

Categorize Then Choose: Boundedly Rational Choice and Welfare

Paola Manzini

Marco Mariotti*

University of St Andrews and IZA

University of St Andrews

This version: September 2010

Abstract

We propose a boundedly rational model of choice where agents categorize alternatives before choosing. The model explains some behavioral anomalies, and it is fully characterised by a property of choice data: a categorizer can never exhibit certain patterns of ‘revealed preference reversals’. This model offers clues on the problem of making welfare judgements in the presence of boundedly rational agents.

*School of Economics & Finance, Castlecliffe, The Scores, University of St Andrews KY16 9AL, UK. Email: paola.manzini@st-andrew.ac.uk, marco.mariotti@st-andrews.ac.uk. This paper supersedes our earlier paper “Two-stage boundedly rational choice procedures: Theory and experimental evidence”. The comments by a set of extraordinarily thorough referees, and discussions with Douglas Bernheim and Tim Feddersen have led to a substantial improvement on earlier versions: special thanks to them. For insightful discussions and comments we are grateful to Gianni de Fraja, John Hey, Nagore Iriberry, Jonathan Leland, Michele Lombardi, Luigi Mittone, Daniel Read, Ludovic Renou, Karl Schlag, Nick Vriend, Quan Wen and seminar audiences at the Universities of Amsterdam, Brescia, 2009 CRETA Workshop in Warwick, EUI, Helsinki, Lausanne, Leicester, LUISS, Malaga, Manchester, Padova, St. Andrews, Siena, Verona, Zurich and at the ESA 2007, APET 2007, FUR 2008, Games 2008 conferences. Part of this work was carried out while visiting the University of Trento, supported by ESRC grant RES-000-22-0866. We wish to thank both institutions for their support. Any error is our own.

1 Introduction

Categorization is a central cognitive mechanism: it is a natural simplifying operation to perform when making choices in complex situations. Suppose you find yourself in a new town, and the hotel provides you with a long list of restaurants to choose from for dinner. You don't compare each restaurant with every other restaurant. Rather you compare 'coarse' objects (categories) such as areas (City of London, West End,...) or cuisine types (Italian, Mexican...). You decide to focus on Italian cuisine and you want to be in East London, and then pick (what you believe to be) the best Italian in East London.

Observe that this is perfectly compatible with your picking a Mexican restaurant over an Italian one, were you to focus exclusively on just these two, sitting at the two sides of the hotel, and having the time to discover that the Mexican menu is in fact more attractive than the Italian one.

We study a general choice procedure that uses categorization in this way as a core component. In the textbook description, choice behavior results from the maximization of a binary preference relation (possibly summarized by a utility function). In our model the agent's decision process takes place instead in two stages. The first stage involves a coarse form of maximization, using a binary relation (interpreted as a psychological *shading relation*) defined on categories, namely *sets* of alternatives. In the previous example, {Italian restaurants} shades {Mexican restaurants} and {restaurants in East London} shades {restaurants in the West End}. Then in the second stage the agent picks an alternative which is preferred to all surviving alternatives. So in both stages the agent does maximize, but only the latter stage involves the maximization of binary relation

on alternatives, interpreted as a preference. The agent never needs to make detailed binary comparisons between alternatives within a category and across categories, and herein lies the simplifying power of categorization.

We call this enriched model of preference maximization Categorize Then Choose, or CTC in short. We argue that the CTC model captures in a natural way many choice situations. When looking for a mortgage plan and trying to find your way in a maze of product offerings, you first focus on a category of products (say, fixed rate mortgages), and then pick the best in that category. When purchasing an airline ticket online you focus solely on flights that are at convenient times, or on cheap flights. Also, while we do not explicitly address strategic behaviour, categorization is also likely to play an important role in situations of strategic interaction: in many relevant games the number of strategies is astronomical, and it is simply impossible for players to consider all strategies one by one. As argued by Arad and Rubinstein [2], in such circumstances players will group strategies into categories.¹

The CTC model is a possible way of capturing some empirical evidence that contradicts the textbook model. Let us identify ‘behavior that is incompatible with the maximisation of a preference ordering’ with ‘behavior that violates Samuelson’s Weak Axiom of Revealed Preference (WARP)’.² A difficulty commonly associated with modelling bounded rationality is that while there is only one way to be rational, there are many ways to be irrational. To place some structure on violations of rationality, our paper begins with a classification of WARP

¹Our thesis is that the large size of the strategy space and its structure force a player to consider the features of a strategy rather than concrete strategies’ (Arad and Rubinstein [2], p. 4). They however go on to study strategic decision procedures which are not related to CTC.

²When choices are single-valued as in this paper, WARP states that if an alternative x is chosen when y is available, then it can never be the case that alternative y is selected when x is available.

violations which is purely behavioural (rather than psychological). If you pick option A over option B, option B over option C and option C over option A, you exhibit a *pairwise cycle of choice*. If you pick option A over both option B and option C in binary comparisons, but you do not pick option A when choosing between A, B and C, then you exhibit an elementary form of *menu dependence*. We note that a choice behavior is not consistent with WARP if and only if it either exhibits pairwise cycles or menu dependence (or both). This is a useful taxonomy of ‘irrationality’, because it zeroes in on two very different behavioral aspects: one involving only binary comparisons, and one involving the ability to use binary comparisons to make higher order choices from larger sets.³

The CTC model can explain both types of WARP violations. The appealing feature is that it does so using just two binary relations. Its reach is wider than the more complex model of ‘sequentially rationalizable choice’ we proposed in Manzini and Mariotti [22], which (despite the use of an arbitrarily large number of binary relations and of decision stages) cannot accommodate menu effects.⁴ We have become convinced that menu dependence can be pervasive, both by an experiment we ourselves conducted ([23]) and by scrutiny of other evidence and introspectively compelling examples. For instance, the following pattern has been observed in doctors’s choices (Redelmeier and Shafir [35]): prescribe medication A when the only alternative option is no treatment, and prescribe

³See e.g. Roelofsma and Read [36], Tversky [48], and Waite [50] who find evidence of pairwise cycles of choice. Menu effects are widely discussed in the marketing and consumer research literature in several guises (e.g. ‘attraction effects’ and ‘compromise effects’), as well as in economics: see e.g. Masatlioglu and Nakajima [27], Masatlioglu, Nakajima and Ozbay [28], Eliaz and Spiegler [10]. The evidence we present in [23] and the later discussion in this paper points to further instances of menu effects.

⁴Apestegua and Ballester [1] offer very general extensions and characterizations of the idea of sequential rationalizability. Even these extensions however cannot accommodate menu effects in the sense meant here.

medication B when the only alternative option is again no treatment, but opt for no treatment when both medications are available. This is a clean case of menu dependence, obtained in addition in a small choice set of only three alternatives.

Despite its wide explanatory power, the CTC model does place restrictions on choice data. The process of categorization has a precise behavioral counterpart: interestingly, a categorizer can never exhibit certain patterns of ‘revealed preference reversals’. Formally, this is captured by a weaker property than WARP, WWARP in short, introduced in Manzini and Mariotti [22]), which fully characterizes our model in terms of observable choice data.⁵ From an empirical viewpoint, this offers a direct, simple nonparametric test of our model. We demonstrate the effectiveness of this test in Manzini and Mariotti [23] using experimental data.⁶

WARP violations are problematic not only at the descriptive level, but also at the normative one: if there is no utility, how can an external observer make welfare judgements on the basis of choice data? How to perform welfare judgements in the presence of boundedly rational decision makers is an ongoing puzzle. Recent important attacks to this problem (Bernheim and Rangel [4], Green and Hojman [12]) have both proposed methods that allow an external observer to make partial welfare rankings on the basis of standard observed choice data. However, they do so on the basis of a framework that while normatively meaningful, does not exclude any conceivable choice behavior and is therefore not falsifiable. In section 7 we discuss some implications of our falsifiable model for

⁵WWARP adds to WARP the clauses between brackets in the following definition: if x is chosen when y is available [both in pairwise contests and when a ‘menu’ of other alternatives is also available], then y cannot be chosen when x is available [when a smaller menu of alternatives is available].

⁶The CTC model performs very well in that particular experiment.

welfare analysis.⁷

2 Preliminaries: Rationality, Pairwise consistency and Menu effects.

Let X be a set of alternatives. The situations in which the decision maker may find himself are described by the domain of choice, a collection of nonempty subsets (choice sets) $\Sigma \subset 2^X \setminus \emptyset$. A choice function on Σ is a function $\gamma : \Sigma \rightarrow X$ such that $\gamma(S) \in S$ for all $S \in \Sigma$, which describes the decision maker's behavior, namely his selection from each choice set in the domain. The only additional assumption on the domain Σ is that all pairs of alternatives and all triples are included in the domain, that is: for all distinct $x, y, z \in X$, $\{x, y\} \in \Sigma$ and $\{x, y, z\} \in \Sigma$.

