal{C}_*}(\cdot)$. Since $U^{\mathcal{C}_*}(\cdot)$ represents $\succeq$ it follows that $U^{\mathcal{C}^1}(\cdot)$ represents $\succeq$, implying $\mathcal{C}^1 \equiv \mathcal{C}_1$ by sharpness. I now claim that the sets $\mathcal{C}_1(x_i)$ are in one-to-one correspondence with the undominated sets of the form $B_t(x_i) \cup \{x_i\}$, via $B_t(x_i) \cup \{x_i\} = \mathcal{C}_1(x_i) \cap (-\infty, x_i]$. Towards contradiction, assume there is a pair $\{x_i, x_j\}$ with $\mathcal{C}_1(x_i) = \mathcal{C}_1(x_j)$ (say $x_i \succ x_j$). Then, $B_t(x_i) \cup \{x_i\} = \mathcal{C}_1(x_i) \cap (-\infty, x_i]$ and

$$\begin{aligned}
B_t(x_j) \cup \{x_j\} &= \mathcal{C}_1(x_j) \cap (-\infty, x_j] \\
&= \mathcal{C}_1(x_i) \cap (-\infty, x_j] \\
&= (\mathcal{C}_1(x_i) \cap (-\infty, x_i]) \cap (-\infty, x_j] \\
&= (B_t(x_i) \cup \{x_i\}) \cap (-\infty, x_j]
\end{aligned}$$

This contradicts the assumption that $B_t(x_j) \cup x_j$ is undominated. It follows that the category $\mathcal{C}_1$ is a prolongation of the category $\mathcal{B}^*$ and similarly $\mathcal{C}_2$ is a prolongation of $\mathcal{B}^*$. It is straightforward to check that $(u, \mathcal{B}^*)$ is a sharp representation of $\succeq$. Representability is trivial given the sufficiency proof of Theorem 3, so that the only issue is sharpness. Towards contradiction, say that there is some undominated set $B_t(x_i) \cup \{x_i\}$ which is not relevant. Then, we must have $X = \cup_{j \neq i} (B_t(x_j) \cup \{x_j\})$. Since $B_t(x_i) \cup \{x_i\}$ is not dominated by any of these sets, for each $j \neq i$ with $x_i \in B_t(x_j) \cup \{x_j\}$ find $b_j \in (B_t(x_j) \cap (-\infty, x_i)) \backslash (B_t(x_i) \cup \{x_i\})$ and consider the menu $A = \{b_j\}_j \cup \{x_i\}$. Note that $A \sim x_i$, by $A1'$ and $A2^*$. OTOH, evaluating the function $U^{\mathcal{C}}(A)$, where $\mathcal{C} := \mathcal{B}^* \backslash (B_t(x_i) \cup \{x_i\})$ we get $U^{\mathcal{C}}(A) < u(x_i)$ - contradiction. □

Proof of Theorem 3. Necessity of NAT and Reduction is straightforward.

Necessity of *A3a*

Let $(u(\cdot), \mathcal{C})$ be a representing narrow category and say that $x \succ y \succ z$ with $x \to_t y$ and $x \to_t z$. Note that this implies that whenever $x \in \mathcal{C}_i$, then we must have $y \in \mathcal{C}_i$ and $z \in \mathcal{C}_i$. To show that $y \to_t z$ it suffices to show that $z \in \mathcal{C}_j$ whenever $y \in \mathcal{C}_j$. Thus, let $\mathcal{C}_j$ be a neighborhood that contains y. Let $\mathcal{C}_i$ contain x and note that we must have $y, z \in \mathcal{C}_i$. Since the category is narrow we must then have, $\{z' \in \mathcal{C}_i : y \succ z'\} = \{z' \in \mathcal{C}_j : y \succ z'\}$. Thus, $z \in \mathcal{C}_j$ so that $y \to_t z$.

Sufficiency

Let $\mathcal{C} \equiv \{B_t(x) \cup x\}$ denote the NAT category constructed in the proof of Theorem 2. We know that the corresponding model $(u, \mathcal{C})$ represents $\succeq$. I check that the category is, in fact, narrow. To see this, put $\mathcal{C}_x = B_t(x) \cup \{x\}$ and let $z \in \mathcal{C}_x \cap \mathcal{C}_y$. Either $\inf(x, y) \succ z$ or $z = \inf(x, y)$. In the former case (the other one is similar, hence we omit the proof), say that $x \succ z$ and consider the set $\{z' \in \mathcal{C}_x : z \succ z'\}$. Thus, for each such z' we have $x \succ z \succ z'$ and $x \to_t z, x \to_t z'$. By *A3a* we obtain $z \to_t z'$. Since $z \in \mathcal{C}_y$, Lemma 2 (when $z \neq y$) then implies that $y \to_t z'$, so that $z' \in \{z' \in \mathcal{C}_y : z \succ z'\}$. Similar reasoning shows the inclusion $\{z' \in \mathcal{C}_y : z \succ z'\} \subseteq \{z' \in \mathcal{C}_x : z \succ z'\}$. Thus, the NAT category is narrow. □

Proof of Corollary 3. Fix a sharp model $(u, \mathcal{C}_1)$, where $\mathcal{C}_1 \equiv \{\mathcal{C}_1^1, \ldots, \mathcal{C}_1^n\}$ and let $\succeq$ be the underlying menu preference. Put $z_i := \sup(\mathcal{C}_1^i)$ and note that, since the category is narrow, we have $\mathcal{C}_1^i = B_t(z_i) \cup \{z_i\}$ (choose the labeling such that $z_1 \succ z_2 \succ \ldots \succ z_n$). Let $A := \{z_1, \ldots, z_n\}$. I claim that A is a maximal, simple subset of X. First, check the simplicity claim. To this end, I claim that $z_i \notin \cup_{j \neq i} (B_t(z_j) \cup \{z_j\})$. If not, let us allege that $z_i \in B_t(z_1) \cup \{z_1\}$ (by the labeling we can only have $z_i \in B_t(z_j) \cup \{z_j\}$ for some $z_j \succ z_i$, so we are just picking z_1 wlog). Take any $y \in B_t(z_1)$ with $z_i \succ y$ and note that we have $z_1 \to_t z_i, z_1 \to_t y$. By *A3a*, this implies $z_i \to_t y$, so that $B_t(z_i) \cup \{z_i\} = (-\infty, z_i] \cap (B_t(z_1) \cup \{z_1\})$. That is, the set $B_t(z_i) \cup \{z_i\}$ is dominated by the set $B_t(z_1) \cup \{z_1\}$. This contradicts sharpness of the model $(u, \mathcal{C}_1)$.[16] It follows that we cannot have $z_i \to_t z_j$ for any pair (z_i, z_j) in the set A. By NAT and Reduction, this implies that A is simple. I now claim that A is a maximal simple set. To see this note that $X = \cup_i (B_t(z_i) \cup \{z_i\})$. Thus, if there is some $z \in X \backslash \{z_1, \ldots, z_n\}$ then $z_i \to_t z$ for some $z_i \in A$ - implying that $A \cup z$ cannot be simple. Let $(u, \mathcal{C}_2)$ be any other sharp category representation of $\succeq$. I show that the category $\mathcal{C}_1$ is a sub-category of $\mathcal{C}_2$. Sharpness of $\mathcal{C}_2$ then implies equality of the categories. Labeling $\mathcal{C}_1 \equiv \{\mathcal{C}_1^1, \ldots, \mathcal{C}_1^n\}$, I show this by inducting downwards on the sets $\mathcal{C}_1^i$. For notational brevity, from here on put $\mathcal{C}_1^i = \mathcal{C}(i)$ and denote generic elements of $\mathcal{C}_2$ as $\mathcal{C}_2(j)$. First check the claim for $i = 1$, via contradiction. Fix z_1 and say that for each $\mathcal{C}_2(j)$ with $z_1 \in \mathcal{C}_2(j)$ we have $\mathcal{C}(1) \neq \mathcal{C}_2(j)$. Then, for each j find $b_j \in \mathcal{C}_2(j) \backslash \mathcal{C}(1)$ and

[16] To see this, note that the category $\mathcal{C}_1' = \mathcal{C}_1 \backslash \mathcal{C}_1^i$ is complete. Moreover, take any menu A and consider the value of the function $U^{\mathcal{C}_1'}(A)$. Assume that the value of $U^{\mathcal{C}_1}(A)$ is attained on the set $\mathcal{C}_1^i$. Since $\mathcal{C}_1^1 (= (B_t(z_1) \cup \{z_1\}))$ dominates $\mathcal{C}_1^i$ it follows that the maximum is also attained on $\mathcal{C}_1^1$. Thus, $U^{\mathcal{C}_1'}(A) = U^{\mathcal{C}_1}(A)$.

consider the menu $B(1) := \{b_j\}_j \cup \{z_1\}$. Since $z_1 \succ b_j, \forall b_j$, representability implies that $z_1 \notin B(1)^*$ (since $U^{\mathcal{C}_2}(B(1)) < u(z_1)$). But this implies, by iterative application of Strong Reduction and No Attractions, that $z_1 \rightarrow_t b_j$ for some $b_j \in \mathcal{C}_2(j)\backslash\mathcal{C}(1)$ - contradiction (since $B_t(z_1) \subseteq \mathcal{C}(1)$). Now consider $\mathcal{C}(2)$. Apply the same argument. I claim that there is some $\mathcal{C}_2(j)$ with $\mathcal{C}(2) = \mathcal{C}_2(j)$. Otherwise, for each $\mathcal{C}_2(j)$ with $z_2 \in \mathcal{C}_2(j)$ we have $\mathcal{C}(2) \neq \mathcal{C}_2(j)$. For each such j select any $b_j \in \mathcal{C}_2(j)\backslash\mathcal{C}(2)$. Note that if $z_2 \succ b_j, \forall b_j$, then putting $B(2) := \{z_2\} \cup \{b_j\}_j$ we find $z_2 \succ B(2)$. Thus, $z_2 \rightarrow_t b_j$ for some b_j - contradiction. If $b_j \succ z_2$ for some b_j, then $b_j \in \mathcal{C}(1)$. Since the category $(u(\cdot), \mathcal{C}_2)$ is narrow this implies $\{x \in \mathcal{C}(1) : b_j \succ x\} = \{x \in \mathcal{C}_2(j) : b_j \succeq x\}$. As $z_2 \in \{x \in \mathcal{C}_2(j) : b_j \succ x\}$ we obtain $z_2 \in \mathcal{C}(1)$, implying that $z_1 \rightarrow_t z_2$ - contradiction. Inductively assume that we have shown that $\mathcal{C}(i) \in \mathcal{C}_2, \forall i \leq n-1$. Now claim that $\mathcal{C}(n) \in \mathcal{C}_2$. Otherwise, for each $\mathcal{C}_2(j)$ with $z_n \in \mathcal{C}_2(j)$ there is some $b_j \notin \mathcal{C}(n)$. Collect all these b_j's and consider the menu $B(n) := \{b_j\}_j \cup \{z_n\}$. If $z_n \succ b_j, \forall b_j$, then by representability we have $z_n \notin B(n)^*$. Thus, $z_n \rightarrow_t b_j$, for some b_j - contradiction. If there is some b_j with $b_j \succ z_n$, then $b_j \in \cup_{i=1}^{n-1} \mathcal{C}(i)$, say $b_j \in \mathcal{C}(i_n)$. By the induction hypothesis, $\mathcal{C}(i_n) \in \mathcal{C}_2$. Thus, since $(u(\cdot), \mathcal{C}_2)$ is narrow we have $\{x \in \mathcal{C}(i_n) : b_j \succ x\} = \{x \in \mathcal{C}_1(j) : b_j \succ x\}$. Since $z_n \in \{x \in \mathcal{C}_2(j) : b_j \succ x\}$ it follows that $z_n \in \mathcal{C}(i_n)$, implying that $z_{i_n} \rightarrow_t z_n$ - contradicting simplicity of A. It follows that for all i the sets $\mathcal{C}(i) \in \mathcal{C}_2$. By sharpness, it follows that $\mathcal{C}_1 = \mathcal{C}_2$. □