For a binary relation $\succ \in X \times X$ denote the \succ -maximal elements of a set $S \in \Sigma$ by $\max(S, \succ)$, that is:

$$\max(S, \succ) = \{x \in S \mid \nexists y \in S \text{ for which } y \succ x\}$$

Definition 1 *A choice function is fully rational if there exists a complete order \succ on X such that $\{\gamma(S)\} = \max(S, \succ)$ for all $S \in \Sigma$.*

As is well-known,⁸ in the present context the fully rational choice functions are exactly those that satisfy Samuelson's [42] Weak Axiom of Revealed Preference (WARP):

⁷Another recent important contribution by Chambers and Hayashi [6] also proposes a theory of welfare analysis for (possibly) boundedly rational decision makers. However, the nature of that model is rather different from our and the other mentioned contributions, in that the primitive is a stochastic choice function.

⁸See e.g. Moulin [31] or Suzumura [47]

WARP: If $x = \gamma(S)$, $y \in S$ and $x \in T$ then $y \neq \gamma(T)$.

WARP says that if steamed salmon is directly revealed preferred to (i.e. chosen in the presence of) fried chicken, then fried chicken can never be directly revealed preferred to steamed salmon (i.e. revealed preference is an asymmetric relation).

Failures of WARP may mix together more than one elementary form of inconsistency. To reduce the lack of full rationality to its building blocks, we consider two conceptually distinct and very basic violations of WARP, menu dependence and pairwise inconsistency. The latter category involves exclusively choices between *pairs* of alternatives. The former category involves failures of aggregation, as it were, in choices from larger sets.

The following property captures the elementary form of *menu independence*:

Condorcet consistency: If $x = \gamma(\{x, z\})$ for all $z \in S \setminus \{x\}$ for some $S \in \Sigma$, then $x = \gamma(S)$.

Condorcet consistency says that if steamed salmon is chosen in pairwise contests against any other item in a menu, then steamed salmon will be chosen from the entire menu.

Let \succ_γ denote the *base relation* of a choice function γ , that is $x \succ_\gamma y$ if and only if $x = \gamma(\{x, y\})$. A set $\{x_1, \dots, x_n\}$ is a *base cycle* of γ if $x_i \succ_\gamma x_{i+1}$ for all $i = 1, \dots, n - 1$ and $x_1 = x_n$.

Pairwise consistency: There is no base cycle of γ .

Our classification result decomposes WARP into Condorcet and Pairwise

consistency:⁹

Proposition 1 *Let X be finite. Then a choice function on Σ satisfies WARP if and only if satisfies both Condorcet consistency and Pairwise consistency.*

Proof: It is obvious that a choice function that violates Condorcet consistency also violates WARP. Suppose it violates pairwise consistency. Then, since each pair of alternatives is in Σ by assumption, any \succ_γ -cycle includes a \succ_γ -cycle involving only three alternatives. That is, there exist $x, y, z \in X$ for which $x \succ_\gamma y$, $y \succ_\gamma z$, and $z \succ_\gamma x$, so that WARP is contradicted (since $\{x, y, z\} \in \Sigma$ by assumption).

For the converse implication, suppose that γ violates WARP, and let in particular $S, T \in \Sigma$ be such that $x = \gamma(S) \neq y = \gamma(T)$, $x, y \in T \cap S$. Suppose that γ satisfies pairwise consistency: we show that then γ must violate Condorcet consistency. By pairwise consistency there exist \succ_γ -maximal elements in S and T (recall that X is finite). Since \succ_γ is asymmetric and complete (all pairs are in the domain), these elements are unique. So there are unique and distinct $s \in S$ and $t \in T$ such that

$$s = \gamma(\{s, z\}) \text{ for all } z \in S \setminus \{s\}$$

$$t = \gamma(\{t, z\}) \text{ for all } z \in T \setminus \{t\}$$

If $s \neq x$, then Condorcet consistency is violated on S . If $s = x$, then in particular $x = \gamma(\{x, y\})$. So $y \neq t$ and Condorcet consistency is violated on T .¹⁰ ■

⁹In Manzini and Mariotti [22] we prove an analogous but weaker result on a less general domain (the inductive method of proof used there would not work on the present domain).

¹⁰Note that this method of proof would yield a similar conclusion also on infinite domains,

In Manzini and Mariotti [22] we focussed on explaining pairwise inconsistency. We submit now that a successful theory must be capable of explaining both types of violations of full rationality. This assertion is supported in particular by an experiment (reported in Manzini and Mariotti [23]), in which we demonstrate how diffuse phenomena of menu dependence are. In that specific choice experiment, it is even the case that the bulk of WARP violations is driven by violations of the seemingly mild Condorcet consistency requirement.

In addition, other experimental evidence describes contexts in which such violations look psychologically natural. Bernheim [3] discusses an example based on experimental findings by Iyengar and Lepper [16]: strawberry jam is chosen when the only alternative options are a different variety of jam or ‘no jam’; however ‘no jam’ is chosen when all varieties of jam are available simultaneously. While the data are not literally a violation of Condorcet consistency (as no pairwise choices are involved) this is nevertheless a very acute manifestation of menu effects - presumably strawberry jam would be chosen in every pairwise choice. Significantly, the example points to a violation *only* of Condorcet, not of Pairwise, consistency.

We argue that in such cases an explanation based on group rankings, rather than on detailed pairwise comparisons, resolves the tension the agent faces in a complex choice situation. It seems natural to assert that, faced with ‘too many’ options, the jam picker contrasts {no jam} with {some jam}, and the doctor of the introduction contrasts {not prescribing} with {prescribing something}. Conversely, claiming that the jam picker chooses ‘no jam’ because this option

provided some regularity condition held to guarantee that the base relation has a maximal element in each choice set.

looks superior in a detailed binary comparison to every other jam looks suspect; and in the case of strawberry jam it is flatly contradicted by the actual choice. And in the doctors' predicament 'not prescribing' has been revealed to be inferior, in a straight binary comparison, to both medications.

3 Categorize Then Choose

Categorization appears very natural as soon as the set of alternatives exhibits some complexity.¹¹ Indeed in psychology categorization is a central concept for the analysis of human reasoning.¹² The grouping into categories occurs on the basis of some criterion. A traditional criterion considered in psychology is 'similarity', with each category comprised of alternatives which are 'similar' to each other.¹³ However many other methods for categorization may apply, depending on context.¹⁴ In our formal model we eschew this difficult and still open issue to focus directly on the choice theoretic consequences of categorization, rather than on the process of category formation (we discuss some of the economic approaches to categorization in the concluding section).

Formally, we assume that categories (subsets of X) can be (partially) com-

¹¹The complexity of decisions is certainly related to the number of options to be considered, but not exclusively so. Even simple choice sets may present difficult choices. In the medical examples above, although the decisions involve at most three alternatives, it is clear that any choice set including two medications is far more complex for the doctor than the choice between prescribing a given medication and not prescribing: the doctor's predicament does not stem from the *number* of alternatives, but from their nature. Similarly, modern cameras have dozens of features to be considered by the consumer. Even a binary choice may be complex, and it may be determined by 'categorising' the cameras (e.g. by brand), rather than by a detailed pairwise comparison.

¹²See e.g. Smith, Patalano and Jonides [45] and Medin and Aguilar [24].

¹³Rubinstein [37] pioneered in economic theory the analysis of similarity considerations in decision-making. See also Leland [17].

¹⁴Smith, Patalano and Jonides [45] even find indirect evidence of neurological differences between 'rule-based categorization' and 'similarity-based categorization' (with the former involving frontal regions, and the latter involving posterior areas). See Pothos [34] and the ensuing discussion for a more recent overview of the relationship between these two types of categorization.

pared:

Definition 2 A *shading relation* on X is an asymmetric (possibly incomplete) relation \succ on $2^X \setminus \emptyset$.

When $R \succ S$, category S is ignored if category R is available: we say that R *shades* S .

More than one criterion may be relevant in the categorization stage, hence we do not assume that categories are mutually disjoint (think of categorizing cameras by price band and brand, or flights by cheapness and convenience). We study some variants of this basic model in section 4.

Definition 3 Given a shading relation \succ and $S \in \Sigma$, the \succ -maximal set on S is given by:

$$\max(S, \succ) = \{x \in S \mid \text{for no } R', R'' \subseteq S \text{ it is the case that } R' \succ R'' \text{ and } x \in R''\}$$

We shall call any asymmetric and complete binary relation on X a *preference*.

We are now ready for our main definition:

Definition 4 A choice function γ is *Categorize-Then-Choose (CTC)* if and only if there exists a shading relation \succ and a preference \succ^* such that for all $S \in \Sigma$:

$$\gamma(S) \in \max(S, \succ) \text{ and } \gamma(S) \succ^* y \text{ for all } y \in \max(S, \succ) \setminus \{\gamma(S)\},$$

In this case \succ and \succ^* are said to *rationalize* γ .

In words, the CTC agent looks first at category rankings within the feasible set, eliminating all alternatives (e.g. expensive flights, Mexican restaurants) which belong to a category dominated by another category. Then, he picks among the remaining alternatives the one that he prefers to all others. When this procedure leads to a single chosen alternative for each choice set, the resulting choice function is CTC.¹⁵

The definition is meant to accommodate the reasonable possibility that some groups of alternatives are not comparable to others, and thus survive the first stage even though they are not a ‘most salient’ category. For example, suppose that beside Italian and Mexican restaurants there are also Alaskan ones. The agent neglects Mexican restaurants given the availability of Italian ones. But he has no idea about Alaskan cuisine, and he is willing to compare Alaskan restaurants in detail with the Italian ones in the second stage. Alaskan restaurants are not specially salient, they are simply not obviously inferior to any other category. So we are taking a ‘negative’ view of the first stage elimination: the agent ignores losers. Later we shall also consider a ‘positive’ view, in which the agent focusses on winners, that is elements of a most salient category that trumps all others.

CTC choice functions can explain menu dependence, as they need not satisfy Condorcet consistency. For instance, take a choice function with the base relation as indicated in figure 1 (where an arrow going from a to b indicates that a is selected in pairwise choice between a and b), and $\gamma(\{x, y, w, z\}) = y$. Condorcet

¹⁵Needless to say, there are ways of categorising and of constructing the shading relation such that the set $\max(S, \succ)$ is empty (for example if $\{x\} \succ \{y, z\}$ and $\{y\} \succ \{x\}$). But we need not worry about these cases: we will just say that the choice function is not CTC. Presumably an agent with categories that result in an empty selection will revise his categorisation method.

consistency is violated, since x is chosen in pairwise comparisons over each of the other alternatives but is not chosen from the grand set. However, this choice function is CTC with, for example, $\{y\} \succ \{x, w, z\}$, and \succ^* coinciding with the base relation \succ_γ .

CTC choice functions can also explain pairwise cycles of choice. Consider the basic three-cycle $\gamma(\{x, y, w\}) = \gamma(\{x, y\}) = x$, $\gamma(\{y, w\}) = y$, $\gamma(\{x, w\}) = w$. This is a CTC using $\{x, y\} \succ \{w\}$ and with the preference \succ^* coinciding with the base relation.

But CTC choice functions are not a vacuous notion: they do provide testable restrictions on behavior. As an example, let $X = \{x, y, w, z\}$ and let $\gamma(X) = x$ and $\gamma(\{x, y, z\}) = y$, with the base relation as in figure 2. Then, since y is chosen in $\{x, y, z\}$, there must be categories $R, R' \subset \{x, y, z\}$ with $R \succ R'$ and $x \in R'$, so that x is eliminated before it can eliminate y . Since $\{x, y, z\} \subset X$, the fact that $\gamma(X) = x$ is a contradiction.

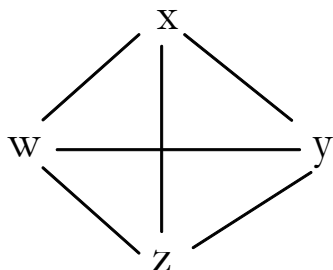


Figure 1

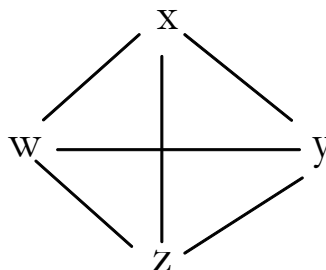


Figure 2

CTC choice functions are characterized by one single property, Weak WARP (WWARP), which we introduced in [22].

WWARP: For all $R, S \in \Sigma$: If $\{x, y\} \subset R \subset S$ and $x = \gamma(\{x, y\}) = \gamma(S) \neq y$ then $y \neq \gamma(R)$.

WWARP weakens WARP. It says that if you choose steamed salmon over steak tartare when these are the only available choices, and you also choose steamed salmon from a large menu including steak tartare, then you cannot choose steak tartare from a small menu including steamed salmon. In other words, if adding a large number of alternatives to the menu does not overturn a revealed preference, then adding just a subset of those alternatives cannot overturn the revealed preference either. In this sense WWARP can be seen alternatively as a ‘monotonicity’ restriction on menu effects.

Remark 1 *WWARP is equivalent (on this domain) to the following stronger looking property: If $x, y \in R \subset S \subset T$ and $y \neq x = \gamma(R) = \gamma(T)$, then $y \neq \gamma(S)$. In other words, the ‘small set’ in the statement of WWARP needs not be binary. To see that WWARP implies the property, suppose the latter fails, that is: $x, y \in R \subset S \subset T$, $y \neq x = \gamma(R) = \gamma(T)$, and $y = \gamma(S)$. If $x = \gamma(\{x, y\})$, then WWARP is violated using the sets $\{x, y\}$, S and T . If $y = \gamma(\{x, y\})$, then WWARP is violated using the sets $\{x, y\}$, R and S . This equivalence breaks down on domains that do not include all pairs.*

Theorem 1 *A choice function γ is CTC if and only if it satisfies WWARP.*

Proof: Necessity. Suppose that γ is CTC with shading relation \succ and preference \succ^* . Suppose $x = \gamma(\{x, y\})$ and $x = \gamma(S)$ with $y \in S$. Now suppose by contradiction that $y = \gamma(R)$ with $x \in R \subset S$. This means that $x \notin \max(R, \succ)$, since $x = \gamma(\{x, y\})$ implies $x \succ^* y$ (the possibility $\{x\} \succ \{y\}$ yields an immediate contradiction with $y = \gamma(R)$). In particular there exist $R', R'' \subseteq R$, such that $R' \succ R''$ and $x \in R''$. Since $R', R'' \subset S$ this contradicts $x = \gamma(S)$.

Sufficiency. The proof is written with an eye to the model variants suggested

in Proposition 2 below. Define: $x \succ^* y$ if and only if $x = \gamma(\{x, y\})$. \succ^* is obviously asymmetric and complete. Fixing the choice function γ , we define the upper and lower contour sets of an alternative on a set $S \in \Sigma$ as

$$Up_\gamma(x, S) = \{y \in X | y \succ_\gamma x\} \cap S$$

and

$$Lo_\gamma(x, S) = \{y \in X | x \succ_\gamma y\} \cap S$$

respectively. Define: $R \succ S$ if and only if there exists $T \in \Sigma$ such that

$$R = \{\gamma(T)\} \cup Lo_\gamma(\gamma(T), T)$$

and

$$S = Up_\gamma(\gamma(T), T) \neq \emptyset$$

The relation \succ is also obviously asymmetric (with an eye to Proposition 2, note that $R \cap S = \emptyset$ and that $|R \cup S| > 2$ whenever R and S are related by \succ).

Now let $S \in \Sigma$ and let $x = \gamma(S)$. We show that x is not eliminated in either round. Suppose first that $y \succ^* x$ for some $y \in S$. Then by construction

$$\{x\} \cup Lo_\gamma(x, S) \succ Up_\gamma(x, S)$$

and $y \notin \max(S, \succ)$.

Next, suppose by contradiction that $x \notin \max(S, \succ)$. Then there exists $R', R'' \subset S$ with $R' \succ R''$ and $x \in R''$. Define $R = R' \cup R''$. By construc-

tion of \succ it must be

$$R' = \{\gamma(R)\} \cup Lo_\gamma(\gamma(R), R)$$

and

$$R'' = Up_\gamma(\gamma(R), R)$$

(Notice that here a separate assumption of closure under union of the domain is not needed). This means that

$$x = \gamma(\{x, \gamma(R)\})$$

Together with $x = \gamma(S)$ (and noting that $R = R' \cup R'' \subseteq S$) this contradicts either $x = \gamma(S)$ (if $R' \cup R'' = S$) or WWARP (if $R' \cup R'' \subset S$). Finally note that by construction, $x \succ^* y$ for all $y \in \max(S, \succ)$ (since then $y \in Lo_\gamma(x, S)$). ■

4 Alternative formulations

In this section we study two variants of the core model. This allows us to bring to the fore several subtleties in the definition of categorisation.

4.1 Choosing from a maximally salient category

Recall our interpretation of the relation \succ as a psychological shading relation. There is some ambiguity in the verbal examples we have provided which has to be resolved in the formal model. In particular, we have assumed so far that the agent simply ignores all alternatives which are in a shaded category. But we could have instead considered a distinct cognitive process, that of focussing on a ‘maximally salient’ category. In the CTC model, the set of alternatives

considered by the agent may not form a category that shades all other categories (it is simply not shaded by them); indeed it may not form a category at all. To clarify, suppose all restaurants are either cheap or expensive, and either Italian or Mexican. {cheap restaurants} shades {expensive restaurants}, and {Italian restaurants} shades {Mexican restaurants}. There are no other psychologically meaningful categories. Then the agent ends up considering (i.e. applying a preference relation \succ^* to) only cheap Italian restaurants, even though we did not assume there was a category {cheap Italian restaurants} to begin with; let alone one that served to eliminate the other alternatives. We now consider a model in which the agent picks instead from a maximally salient category.