5.2 Omitted Proofs for Section 3

Proof of Theorem 4. This proof extends the argument given in the text to allow for ties in the singleton ranking. We will just indicate the where we need to make changes in the proof presented in the text. Fix a menu A and recall the sets $D^i_A(k)$. Recall the notation $I^A(x) = \cup\{z : z \succeq x, z \in I_{A'}(x)\}$ where the union is over all subsets $A' \subseteq A$ with $x \in \Sigma_w(A')$. Using these sets we recursively put (with $D^i_A(0) = I^i_A$)

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

When there are ties among singletons we amend this definition as follows. First, let $I^A_{\succ}(x) := \{x\} \cup \cup\{z : z \succ x, z \in I_{A'}(x)\}$ and put $\{x^i_1, x^i_2, \ldots, x^i_{n_i}\} = \inf(I^i_A)$. For each x^i_j recursively define (put $D^{(j,i)}_A(0) = \{x^i_j\} \cup \{z \in I_A(x^i_j) : z \succ x^i_j\}$)

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

As before the sequence of these sets is increasing, i.e. $D^{(j,i)}_A(k) \subseteq D^{(j,i)}_A(k+1)$ - hence, terminates at some $D^{(j,i)}_A$. Also note that $\cup_{(j,i)} D^{(j,i)}_A = A$, so that completeness holds. Representability follows as before, so that the only possible issues are coherence and downwards rigidity. We first verify coherence. Fix $A' \subseteq A$ and take $D^{(j,i)}_{A'}$. Note that, by construction, $\inf(D_{A'}^{(j,i)}) = x_j^i \in \inf(I_{A'}^i)$. 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 the text 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. Next, we verify downwards rigidity. Note that we can enumerate the sets in the local category

$$\mathcal{B}_A = \{D_A^{(j,i)}\}_{j \leq n_i, i \leq k}$$

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, $x_j^i \notin D_A^{(n,i)}$ for any other $n \neq j, n \leq n_i$. The only way downwards rigidity could fail is if $x_j^i \in D_A^{(q,l)}$ for some $l > i$, i.e. $x_j^i \succ x_q^l = \inf(D_A^{(q,l)})$. 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 $x_j^i \in D_A^{(q,l)}(n) \backslash D_A^{(q,l)}(n-1)$. 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 $A' \subseteq A$.[17] This contradicts the fact that $u(x_j^i) = u_{\min}(x_j^i, A)$. Hence, downwards rigidity also holds. □