Formally, given a shading relation \succ , define the set of categories in S by

$$\mathcal{C}(S, \succ) = \{R \subset S \mid R \succ R' \text{ or } R' \succ R \text{ for some } R' \subset S\}$$

Say that a category R_S^* is *maximally salient* in S if

$$R_S^* \succ R \text{ for all } R \in \mathcal{C}(S, \succ) \setminus R_S^*$$

Obviously, if there is a maximally salient category in S , it is unique. Consider now the following alternative procedure:

Definition 5 A choice function γ is a *Saliency CTC* if and only if there exists a shading relation \succ and a preference \succ^* such that for all $S \in \Sigma$:

1. If there exists a maximally salient R_S^* , then

$$\gamma(S) \in R_S^* \text{ and } \gamma(S) \succ^* y \text{ for all } y \in (S \cap R_S^*) \setminus \{\gamma(S)\}$$

2. Otherwise $\gamma(S) \succ^* y$ for all $y \in S \setminus \{\gamma(S)\}$

Note that we are requiring that the agent must pick *exclusively* from the maximally salient category R_S^* (if there is one) in the second stage. This is a strong requirement, since a maximally salient category must shade all other categories but it might not eliminate all alternatives which do not belong to it. If some uncategorized alternatives (say an Alaskan and a Nasi Padang restaurant) survived the first stage, the category {Italian restaurants} could hardly have a claim to maximal salience even if it shaded the only other category {Mexican restaurants}, since the Alaskan and Nasi Padang restaurants would still be competing for attention with the Italian restaurants. Furthermore, if there is no maximally salient category in S , the chosen alternative is forced to beat pairwise all other alternatives in S , since no alternative is eliminated in the first stage.

Despite these strong requirements, a Saliency CTC hardly imposes any constraint on choices in terms of the properties we have considered. The following examples show in fact that a Saliency CTC may accommodate violations of both WWARP and Condorcet Consistency (and therefore Expansion).

A violation of Condorcet consistency: Let $X = \{x, y, z\}$, and let $x = \gamma(X)$ and $y = \gamma(\{x, y\}) = \gamma(\{z, y\})$. This is a Saliency CTC with shading relation $\{x\} \succ \{y, z\}$ and preference $y \succ^* x, z$.

A violation of WWARP: Let $X = \{x, y, w, z\}$, and let $x = \gamma(\{x, y\}) = \gamma(X)$ and $y = \gamma(\{x, y, z\})$. This is a Saliency CTC with shading relation $\{y\} \succ \{x, z\}$, $\{x, w\} \succ \{y, z\}, \{x, z\}, \{y\}$ and preference $x \succ^* y, w$.¹⁶

¹⁶The relation $\{x, w\} \succ \{x, z\}$, in which x appears in the sets on both sides, is not as strange as it may initially seem. Suppose that x and w are the flights on your favourite airline, whereas

The second example in particular shows that a Saliency CTC is not a strengthening of a CTC. A characterization of a Saliency CTC remains an open question. But we shall see (in section 5) that, strikingly, when an additional consistency requirement is imposed on the categorization process, a Saliency CTC - unlike a CTC with the same consistency requirement - allows no deviation whatsoever from fully rational behaviour.

4.2 Comparing only disjoint or only nontrivial categories

So far, apart from asymmetry, we have not imposed any restriction on the shading relation \succ on categories: any collection of subsets of alternatives qualified as a category; and any two categories could be compared. However in some contexts it may make sense to impose restrictions across categories. We note here that Theorem 1 holds true even under two additional, often natural, assumptions on the shading relation.

Definition 6 *A choice function γ is a **Restricted CTC** if and only if it is a CTC and the shading relation \succ satisfies the following two conditions:*

- (i) $R \cap S = \emptyset$ whenever $R \succ S$;
- (ii) $|R \cup S| > 2$ whenever $R \succ S$.

Condition (i) above says that the decision-maker can only compare disjoint categories. This is natural when a category is identified by a property (or a set of properties). For the agent who focusses solely on cheap flights (R), for example, the only relevant comparison in $S \in \Sigma$ might be $R \succ S \setminus R$. The assumption also

x and z are the flights which are both cheap and convenient. Flights with your favourite airline shade all cheap and convenient flights, which is of course compatible with flight x belonging to both categories.

solves the predicament an agent might find himself in when comparing categories that overlap: if x is both in a shading and in a shaded category, should x be considered or not?

Condition (ii) requires that the relation \succ makes comparison involving at least one genuine *group*: degenerate comparisons between singletons are not allowed. This is a benchmark, extreme case to capture situations in which complexity is driven at least in part by the number of objects in the choice set.

These restrictions do not constrain choice data in any different way. The proof of Theorem 1 has also proved:

Proposition 2 *A choice function γ is a Restricted CTC if and only if it satisfies WWARP.*

The main interest of the Restricted CTC model is that if there is a preference relation \succ^* that rationalizes γ , it is unique: that is, if (\succ, \succ^*) and (\succ', \succ'^*) both rationalize γ , then $\succ^* = \succ'^*$. This is because in this model the only way to rationalize pairwise choices is by means of the preference, not the shading relation, and therefore both \succ^* and \succ'^* must coincide with the base relation. So, while there is some leeway in recovering the categories from choice data and the ranking between them, Restricted CTC choice functions are designed in such a way that their preference relation is pinned down uniquely. This preference relation may be non-standard as it may include cycles, but *these cycles never involve the alternative chosen in the post-categorization stage*: so, the chosen alternative is always strictly preferred to all rejected alternatives which survive to the post-categorization stage.¹⁷

¹⁷Ehlers and Sprumont [9] and Lombardi [19] also consider possibly cyclical preference relations. They study which choice functions can be ‘rationalized’ as the top cycle or the uncovered

5 Consistent categorizers look like rational (or sequentially rational) agents

In all versions of the CTC model considered so far, the shading relation may not be consistent across subsets. It may occur that {Antonio’s place, Gino’s place, Maria’s place} shades {El Taco} in the choice set comprising all four restaurants, but that the agent does not use {Maria’s place} as a category to shade {El Taco} in the choice set {Maria’s place, El Taco}: in this latter case he just compares the two restaurants in detail, without categorization. So it is possible that alternatives in a shading category in some set S are in no category in a subset of S , or that in such subset they are now in a shaded category. This possibility is descriptive of many examples. In a ‘scarcity bias’ example (Mittone and Savadori [29], Mittone, Savadori and Rumiati [30]), for a small child in a toy shop {three brown teddy-bears} shades {fifteen beige teddy bears} and {three beige teddy bears} shades {fifteen brown teddy bears}, but presumably no shading relation exists between {three beige teddy bears} and {three brown teddy bears}. In a restaurant with a wide selection of fish (suggesting the chef’s ability to cook it properly), it would be unsurprising if {bream, carp, sole, lobster,...} shaded {grilled chicken breast}, but {grilled chicken breast} may shade {carp} when carp is the only fish item on the menu (in which case chicken is the safer choice). Recall also the medication example: in that case there is an intuitive sense in which the contrast between {Prescribing some medication} and {Not prescribing} is of a different psychological nature from that between {Prescribing medication A} and {Not prescribing}.

set of an asymmetric and complete preference relation (a tournament).

Nevertheless, in other leading cases in which categories include similar objects, such as {Italian restaurants}, it may become a sensible requirement that categories exhibit consistency across subsets: no matter how many Italian restaurants are closed, the ones that are open still shade all the available Mexican ones. And at the theoretical level, the question of what restrictions consistent categories imply for an agent's choices merits investigation.¹⁸

Formally, consider the following condition on the shading relation:

Consistency: (i) If $R \in \mathcal{C}(X, \succ)$ and $R \cap T \neq \emptyset$ for some $T \in \Sigma$, then $R \cap T \in \mathcal{C}(T, \succ)$. (ii) If $R \succ S$ then, for any $T \in \Sigma$ for which $R \cap T \neq \emptyset \neq S \cap T$, it is the case that $R \cap T \succ S \cap T$.

We apply the idea of consistency to both versions of the model, CTC and Saliency CTC. We shall see that the consistency condition has quite different effects in the two cases.

Definition 7 *A choice function γ is a **Consistent CTC** (respectively, a **Consistent Saliency CTC**) if and only if it is a CTC (Saliency CTC) and the shading relation \succ satisfies Consistency.*

The Consistency requirement turns out to greatly limit the power of categorization to explain non-standard choices. In order to characterise Consistent CTCs, we need to state another axiom:

Expansion: Let $\{S_i\}$ be a class of sets such that $S_i \in \Sigma$ for all i , and $\cup_i S_i \in \Sigma$.

If $x = \gamma(S_i)$ for all i then $x = \gamma(\cup_i S_i)$.

¹⁸We thank two referees for directing our attention to this category of models, which was shaded in an earlier version of the paper.

Expansion says that if steamed salmon is chosen from each of a series of menus, then it is also chosen when all the menus are merged; it obviously implies Condorcet consistency.

Proposition 3 *A Consistent CTC satisfies Expansion.*

Proof: Suppose that γ is a Consistent CTC with shading relation \succ and preference \succ^* , and let $x = \gamma(T_i)$ for all T_i in a class of sets $\{T_i\}$. Suppose by contradiction that $\bigcup_{i \in I} T_i \in \Sigma$ and $x \neq \gamma(\bigcup_{i \in I} T_i)$, so that there exist $R, S \subset \bigcup_{i \in I} T_i$ such that $R \succ S$ and $x \in S$. There is at least one j for which $R \cap T_j \neq \emptyset$. Also, since by assumption $x = \gamma(T_j)$, $x \in S \cap T_j$. Thus by Consistency $R \cap T_j \succ S \cap T_j$, and thus $x \notin \max(T_j, \succ)$, a contradiction. Therefore $x \in \max(\bigcup_{i \in I} T_i, \succ)$. And for any y and T_i with $y \in T_i$ and *not* $(x \succ^* y)$ it must be the case (since $x = \gamma(T_i)$) that $y \notin \max(T_i, \succ)$, which implies $y \notin \max(\bigcup_{i \in I} T_i, \succ)$. We conclude that $x = \gamma(\bigcup_{i \in I} T_i)$. ■

Since Expansion is a strengthening of Condorcet consistency, this result shows that the Consistent CTC model ends up forbidding *any* form of menu dependence in choice. It thus has no more explanatory power than the notion of Rational Shortlist Method, a special two-stage version of Sequential Rationalizability, which we introduced in Manzini and Mariotti [22] (RSM, on the full domain, is characterised by WWARP and Expansion).¹⁹

But the consequences of the Consistency requirement are even more dramatic if we consider the alternative version of the model, Saliency CTC.

¹⁹A choice function γ is a Rational Shortlist Method if and only if there exists an ordered pair (\succ_1, \succ_2) of asymmetric binary relations on X such that:

$$\text{For all } S \in \Sigma: \{\gamma(S)\} = \max(\max(S, \succ_1), \succ_2)$$

Proposition 4 *A Consistent Saliency CTC is fully rational.*

Proposition 4 is implied by a more general result, which holds on any arbitrary domain (and which in particular does not necessarily include all pairs and triples). A standard rationality property is:

IIA: If $\gamma(T) \in S \subset T$ then $\gamma(S) = \gamma(T)$.

IIA (for Independence of Irrelevant Alternatives) says that if steamed salmon is chosen in a large menu, it remains chosen in a smaller menu that includes it. This property is equivalent to WARP if the domain is finite.

Lemma 1 *A Consistent Saliency CTC satisfies IIA.*

Proof: Let γ be a Consistent Saliency CTC with shading relation \succ and preference \succ^* . Suppose by contradiction that $\gamma(T) \in S \subset T$ but $\gamma(S) \neq \gamma(T)$. Since γ is a Consistent Saliency CTC it must be either (a) $\gamma(T) \succ^* y$ for all $y \in T \setminus \{\gamma(T)\}$ or (b) $\gamma(T) \in R_T^* \succ R$ for all $R \in \mathcal{C}(T, \succ) \setminus R_T^*$, where R_T^* is the maximally salient category in T . Moreover note that by Consistency $\mathcal{C}(S, \succ) \subset \mathcal{C}(T, \succ)$. In case (a) $\gamma(T) \succ^* y$ for all $y \in S \setminus \{\gamma(T)\}$ and therefore $\gamma(T) = \gamma(S)$. In case (b) we have $R_S^* = R_T^* \cap S$ (note that $R_T^* \cap S \neq \emptyset$ since $\gamma(T) \in R_T^* \cap S$). This implies that $\gamma(S) \in R_T^*$. But then: If (i) $\gamma(T) \succ^* \gamma(S)$, then γ violates the definition of a Consistent Saliency CTC on S ; if (ii) $\gamma(S) \succ^* \gamma(T)$, then γ violates the definition of a Consistent Saliency CTC on T ; and if neither (i) nor (ii) the definition is violated on both S and T .

■

Proposition 4 now follows from the fact that on our domain IIA and WARP are equivalent.

Thus, the choice behaviour of a consistent categorizer is in this case indistinguishable from that of a textbook preference maximizing agent!

One interesting feature of this result is that we did not explicitly impose, in the definition of a Consistent Saliency CTC, any restriction concerning transitivity, either on the shading relation or on the preference relation. Transitive behaviour emerges purely from the assumed consistency in the way categories are formed and compared.

We draw two lessons from this analysis. First, the power of the CTC model to explain irrational choice behaviour stems from the freedom in the cognitive process of categorization: as soon as that process is required to possess consistency, at least menu effects, and possibly even pairwise cycles, can no longer be explained. The second lesson concerns the limitations of ‘revealed preference axioms’ *alone* in pinpointing the exact decision process followed by the agent: categorizing agents may look identical, in terms of observable choice behaviour, to RSM users or even to preference maximizers. We return to this theme in section 7.

6 Related choice-theoretic literature

The literature on categorization in a broad sense is huge, and a complete survey would go beyond the scope of the present paper. We discuss our contribution to the far less developed choice theoretic literature that studies two-stage ‘focus and choose’ procedures. Three main papers are related to ours in this respect: Cherepanov, Feddersen and Sandroni’s ‘Rationalization’ model [7]; Masatlioglu, Nakajima and Ozbay [28] model of ‘Limited attention’; and Lleras, Masatlioglu,

Nakajima and Ozbay [18] ‘Limited consideration’ model. There are two features that these papers share with ours:

1. decision makers follow a two stage procedure to arrive at choice;
2. in the first stage, some psychological mechanism selects only a subset of the initial set of alternatives as psychologically relevant.

All three papers mentioned above formalise the mechanism as some map $\mu : \Sigma \rightarrow \Sigma$ with the property that $\mu(S) \subseteq S$ for all $S \in \Sigma$. The standard rationality model can be seen as one where $\mu(S) = S$ for all S in the domain: the decision maker considers all the alternatives in the set. Different psychological mechanism that select in the first stage the alternatives from which the decision maker will make his choice in the second stage call for different restrictions on μ . The properties imposed on μ are what differentiates each of the three aforementioned models from one another.

In their Revealed attention model, Masatlioglu, Nakajima and Ozbay [28] consider the map μ^{MNO} by imposing on μ the following property:

$$s \notin \mu^{MNO}(S) \Rightarrow \mu^{MNO}(S) = \mu^{MNO}(S \setminus \{s\}) \quad (1)$$

The map μ is interpreted as an *attention filter*, that selects only those alternatives in a set to which the decision maker effectively pays attention. In this interpretation, condition (1) is very plausible: if an alternative is not attended to (i.e. $s \notin \mu^{MNO}(S)$) in a set, then removing it from the set does not affect which other alternatives are attended to. The decision maker then selects in the second stage, from those alternatives which are attended to, the one which

maximises a complete and transitive binary relation. The authors show that this choice procedure is fully characterised by a single axiom which is a weakening of WARP (Limited Attention WARP), but that has no logical connections with WWARP.

In a companion paper, Lleras, Masatlioglu, Nakajima and Ozbay [18] provide a characterization of their choice procedure when μ^{MNO} is a *consideration filter*, that is when for any pair of sets S and T in the domain,

$$x \in S \subset T \text{ and } x \in \mu^{LMNO}(T) \Rightarrow x \in \mu^{LMNO}(S) \quad (2)$$

A consideration filter imposes a form of consistency. If an alternative x is considered in a superset, then it must be considered also in a subset. In practice it imposes IIA on the filtering map μ . It is somewhat reminiscent of our Consistency restriction, in the sense that we, too, require that if an alternative is in a category in a superset, it is also in a (intersection) category in a subset. Interestingly, however, the adoption of property (2) leads to a model that is *not* equivalent to full rationality. If the choice from each set maximises some linear order over the set $\mu^{LMNO}(S)$ for each set S , then it is dubbed choice by Limited Consideration. This choice procedure is fully characterised by a single axiom that the authors introduce, Limited Consideration WARP. This property implies WWARP, so that choice by Limited Consideration is also CTC.²⁰ Eliaz and Spiegler [10] independently study a full blown economic application of this idea.

²⁰Details available from the authors.

Cherepanov, Feddersen and Sandroni [7] consider instead the map μ^{CFS} with the following property: given a set of asymmetric and transitive relations (rationales) $\{\succ_1, \succ_2, \dots, \succ_K\}$,

$$\mu^{CFS}(S) = \{s \in S : s \succ_i s' \text{ for all } s' \in S \setminus \{s\} \text{ for some } i = 1, \dots, K\}$$

Each rationale is interpreted as a possible motivation underpinning the agent's decision, and only alternatives that are best according to some motivation survive to the next stage.