Proof of Proposition 4. I only present the proof when there are no ties on singletons. As in the extension for the proof of Theorem 4, the argument is easily adapted (with more notation) to deal with ties. We will require the results of Claims 1 and 2 that were stated in course of proving Theorem 4. From the fact that $\Sigma_w(A) = \inf(I_A^1)$ we obtain $\inf(I_A^1) = \arg\min_{x \in A} L(x)$. Similarly, we have $\inf(I_A^i) = \arg\min_{x \in A^i} L(x)$ where we have set $A^i := A \backslash (\cup_{j \leq i-1} I_A^j)$. This implies that for a given $x \in I_A^j$, $L(x) \leq L(y), \forall y \in I_A^j, j > i$. Thus, $D_A^i \supseteq \{y \in A^i : y \succeq x^i\}$. To show the converse, we check by induction that $D_A^i(n) \subseteq \{y \in A^i : y \succeq x^i\}$. First, we check that $D_A^i(1) \subseteq \{y \in A^i : y \succeq x^i\}$: If $y \in I_{A'}(x)$ for some $x \in I_A^i$ (with $x \sim A' \subseteq A$), then $y \succeq x$ by Claim 1. Since $x \succeq x^i$ we obtain $y \succeq x^i$. We check that $y \in A^i$. Note that $L(y) \geq L(x)$ since $x \in \Sigma_w(A') = \arg\min_{z \in A'} L(z)$. Moreover, $L(x) \geq L(x^i)$ so that $L(y) \geq L(x^i)$. If $y \in I_A^k$ for some $k < i$, then $\{y, x^i\} \sim \{x^i\} \sim A_{x^i} \prec A_y$ - contradiction. Thus, $y \in I_A^j$, for some $j \geq i$ so that $y \in A^i$. For the inductive step assume that $D_A^i(n-1) \subseteq \{y \in A^i : y \succeq x^i\}$. The above argument then applies verbatim, so that we obtain $D_A^i(n) \subseteq \{y \in A^i : y \succeq x^i\}$ (here we are using the fact that $z \succeq x^i, \forall z \in D_A^i(n-1)$). Since the sets $D_A^i(n)$ eventually stabilize we obtain $D_A^i \subseteq \{y \in A^i : y \succeq x^i\}$, and hence equality. To check the representability claim note that $x^1 \succ x^2 \succ \cdots \succ x^k$ by Claim 2, so that $U^{\mathcal{B}_A}(A) = u(x^1)$. On the other hand, $\inf(I_A^1) = \arg\min_{x \in A} L(x)$ so that $u(x^1) = U^L(A)$. □

[17] This is the step where we require a different construction when there are ties. If we had used exactly the same construction of sets D_A^i as in the main text, then we could not claim $x_j^i \succ z$. It is possible for x_j^i and z to be "normatively" equivalent, but for z to be susceptible to compromise (where x_j^i is a no-compromise element of A). Thus, to pick up this distinction we need to use the modified construction we give here.

Proof of Lemma 3. Break into two cases, (i) $y \succ A$, and (ii) $y \preceq A$. In the former case, if $\{y\} \succ A \cup \{y\}$ then $(A \cup \{y\})^* \subseteq A$ so that $A \sim A \cup \{y\}$ by $A2$. In the latter case, if $y \sim A$ and $A \cup \{y\} \prec A$ then we again have $(A \cup \{y\})^* \subseteq A$ so that $\{y\} \sim A \sim A \cup \{y\}$ - contradiction, so that $y \sim A \sim A \cup \{y\}$ in this case. If $A \succ y$, then note that if $A \cup \{y\} \prec \{y\}$ we must have $(A \cup \{y\})^* \subseteq A$, which implies $A \sim A \cup \{y\}$ by $A2$ - a contradiction. Thus, $A \cup \{y\} \succeq \{y\}$. □

Proof of Lemma 4. Since $x \rightarrow_t y$ we have $x \notin (x \cup y \cup A(y))^*$, by Reduction (which is implied by $A1+A2$). Moreover, $y \rightarrow_t A(y)$ implies that $y \notin (x \cup y \cup A(y))^*$. Therefore, $(x \cup y \cup A(y))^* \subseteq A(y)$. By Strong Reduction, this implies that $x \cup y \cup A(y) \sim x \cup A(y) \sim A(y)$. In particular, $x \succ x \cup A(y)$. □

Remark: In the forthcoming proof we will commit a wild 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. That the notation denotes a node vs. an element of a node will be clear from the context of the argument. 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.

Proof of Corollary 4. The argument follows two steps, resp. labeled (i) embedding and (ii) pruning. The first step shows that every category model $(u, \mathcal{C})$ that is sharp is a prolongation of some sharp sub-category of the tree category - hence, it can be embedded in the tree category. The second step shows that there is a unique sharp sub-category of the tree category.

Step 1: Embedding.
For each $x \in X$ consider $\Sigma'(x) = \{\mathcal{C}' \in \mathcal{C}_1 : x \in \mathcal{C}', x \in \sup(\mathcal{C}')\}$. If $\Sigma'(x) \neq \emptyset$, I show that $\Sigma'(x)$ is a sub-collection of the terminal nodes in the x-tree associated with $\succeq$. With labeling determined by the x-tree construction, let $A_1(x), \ldots, A_k(x)$ be the attraction sets associated with x. I will reconstruct a set of paths through the x-tree whose associated set of terminal nodes is exactly $\Sigma'(x)$. Let Φ be the set of all paths in the x-tree and for each $\ell \in \Phi$, let $\ell^{-1}(x(1))$ denote the set of all terminal nodes which have $x(1)$ as a predecessor node.[18] 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 $x(1) \in A_1(x)$. We do not know yet that the sets in $\Sigma'(x)$ are 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 $x(1) \in A_1(x)$ we take $\ell^{-1}(x(1))$ to be the set of all elements (terminal nodes) in $\Sigma'(x)$ that contain the element $x(1)$. Similarly, for any $x_2(i) \in A_2(x)$ let $\ell^{-1}(x(1), x_2(i))$ denote all elements in $\Sigma'(x)$ which contain both $x(1)$ and $x_2(i)$. Note that, by representability,

[18]Note that, by the tree structure, there is a one-to-one correspondence between (complete) paths in the x-tree and terminal nodes of the x-tree.