The agent then selects, in the second stage, the alternative that is best according to an asymmetric relation (i.e. a preference). Interestingly, this procedure generates (on the full domain) exactly the same choice data as a CTC choice function, since Cherepanov, Feddersen and Sandroni show that a decision maker follows this procedure if and only if his choices satisfy WWARP.

The last model, then, is observationally equivalent to our CTC model, thus indistinguishable from it on the basis of choice data alone. Obviously, though, the underlying choice procedures are very different positive models of behaviour. Nevertheless they may have concordant implications in terms of welfare analysis. We discuss this point in the next section.

To conclude, we note that the models studied in this paper fall into a more general choice theoretic literature, which studies procedures whereby the decision maker discards round after round (possibly more than two) those alternatives that are inferior according to some criterion, with only the chosen alternative surviving all rounds of elimination. The criteria applied in the existing literature are either properties (as in Mandler, Manzini and Mariotti [21]); or binary

relations (as in Manzini and Mariotti [22], Apesteguia and Ballester [1], Houy and Tadenuma [14]). In the notation introduced here, the ‘checklist’ procedure of Mandler, Manzini and Mariotti [21] corresponds to a sequence of maps μ_i applied one after the other to each previous ‘survivor set’. In terms of predicted choice behaviour, the checklist model is broadly equivalent to full rationality, as is the Consistent CTC. When instead the decision maker applies sequentially binary relations, depending on the restrictions on such binary relations, we obtain models in which CTC is not necessarily nested (WWARP is not necessary for sequential rationalizability).

7 Remarks on welfare analysis

When WARP is violated the decision maker may ‘reveal preferences’ which are contradictory. Basing welfare judgements on a revealed preference approach, as is standard in normative economics, might seem a desperate task in such circumstances. We would like to use our model as a concrete setting for some remarks on welfare analysis in the presence of boundedly rational agents. These comments are not to be seen as a full-blown welfare theory, but rather as first steps in approaching a difficult and pressing problem.²¹

7.1 Identification problems

Suppose that through choice evidence and other types of evidence we have confidence that the CTC model is the correct description of the agent’s decision

²¹In a paper in preparation, we offer a more general study of the issue of welfare and bounded rationality. On this issue in particular we have greatly benefitted from exchanges with Doug Bernheim and from penetrating comments by the referees, which have led to a deeper understanding both of the Bernheim and Rangel [4] approach and of our own model.

making procedure (non-choice evidence is needed, for example, to distinguish our model from the Rationalization model or the Limited consideration model). How would we characterise the agent's welfare? It may appear obvious that the preference \succ^* is always welfare-relevant, but even so it is not obvious that *only* this component is relevant. One can argue that the shading relation \succ might also sometimes include preference, and thus welfare, components. Because \succ is in any case not a formally standard preference, the judgement has to be necessarily nuanced according to circumstances, including at least three broad varieties:

- If your department decides that the hiring priority must go to Micro candidates, so that the category {Micro candidates} shades {Macro candidates}, we can perhaps say that your department is better off with a micro candidate: here the shading relation has a clear welfare component. Similarly, if an investor ignores risky portfolios or a chess player ignores a class of stupid moves, the shading relation reflects the agent's welfare.
- If {Italian restaurants} shades {Mexican restaurants}, it may be because in general you eat better in Italian restaurants. But this does not exclude that the occasional Mexican restaurant offers a menu which you would judge superior to that of some Italian restaurants. Here the shading relation incorporates a sort of 'average' or 'expected' welfare component.
- If the category {items on the first page of a google search} shades the category {other items}, the shading relation does not have any immediate preference content (as noted, Masatlioglu, Nakajima and Ozbay [28] study in great detail this pure 'attention' aspect of consideration sets.)

Even these statements presuppose that we have been able to identify that \succ and \succ^* are what we say they are. It might be, for example, that while formally the agent follows the CTC procedure, \succ^* represents in fact a salience ordering and not a preference. Finally, as we have observed, other models with different primitives might explain the same set of data. These ‘identification problems’ have been discussed extensively by Bernheim [3]. They are common to all models of boundedly rational choice and appear to present a formidable challenge to welfare analysis.

Before proceeding, let us note an important point. If one believes that the actual decision procedure followed by the agent is relevant in order to make normative judgments, then identification problems pose a challenge of exactly equal force to the textbook model of choice as preference maximization and to standard welfare analysis based on such preferences. An external observer faced with choice data satisfying WARP is not automatically legitimised to take the inferred ‘revealed preference’ relation P as a welfare ranking. He might be facing a boundedly rational agent who is one of the consistent categorizers of section 5, or one of the checklist users of Mandler, Manzini and Mariotti [21]: both these models satisfy WARP and are thus indistinguishable from preference maximisation by means of choice data alone.²² Or, P might in fact just reflect a salience ordering (Tyson [49]), not a preference ordering; in an extreme (though not very plausible) case, the agent might be minimizing his welfare.

²²Salant [43] and Mandler [20] argue that there is a sense in which utility maximization is computationally efficient.

7.2 Possible solutions

One way to address these problems, advocated by Bernheim and Rangel [4] (henceforth BR) and Bernheim [3], is to give up trying to identify the decision procedure actually followed by the agent, and to simply respect choices ‘*qua* choices’ for the purpose of welfare analysis. In this approach, the challenge is that while the choices of a rational agent offer consistent guidance (P is transitive), those of a boundedly rational agent do not. However, BR have demonstrated that there exist ways of constructing consistent rankings starting from *any* set of choice data, whether they satisfy WARP or not.²³ They suggest to take as a welfare relation the relation P^* (strict welfare dominance) defined by xP^*y iff there is no $S \in \Sigma^*$ such that $x \in S$ and $y = \gamma(S)$, where Σ^* is a subdomain of Σ which is deemed relevant for welfare analysis.²⁴ The set of weak welfare optima in a given choice set is defined as the set of alternatives which are not P^* -dominated.

Let us try to apply the BR idea in the context of our CTC model. We consider a success case and a problematic case, starting from the latter.

The issue is *which* Σ^* should be adopted, since taking $\Sigma = \Sigma^*$ may be entirely inappropriate when the shading relation has no welfare content. For example, suppose that $X = \{x, y, z\}$ and that γ is CTC with preference given by $x \succ^* y, z$ and shading relation $\{y\} \succ \{x, z\}$, and let us add a further ingredient of rationality: \succ is consistent, so that also $\{y\} \succ \{x\}$ and $\{y\} \succ \{z\}$ hold. Suppose further that \succ has no welfare content (through marketing manipulation y just

²³In a paper in preparation, we argue that the Bernheim and Rangel methodology, while convincing, is not the only conceivable one that respects choices.

²⁴We have adapted the original definition, which applies to broader contexts, to our more limited setting.

happens to belong to a salient brand), whereas \succ^* expresses the true agent's welfare (BR would probably deny that we can have knowledge of this fact, and that is why they adopt the normative position described above of 'respecting choices *qua* choices'). Is there a subdomain Σ^* that reflects, through the BR methodology, the true welfare ranking $x \succ^* y, z$? In this case we have yP^*x, z , so that y is a unique weak welfare optimum in $\{x, y, z\}$ when $\Sigma^* = \Sigma$. The problem is that there is *no subdomain* Σ^* which can reflect the welfare inferiority of y with respect to x . In any subdomain that includes choice sets where x and y are available, we will have yP^*x . One possible way out is to simply disqualify y from the competition, refusing to consider any choice situation in which y appears (and thus is chosen). Another possibility is to reinterpret the choice sets $\{x, y, z\}$, $\{x, y\}$, and $\{y, z\}$ as $\{y\}$ before P^* is constructed.²⁵ After this reinterpretation, we can conclude that xP^*z , and we cannot rank y against x or z . It follows that x is now a welfare optimum; but so is y .

One can debate the respective merits of these strategies, but the crucial point we want to make here is that such reinterpretations of the choice correspondence would need a justification, and we would argue that a rigorous justification will be close to offering a model of the agent cognitive process and of his decision making procedure (e.g. 'the agent categorises products by brands and picks the most salient brand').²⁶ We are thus back to the issue of model identification.

While we are unable to offer a systematic procedure to resolve this issue, we

²⁵This is consistent with the guidance offered in footnote 43 of Bernheim and Rangel [4] (p. 84).

²⁶Of course, coarser strategies of elimination will need less information/assumptions. For example, one might simply delete from the domain all situations for which there is evidence that the set of alternatives considered by the agent is different from the set of alternatives that are actually available. Such coarser elimination strategies, however, run the risk of throwing away relevant choice data.

believe it is inescapable to some extent, even within the BR approach (unless one sticks rigidly to the view that disqualifications are unjustifiable and that one always needs to work with the unrefined choice domain).²⁷

Suppose now that in the example above the agent is CTC but not a consistent categorizer, and that the shading relation only asserts that $\{y\} \succ \{x, z\}$ (while \succ^* is as before). Here the set of weak welfare optima in $\{x, y, z\}$ is $\{x, y\}$, since $y = \gamma(\{x, y, z\})$ and $x = \gamma(\{x, y\})$. If we were somehow able to obtain the information that the choice in the grand set is ‘suspect’, then in applying the BR method we would exclude $\{x, y, z\}$ from the construction of P^* . Thus we would reach the conclusion that xP^*y , narrowing the set of welfare optima to $\{x\}$. Thus in this case information on the agent’s cognition easily permits the BR method to uncover the ‘true’ welfare relation by ruling out ‘suspect’ choice problems.