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 non-empty valued correspondence $\gamma_1 : A_1(x) \rightrightarrows \Sigma'(x)$ 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 $\gamma_2 : A_1(x) \times A_2(x) \rightrightarrows \Sigma^{-1}(x)$ given by $(x_1(i), x_2(j)) \mapsto \ell^{-1}(x_1(i), x_2(j))$. Inductively construct correspondences, $\gamma_n : \prod_{j=1}^{n} A_j(x) \rightrightarrows \ell^{-1}(\ell(1,n))$. 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 $x_i(j) \notin x(k)$ 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.[19] Inductively proceeding we obtain a collection of paths in the x-tree.

Some observations about the nodes in $\Sigma'(x)$ that are the terminal ends of these paths. 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 length k_ℓ. Now we apply the concept of prolongation to show that every node in $\Sigma'(x)$ is in fact a terminal node of the x-tree. 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 ℓ_i) such that $x' \in \ell_i^{-1}(1,k), \forall k \leq k_i$ and for all paths ℓ_i. Also let $x(i)$ denote the terminal node of path ℓ_i.

We first make an observation about non-redundant sets $\mathcal{C}_i \in \mathcal{C}$, where the model $(u, \mathcal{C})$ is a sharp representation of the underlying preference. This means that for each set $\mathcal{C}_i \in \mathcal{C}$ one of the following must be true. Either (i) $\mathcal{C}_i \not\subseteq \cup_{j \neq i} \mathcal{C}_j$ or (ii) there is a menu A for which $\arg\max_{\mathcal{C}_j \in \mathcal{C}} \min_{z \in \mathcal{C}_j \cap A} u(z) = \mathcal{C}_i$. Using this observation, I claim that there must be a unique ℓ_i with $x(i) \in x'$. Towards contradiction, assume there are two nodes $x(i), x(j)$ with $x(i), x(j) \subseteq x'$. Applying the preceding observation about sharpness consider two cases, (i) $\Phi(x') := \{A : \arg\max_{x \in \mathcal{C}} \min_{z \in x \cap A} u(z) = x'\} \neq \emptyset$, (ii) $x' \not\subseteq \cup_{x \in \mathcal{C}} x$. Note that the two cases are not mutually exclusive. Consider first the setting where only (i) holds. Then, for each $A \in \Phi(x')$ we know that the (strict) maximum value $U^{\mathcal{C}}(A)$ is attained on the set $x' \cap A$. Let $\Phi_i(x')$ denote the menus for which the maximum is attained on $x(i)$ and similarly denote $\Phi_j(x')$. If both $\Phi_i(x')$ and $\Phi_j(x')$ are non-empty, then find menus $A_1 \in \Phi_i(x'), A_2 \in \Phi_j(x')$ such that $A_1 \succeq A_2$. If $A_1 \succ A_2$, then we obtain on the one hand $U^{\mathcal{C}}(A_1) = u(x_{A_1})$. OTOH, consider the menu $A_1 \cup x_{A_2}$. Note that we still have $U^{\mathcal{C}*}(A_1 \cup x_{A_2}) = u(x_{A_1})$ (since the tree category represents $\succeq$). By sharpness, we know that the value of $U^{\mathcal{C}}(A_1)$ on the set

[19] Here I make an even more egregious abuse of notation and take $x_i(j)$ to be an element, whereas $x(k)$ is a node.

$x' \cap A_1$ is a strict maximum. However, representability implies that there must be a set $x^*(\neq x') \in \mathcal{C}$ such that $\min_{z\in x^*\cap(A_1\cup x_{A_2})} u(z) = u(x_{A_1})$. This contradicts the fact that the strict maximum of the category function occurs on the set $x' \cap A_1$.

This completes the argument for the setting where only (i) holds. Now assume that there is some element $z_i \in x(i)$ with $z_i \notin \cup_{x\neq x':x\in\mathcal{C}} x$. Consider again the sets $\Phi_i(x')$. If $\Phi_j(x') \neq \emptyset$, then take $A \in \Phi_j(x')$ and note that $x_A \neq z_i$. If $x_A \succ z_i$, then apply the preceding argument verbatim to deduce that only one of the nodes $x(i)$ or $x(j)$ can be contained in x'. If $z_i \succ x_A$ consider the menu $A = \{z_i, x_A\}$. Consider two sub-cases, (i) $x_A \in x(i)$ and (ii) $x_A \notin x(i)$. In the latter case, representability implies that $U^{\mathcal{C}}(A) = u(z_i)$. OTOH, since $z_i \notin \cup_{x\neq x':x\in\mathcal{C}} x$ we must have $U^{\mathcal{C}}(A) = u(x_A)$ - contradiction. Note that this happens if there is even a single $A \in \Phi_j(x')$ such that (ii) holds. We are left with the case where $x_A \in x(i)$ whenever $\Phi_j(x') \neq \emptyset$. Note that this implies, in particular, that $x(j) \subseteq \cup_{x\neq x(j):x\in\mathcal{C}} x$. Now apply the prolongation reduction - replace the category $\mathcal{C}$ with the category $\mathcal{C}' := \mathcal{C}\backslash x' \cup x(i)$. The model $(u, \mathcal{C})$ is a prolongation of $(u, \mathcal{C}')$ (call $\mathcal{C}'$ a retract of $\mathcal{C}$), where both models are sharp. Applying this argument to each node $x' \in \Sigma'(x)$ we obtain, by successively retracting if necessary, that each node in the set $\Sigma'(x)$ contains only a single terminal node in the x-tree.

To complete the embedding argument, we verify that, by retracting again if necessary, for each $x' \in \Sigma'(x)$ there are no elements outside of the x-tree. Again via contradiction, say that there is some $z \in x'$ where the z-tree is not embedded in the x-tree. Apply sharpness and consider the case where $\Phi(x') \neq \emptyset$. Let A be a menu such that the strict maximum of $U^{\mathcal{C}}(A)$ is attained on $x' \cap A$ and let x_A be the maximizer. If $x_A \succ z$, then consider the menu $A \cup z$. Since the z-tree is not embedded in the x-tree we have $A \cup z \sim x_A$ - so that $U^{\mathcal{C}}(A \cup x_A) = u(x_A)$. This implies there must be some set $x'' \in \mathcal{C}$ such that $\min_{z\in x''\cap A} u(z) = u(x_A)$. Since $x'' \neq x'$, this then contradicts the fact that $A \in \Phi(x')$. It follows that if there is a $z \in x'$, where the z-tree is not embedded in the x-tree, then we must have $\Phi(x') = \emptyset$, so that there is some $z_* \in x'$ with $z_* \notin \cup_{x\neq x':x\in\mathcal{C}} x$. Choose a $\succeq$-maximal z w.r.t the property that (i) $z \in x'$ and (ii) the z-tree is not embedded in the x-tree. Note that if $z_* \succ z$, maximality (and strictness of the singleton ranking) then implies that the z_*-tree is embedded in the x-tree. Since z_* is not in any other set $x'' \in \mathcal{C}$, representability implies that $z_* \to_t z$, so that the z-tree is embedded in the x-tree. Thus, consider $z \succ z_*$. Since $\Phi(x') = \emptyset$, consider the subset of x' defined by $x'_* := \{y \in x' : z \succeq y\}$ and note that the retracted category given by $\mathcal{C}' := \mathcal{C}\backslash x' \cup x'_*$ is also a sharp representation of $\succeq$. Thus, we may iteratively retract the original category $(u, \mathcal{C})$ to obtain a sharp sub-category of the tree category that also represents $\succeq$.

Step 2: Pruning.
We show that there is a unique sharp sub-category of the tree category. First, we produce a sharp sub-category of the tree category. Second, we prove that this is the

only sharp sub-category. To get a hint as to where to look for a sharp sub-category, consider the example we gave preceding the proof of the corollary. Note that there was only one x-tree required for a sharp representation, and moreover, the only relevant nodes in the x-tree were the *minimal* (w.r.t. set inclusion) terminal nodes. This provides a clue as to what the unique sharp sub-category should be. For each tree $\mathcal{T}_x$ let $\mathcal{C}_x^*$ denote the collection of all minimal terminal nodes. Consider the sub-category of all minimal terminal nodes across all x-trees, i.e. put $\mathcal{C} := \{\mathcal{C}_x^*\}_{x \in X}$. It is straightforward to check that the model $(u, \mathcal{C})$ represents $\succeq$. By deleting nodes, if necessary, pass to a sharp sub-model of this category representation, and abusing notation, denote the pair with the same notation $(u, \mathcal{C})$. This is a sharp sub-category of the tree category that represents $\succeq$. We now check there is no other.

Start with the $\succeq$-maximal x-tree. Take $(u, \mathcal{C}_1)$ to be any sharp sub-category. Let $(u, \mathcal{C})$ be the sharp category constructed in the preceding paragraph. First, observe that for any $\succeq$-maximal x-tree *all* minimal terminal nodes are relevant for the sub-category of all minimal nodes, and hence present in the sharp sub-category $(u, \mathcal{C})$. The argument for this is useful and is implicitly invoked many times in the remainder of this proof. Towards contradiction, assume there is a minimal terminal node, say x_* that is not in the sub-category $(u, \mathcal{C})$. Let $\{x_1, \ldots, x_N\}$ be an enumeration of the nodes in the x-tree with $x_1 \equiv x_*$. For each $x_i (\neq x_1)$ pick some element $z_i \in x_i \backslash x_1$ (by minimality of x_*). Consider the menu $A' := \{z_i : z_i \in x_i\} \cup \{x\}$. Since the x-tree is unobstructed at A' we must have $A' \sim x$. OTOH, consider the value of the function $U^{\mathcal{C}}(A')$. Note that the only nodes in the category that contain x are the terminal nodes from the x-tree (by $\succeq$-maximality). Since $x_* \notin \mathcal{C}$ and $x \succ z_i \in x_i$, every node in $\mathcal{C}$ that contains the singleton x is obstructed by some z_i. It follows that $U^{\mathcal{C}}(A') < u(x)$ - contradicting representability. Thus, all minimal terminal nodes for a $\succeq$-maximal x-tree are present in the model $(u, \mathcal{C})$.

Similar reasoning shows that all minimal terminal nodes are present for any non-embedded x-tree (with the obvious notion of embedding). 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. I first check that the category $(u, \mathcal{C}_1)$ contains all nodes from $(u, \mathcal{C})$ that come from trees in group I. Then, I induct downwards on $\succeq$-rank to show that all nodes in $(u, \mathcal{C})$ that come from group II trees are also in $(u, \mathcal{C}_1)$. Sharpness of $(u, \mathcal{C}_1)$ then implies the equality of categories. First consider trees in group I. Take any x-tree in group I and recall that the previous paragraph showed that all minimal terminal nodes were present in the sharp model $(u, \mathcal{C})$ constructed above. Towards contradiction, say that x_* is a node in $(u, \mathcal{C})$ that is not in $(u, \mathcal{C}_1)$. Let $\{x_1, \ldots, x_n\}$ be an enumeration of the (terminal) nodes in the x-tree and note that, by minimality of x_*, for each x_i there is a $z_i \in x_i \backslash x_*$. Put $A' = \cup_i z_i \cup \{x\}$ and note that $A' \sim x$, whereas $U^{\mathcal{C}_1}(A') < u(x)$ - contradiction. Thus, for all group I trees the nodes in $(u, \mathcal{C})$ are contained in $(u, \mathcal{C}_1)$.