We show how the same welfare conclusion could be reached by the use of our model without discarding choice data. Our approach uses a different type logic from the one we have just considered, and thus exhibits some complementarity with the BR model-free approach. On the assumption that the restricted version of the CTC model holds, the binary choices are uniquely explained by \succ^* , and a welfare conclusion can be attained directly on the only assumption that \succ^* bears welfare content. But we can say something even on the assumption that the non-restricted version of the model holds. For then, $x = \gamma(\{x, y\})$ and $y = \gamma(\{x, y, z\})$ directly reveal that $x \succ^* y$: it cannot be that $\{x\} \succ \{y\}$ (for

²⁷In addition, we perceive the danger that, in the case of data consistent with WARP, an analyst not committed to model identification but rather to the application of a model-less welfare relation might be less inclined to ‘spoil’ data which so clearly appear to ‘reveal’ the maximization of a preference relation. In other words, the practice of attempting a model identification may reinforce a critical attitude towards the interpretation of the data.

in that case it could not be $y = \gamma(\{x, y, z\})$ and therefore the choice in $\{x, y\}$ must be driven by preference, not by categorization. This observation holds in general: whenever there is a conflict between a choice in a binary set and that in a larger set, the choice in the binary set is driven by preference and not by the shading relation. On the assumption that welfare judgements should not contradict preferences, we are home.

The identification issue of course remains. But we think that the most interesting aspect of this analysis is that in fact its main feature (that a binary choice contradicted in a larger set ‘reveals’ preference) does *not even need a full model identification* to hold. It is in fact common to *three* of the two-stage models we have discussed in section 6: ours, Cherepanov, Feddersen and Sandroni’s Rationalization model [7], and Lleras, Masatlioglu, Nakajima and Ozbay [18] Limited consideration model (these authors in fact devote a great deal of their analysis to the question of ‘preference extraction’, which we do not duplicate here). The difference between these models only pertains to the mechanism of the first stage selection (respectively categorization, rationalization and consideration).

We submit that this is a clue to a model-based solution, complementary to the BR approach, of the welfare riddle. Suppose that, as in this case, several plausible but distinct models converge in identifying a preference ‘revelation mechanism’. Then we would argue that this constitutes a strong basis to adopt those revealed preferences as the welfare ranking whenever the choice data support these models. In the example all three mentioned models can be rejected by the single axiom WWARP. Therefore this axiom, while not fully identifying a single model and thus exhibiting some limitation for positive analysis, does pro-

vide a wider basis for welfare analysis. From this angle the dreaded identification problem can be seen as a potential source of strength rather than of weakness.²⁸

Note also that in the Limited Attention model by Masatlioglu, Nakajima and Ozbay [28], the data $x = \gamma(\{x, y\})$ and $y = \gamma(S)$ with $x \in S$ do not necessarily suggest that x is preferred to y . That model admits in fact that y does not receive attention in $\{x, y\}$ while it does receive attention in $\{x, y, z\}$. In this case y might be preferred to x and not be chosen in a binary contest. The Limited Attention model is not characterised by WWARP. This is a clear illustration of how, in some circumstances, even pure choice data can serve a normative as well as a positive testing purpose.

The above remarks are not meant to hide the difficulty in identifying, even partially, decision procedures and their welfare implications. We still lack a systematic framework that guides the gathering and processing of the necessary extra-choice and contextual evidence. The fundamental observation by Bernheim [3] that revealed preference tests may have zero power against plausible alternatives maintains its force. Contextual information is likely to be useful, but there is a great deal of hand-waving in this assertion, until a formal framework is found to incorporate that type of information. Similarly, we believe we also still lack a systematic framework to guide us in the necessary pruning of the choice domain within the Bernheim and Rangel approach.²⁹

²⁸To highlight the complementarity with the BR approach, note that a similar argument could be made in their framework. The task there would be to find a set of justifications (eschewing a fully blown model of cognition) for restricting the welfare-relevant domain to binary (or at least to ‘small’) choice sets.

²⁹There is a number of very interesting contributions to this issue which for space reasons we could not discuss here: those by Chambers and Hayashi [6], Dalton and Ghosal [8], Green and Hojman [12], Rubinstein and Salant [39] and Sugden [46].

7.3 Some simple comparative statics

When the revealed preference $\succ_{\gamma} = \succ^*$ can be interpreted as a welfare ranking like in the examples above, some interesting comparative statics implications hold. In the fully rational model the chosen alternative in each set is better than any other feasible alternative. Therefore, adding an inferior alternative can be neither welfare worsening nor welfare improving. In the CTC model, the alternative chosen in a set is not necessarily preferred in pairwise choices to all other available alternatives. Thus it is in principle possible that if an inferior alternative is added, the decision maker shifts his choice to this or to another inferior alternative already present in the original choice set. Arguably, the abundance/complexity of alternatives may induce the decision maker in categorizations that mentally delete an alternative which would be valued better in pairwise comparisons.

But the other, positive side of the coin is that by *deleting* rejected alternatives (good or bad) the decision maker is welfare protected even if he changes his choice. This may not happen for other choice heuristics: for example in the ‘choose the median procedure’, in which alternatives are first listed according to some order and then one of the median elements in this order is selected, the deletion of a rejected alternative may reduce welfare.

Say that the choice set S is *welfare superior to* T if $\gamma(S) \succ_{\gamma} \gamma(T)$ (this terminology is justified by the previous discussion):

Proposition 5 (*Welfare comparative statics*) *Suppose the CTC model holds. If either (i) $S = T \setminus \{r\}$, $r \neq \gamma(T)$, or (ii) $T = S \cup \{r\}$, $\gamma(S) \succ_{\gamma} r$; then T cannot be welfare superior to S .*

Proof: (i) Immediate, by checking that if $\gamma(T) \succ_{\gamma} \gamma(S)$ WWARP would be violated. (ii) If $\gamma(T) \neq r$ WWARP is violated, and if $r = \gamma(T)$ then $\gamma(S) \succ_{\gamma} \gamma(T)$ by assumption. ■

Viewing the change as one from the small set S to the large set T , Proposition 5 highlights that the decision maker can be manipulated (by suitably inflating the feasible set) into making detrimental choices that constitute a welfare worsening, a phenomenon familiar to marketing experts (see e.g. Schwartz [44], Iyengar, Huberman and Jiang [15], and Reutskaja and Hogarth [41] for a recent example). Viewing the change as one from T to S , Proposition 5 highlights that a ‘simplification’ of the problem that removes (even good) alternatives can lead to violations of WARP but *not* to a welfare worsening: welfare is bounded below by that corresponding to the initial choice, and can possibly increase.

8 Concluding remarks

We have proposed a choice-theoretic model based on the cognitive phenomenon of categorization. Beside the vast psychology literature, some papers in economics have recognised the importance of categorization and have focussed on different aspects from those touched in this paper. Notably, Mullainathan [32], and Mullainathan, Schwartzstein and Schleifer [33] study the way categorisation affects the inferences made by the agents, and how such non-Bayesian’ inferences can be exploited by other economic actors. In the words of these authors, ‘individuals think coarsely’: aspects of the world are grouped together in categories, and inferences are made without distinguishing between situations in the same category. Fryer and Jackson [11] study a model of how categories are formed.

They invoke a sense in which categorization may be ‘optimal’, identifying some resulting biases in the categorization process and in the predictions that the agent makes.

In contrast, in this paper we have instead concentrated on the pure choice-theoretic implications of thinking in terms of categories.

We have provided a simple, one-axiom characterization of our model, that permits empirical testing with an extension of the standard ‘revealed preference’ economic approach, as pioneered by Samuelson [42].³⁰ We are not claiming that *only* direct choice data are relevant for economics (as argued for example by Gul and Pesendorfer [13]). Indeed we have even provided some evidence in the opposite direction, by showing how several highly distinct cognition/decision models (including the standard preference maximization model) place exactly the same restrictions on choice data. In this respect we have illustrated some of the points made by Bernheim [3] on the ‘identification’ difficulty inbuilt in a revealed preference type of approach. In particular, we have shown how consistent categorizers may behave exactly like fully rational agents.

Nevertheless, even these data alone can be extremely helpful in discriminating (1) between the standard choice model and models of boundedly rational choice, and (2) between different models of boundedly rational choice. For instance, our own data (in [23]) draw a very clear-cut and a priori unexpected line between psychologically plausible models such as Rational Shortlist Methods and CTC.