I now consider the terminal nodes in group II trees. I claim that all minimal nodes in the category $\mathcal{C}_1$ (for any group II tree) are present in $\mathcal{C}$ and, moreover, all nodes in $\mathcal{C}$ are present in $\mathcal{C}_1$. For this we apply sharpness of both categories. Induct downwards on the $\succeq$-rank of the x-trees in group II. Begin with a $\succeq$-maximal embedded x-tree (note that there may be more than one) and consider the terminal minimal nodes. First check that all nodes in $\mathcal{C}$ are contained in $\mathcal{C}_1$. Let $x^1 \in \mathcal{C}$ be a node in the sharp sub-category consisting of minimal nodes and assume, towards contradiction, that $x^1 \notin \mathcal{C}_1$. By sharpness, either (i) there is some $z_1 \in x^1$ such that $z_1 \notin \cup_i x^i, \forall x^i (\neq x^1) \in \mathcal{C}$ or (ii) there is a menu A such that the strict maximum of $U^{\mathcal{C}}(A)$ is obtained on the node x^1. Consider the latter case first and let x_A denote the compromise in the menu A (where $x_A \in x^1$). Take any x_i such that the tree $\mathcal{T}_x$ is embedded in $\mathcal{T}_{x_i}$, where the x_i-tree is non-embedded. Let $\kappa_i(x^1) = \{y_i(j) \in \mathcal{C}_1 : x^1 \subseteq y_i(j)\}$ denote the set of all terminal nodes in the x_i-tree that contain the node x^1. Note that for each $y_i(j) \in \kappa_i(x^1)$ one of the following is true: Either (i) the node is minimal or (ii) the node is non-minimal and there is some $z_i \in y_i(j)$ with $x \succ z_i$.

Note that there is a third possibility we are omitting: the node is non-minimal and the elements of the node x^1 form a lower bound order interval for the node $y_i(j)$.[20] I claim this cannot happen if $(u, \mathcal{C}_1)$ is a minimal, sharp sub-category. Else, by sharpness consider a menu A such that $U^{\mathcal{C}_1}(A)$ attains a strict maximum on $y_i(j)$. If $x_A = x_i$, then the maximum is also attained in some minimal node of the x_i-tree contained in the node $y_i(j)$. By the argument for group I trees, this node must be in $\mathcal{C}_1$ - which contradicts sharpness. Thus, $x_i \succ x_A$. If $x_A \succ x$, then by $\succeq$-maximality of x among group II trees, we know that the x-tree is not embedded in the x_A tree. Thus, we must have $A \cup x \sim x_A$. OTOH, the strict maximum of the function $U^{\mathcal{C}_1}(A)$ occurs on $y_i(j)$, which implies that $U^{\mathcal{C}_1}(A \cup x) < u(x_A)$ - contradicting representability. It follows that $x_A \in x^1$. This then implies we can replace the node $y_i(j)$ with the node x^1 and take the retracted category, $\mathcal{C}' := (\mathcal{C}_1 \backslash y_i(j)) \cup x^1$. The model $(u, \mathcal{C}')$ is also a sharp representation of $\succeq$ - contradicting minimality of $\mathcal{C}_1$. It follows that the third possibility cannot occur. Now if $y_i(j)$ is minimal (the first possibility), then by sharpness of $(u, \mathcal{C})$ and the fact that all minimal nodes from group I trees are present in $\mathcal{C}$, there is some $z_i \in y_i(j)$ with $x \succ z_i$. Similarly, find such a z_i in case (ii). Thus, for each non-embedded x_i-tree in which the x-tree is embedded, find a set

$$A_i = \{z_i : x \succ z_i, z_i \in y_i(j), y_i(j) \in \kappa_i(x^1)\}$$

Finally, since x^1 is a minimal node of the x-tree, for any other terminal node x^i in the tree find some $y_i \in x^i \backslash x^1$ and put $A' := \{x\} \cup_i y_i$. Consider the menu $\hat{A} := A' \cup_i A_i$ and note that the x-tree is unobstructed at $\hat{A}$, so that $\hat{A} \sim x$. OTOH, computing the value of $U^{\mathcal{C}_1}(\hat{A})$ note that the embedded image of x-trees in any x_i-tree is obstructed

[20] I say that a set A is a *lower bound order interval* for a set B if $A \subseteq B$ and there is some $x \in B \cap A$ such that $(-\infty, x] \cap B = A$.

by $\hat{A}$ (for the sub-category $\mathcal{C}_1$) and, by construction, the nodes in $\mathcal{C}_1$ from the x-tree are also obstructed by $\hat{A}$. It follows that $U^{\mathcal{C}_1}(\hat{A}) < u(x)$ - contradicting representability. Thus, all nodes in $\mathcal{C}$ from the x-tree are present in $\mathcal{C}_1$.

For the reverse inclusion, take a minimal node $x^1 \in \mathcal{C}_1$ and proceed as above. If $x^1 \notin \mathcal{C}$, then find a menu $\hat{A}$ as above at which all $\mathcal{C}$ nodes in the x-tree and its embedded images are obstructed. Note that since $\mathcal{C}$ contains only minimal nodes, by construction, finding this menu $\hat{A}$ is much easier (and uses only (i) from the above argument) than the direction showing the inclusion of $\mathcal{C} \subseteq \mathcal{C}_1$ (for x-tree nodes). The same argument as above then gives $\hat{A} \sim x$ (since $U^{\mathcal{C}_1}(\hat{A}) = u(x)$; however, $U^{\mathcal{C}}(A) < u(x)$ - contradiction. Thus, the minimal nodes of $\mathcal{C}_1$ and $\mathcal{C}$ agree on the x-tree for any $\succeq$-maximal group II tree. List the group II trees as $\{x_1 \succeq x_2 \cdots \succeq x_n\}$ and inductively assume we have shown the equality of minimal nodes for all x_i-trees where $1 \leq i \leq k$. Consider the x_{k+1}-tree. As above, first check the harder inclusion of $\mathcal{C} \subseteq \mathcal{C}_1$. Let $x^1 \in \mathcal{C} \backslash \mathcal{C}_1$ as before, and proceed via contradiction. Define the sets of nodes $\kappa_i(x^1)$ as before and note that any $y_i(j)$ falls into three cases, (i) $y_i(j)$ is minimal, (ii) $y_i(j)$ is non-minimal but contains some z_i with $x \succ z_i$, and (iii) $y_i(j)$ is non-minimal and elements of the node x^1 form the lower bound order interval on $y_i(j)$. I check that case (iii) is empty. Fix an x_i-tree that contains the x-tree and take a node $y_i(k) \in \kappa_i(x^1)$ that allegedly falls into case (iii).

Consider a menu A for which $U^{\mathcal{C}_1}(A)$ attains a strict maximum on the node $y_i(j)$ and consider $x_i \succ x_A \succ x$. Take the $\succeq$-maximal x_A with the property that the strict maximum of $U^{\mathcal{C}_1}(A)$ is obtained on the node $y_j(k)$. Note that this implies any lower ranked singleton is embedded in the x_A-tree. I claim that $(-\infty, x_A] \cap y_j(k)$ is a terminal node in the x_A-tree. This follows by an identical embedding argument as in step 1, hence we omit the details. Let y_{x_A} denote the corresponding node of the x_A-tree and note that we then obtain a proper retraction of the category $\mathcal{C}$ - contradicting minimality. It follows that case (iii) is empty. Therefore, for every $y_i(j)$ that contains the node x^1, we have either (i) $y_i(j)$ is minimal or (ii) there is some $x \succ z_i \in y_i(j)$. In the former case, apply the induction hypothesis (to the x_i-tree) to claim that the node $y_i(j) \in \mathcal{C} \cap \mathcal{C}_1$. Thus, by sharpness of x^1 (relative to $\mathcal{C}$) there is some $z_i \in y_i(j)$ with $x_A \succ z_i$ (where, as above, x_A is the compromise for some menu A at which the function $U^{\mathcal{C}}(A)$ attains a strict maximum at x^1). Assemble sets A_i and A' as before and put $\hat{A} = \cup_i A_i \cup A'$. Note that the x-tree is unobstructed at $\hat{A}$ - so that $\hat{A} \sim x$. OTOH, the nodes of the category $\mathcal{C}_1$ are obstructed, so that $U^{\mathcal{C}_1}(\hat{A}) < u(x)$ - contradiction. The reverse inclusion similarly follows. Thus, for all x-trees we have (i) all minimal nodes in $\mathcal{C}_1$ are contained in $\mathcal{C}$, and (ii) all nodes (which are minimal by construction) in $\mathcal{C}$ are in $\mathcal{C}_1$. In particular, we have $\mathcal{C} \subseteq \mathcal{C}_1$. By sharpness, $\mathcal{C}_1 \equiv \mathcal{C}$. □

References

[1] Chandrasekher, M. (2009): "A Theory of Local Menu Preferences," Mimeo, Arizona State University.

[2] Cherepanov, V., T. Feddersen, and A. Sandroni (2008): "Rationalization," Mimeo.

[3] de Clippel, G. and K. Eliaz (2010): "Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects," Mimeo.

[4] Dekel, E., B. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, vol. 69, 891 - 934.

[5] Dekel, E., B. Lipman, and A. Rustichini (2009): "Temptation-Driven Preferences," *Review of Economic Studies*, forthcoming.

[6] Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, vol. 69, 1403 - 1435.

[7] Gul. F. and W. Pesendorfer (2005): "The Revealed Preference Theory of Changing Tastes," *Review of Economic Studies*, vol. 72, 429 - 448.

[8] Manzini, P. and M. Mariotti (2010): "Categorize then Choose: Boundedly Rational Choice and Welfare," Mimeo.

[9] Masatlioglu, Y., D. Nakajima, and E. Ozbay (2009): "Revealed Attention," Mimeo.

[10] Ok, E., P. Ortoleva, and G. Riella (2007): "Revealed (P)Reference Theory," Mimeo.