The setup we propose makes room for the widely observed phenomenon of

³⁰Beside those already mentioned in the paper, theoretical applications or discussions of this methodology are Dalton and Ghoshal [8], Rubinstein and Salant [38], Salant and Rubinstein [40], Caplin [5], Masatlioglu and Ok ([25], [26]) Masatlioglu and Nakajima [27], Tyson [49], among others.

menu dependent choice. Categorization has the virtue of simplifying a more complex choice task, but risks pointing the decision maker to alternatives he might not choose if he were to consider all options in greater detail. Our analysis offers some clues for dealing with the hard problem of welfare analysis in the presence of boundedly rational agents.

References

- [1] Apestequia, Jose and Miguel A. Ballester (2009) “Rationalizability of Choice by Sequential Procedures”, mimeo, Universitat Autònoma de Barcelona and Universitat Pompeu Fabra.
- [2] Arad, Ayala and Ariel Rubinstein (2010) “Colonel’s Blotto Top Secret Files”, working paper, Tel Aviv University.
- [3] Bernheim, B. Douglas (2009) “Behavioral Welfare Economics” (2009), *Journal of the European Economic Association*, 7: 267-319.
- [4] Bernheim, B. Douglas and Antonio Rangel (2009) “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *Quarterly Journal of Economics*, 124: 51-104.
- [5] Caplin, Andrew (2008) “Economic Theory and Psychological Data: Bridging the Divide”, in: *The Foundations of Positive and Normative Economics: A Handbook*, A. Caplin and Andrew Schotter (eds.), Oxford University Press.
- [6] Chambers, Christopher P. and Takashi Hayashi (2008) “Choice and Individual Welfare”, mimeo, HSS California Institute of Technology WP 1286.

- [7] Cherepanov, Vadim, Tim Feddersen and Alvaro Sandroni (2008) “Rationalization”, mimeo, Northwestern University.
- [8] Dalton, Patricio and Sayantan Ghosal (2008) “Behavioral Decisions and Welfare”, mimeo, University of Warwick.
- [9] Ehlers, Lars and Yves Sprumont “Weakened WARP and Top-Cycle Choice Rules”, *Journal of Mathematical Economics*, 44: 87-94 2008.
- [10] Eliaz, Kfir and Ran Spiegler (2007) “Consideration Sets and Competitive Marketing”, forthcoming in the *Review of Economic Studies*.
- [11] Fryer, Roland and Jackson, Matthew O. (2008) ”A Categorical Model of Cognition and Biased Decision Making”, *The B.E. Journal of Theoretical Economics*: Vol. 8 (1) (Contributions), Article 6.
- [12] Green, Jerry R. and Daniel Hojman (2008) ”Choice, Rationality and Welfare Measurement”, mimeo, Harvard Institute of Economic Research Discussion Paper No. 2144.
- [13] Gul, Faruk and Wolfgang Pesendorfer (2008) “The Case for Mindless Economics”, in: *The Foundations of Positive and Normative Economics*, by Andrew Caplin and Andrew Shotter (eds.). Oxford University Press.
- [14] Houy, Nicolas and Koichi Tadenuma (2009) “Lexicographic Compositions of Multiple Criteria for Decision Making”, *Journal of Economic Theory*, 144: 1770-1782.

- [15] Iyengar, Sheena S., Gur Huberman and Wei Jiang “How much choice is too much? Contributions to 401(k) retirement plans” In: *Pension Design and Structure*, chapter 5, pages 83-97. Oxford University Press, 2004.
- [16] Iyengar, Sheen S. and Mark R. Lepper (2000) “When choice is demotivating: Can one desire too much of a good thing?”, *Journal Personality and Social Psychology* 79(6):995-1006.
- [17] Leland, Jonathan W. (1994) “Generalized Similarity Judgments: An Alternative Explanation for Choice Anomalies”, *Journal of Risk and Uncertainty*, 9: 151-72.
- [18] Lleras, Juan S., Yusufcan Masatlioglu, Daisuke Nakajima and Erkut Y. Ozbay (2010) “When More is Less: Limited Attention”, mimeo, Michigan University.
- [19] Lombardi, Michele (2008) “Uncovered set choice rules”, *Social Choice and Welfare*, 31: 271 – 279.
- [20] Mandler, Michael (2009) “Rational Agents are the Quickest”, mimeo, Royal Holloway University of London.
- [21] Mandler, Michael, Paola Manzini and Marco Mariotti (2008) “One Million Answers to Twenty Questions: Choosing by Checklist”, mimeo, Queen Mary, University of London.
- [22] Manzini, Paola and Marco Mariotti (2007) “Sequentially Rationalizable Choice”, *American Economic Review*, 97: 1824-1839.

- [23] Manzini, Paola and Marco Mariotti (2009) “Revealed Preferences and Boundedly Rational Choice Procedures: an Experiment”, mimeo, University of St Andrews.
- [24] Medin, Douglas L. and Aguilar, Cynthia (1999) “Categorization”, in Robert A. Wilson and Frank C. Keil (eds) *MIT Encyclopedia of the Cognitive Sciences*, MIT Press, Cambridge, Ma.
- [25] Masatlioglu, Yusufcan and Efe Ok (2005) “Rational Choice with Status-quo Bias”, *Journal of Economic Theory* 121: 1-29.
- [26] Masatlioglu, Yusufcan and Efe Ok (2006) “Reference-Dependent Procedural Decision Making”, working paper, NYU.
- [27] Masatlioglu, Yusufcan and Daisuke Nakajima (2008) “Choice by Iterative Search”, mimeo, University of Michigan.
- [28] Masatlioglu, Yusufcan, Daisuke Nakajima and Erkut Y. Ozbay (2009) “Revealed Attention”, mimeo, University of Michigan.
- [29] Mittone, Luigi and Lucia Savadori (2009) “The Scarcity Bias”, *Applied Psychology* 58: 453-68.
- [30] Mittone, Luigi, Lucia Savadori and R. Rumiati (2005) “Does Scarcity Matter in Children’s Behaviour? A Developmental Perspective of the Scarcity Bias”, CEEL Working Paper 1/2005.
- [31] Moulin, Herve (1985) “Choice Functions Over a Finite Set: A Summary” *Social Choice and Welfare*, 2, 147-160.

- [32] Mullainathan, Sendhil (2000) “Thinking through Categories,” working paper, MIT.
- [33] Mullainathan, Sendhil, Joshua Schwartzstein and Anrei Schleifer (2008) “Coarse Thinking and Persuasion”, *Quarterly Journal of Economics*, 123: 577-619.
- [34] Pothos Emmanuel M. (2005) “The rules versus similarity distinction”, *Behavioral Brain Science* 28(1):1-14 (discussion 14-49).
- [35] Redelmeier Donald A. and Eldar Shafir (1995) “Medical decision making in situations that offer multiple alternatives”, *Journal of the American Medical Association*, 273(4): 302-5.
- [36] Roelofsma, Peter H. M. and Daniel Read (2000) “Intransitive Intertemporal Choice”, *Journal of Behavioral Decision Making*, 13: 161-177.
- [37] Rubinstein, Ariel (1988) “Similarity and Decision Making Under Risk (Is there a utility theory resolution to the Allais Paradox?)”, *Journal of Economic Theory*, 46: 145-153.
- [38] Rubinstein, Ariel and Yuval Salant (2008) “Some thoughts on the principle of revealed preference”, in: *The Foundations of Positive and Normative Economics: A Handbook*, Andrew Caplin and Andrew Schotter (eds.), Oxford University Press.
- [39] Rubinstein, Ariel and Yuval Salant (2010) “Eliciting Welfare Preferences from Behavioral Datasets”, mimeo, Northwestern University.

- [40] Salant, Yuval and Ariel Rubinstein (2008) “ (A, f) : Choice with Frames”, *Review of Economic Studies*, 75: 1287-1296.
- [41] Reutskaja, Elena and Robin M. Hogarth (2009) “Satisfaction in choice as a function of the number of alternatives: When “goods satiate” but “bad escalate””, *Psychology and Marketing*, 26 (3): 197-203.
- [42] Samuelson, Paul (1938) “A Note on the Pure Theory of Consumer’s Behavior”, *Economica* 5 (1): 61-71.
- [43] Salant, Yuval (2007) “Procedural Analysis of Choice Rules with Applications to Bounded Rationality”, forthcoming in the *American Economic Review*.
- [44] Schwartz, Barry (2000) “Self-determination: The tyranny of freedom”, *American Psychologist*, 55: 79-88.
- [45] Smith Edward E., Andrea L. Patalano and John Jonides (1998) “Alternative strategies of categorization” *Cognition* 65:167-96.
- [46] Sugden, Robert (2004) “The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences”, *American Economic Review* 94: 1014-1033.
- [47] Suzumura, Kotaro (1983) *Rational Choice, Collective Decisions, and Social Welfare*, Cambridge University Press, Cambridge U.K.
- [48] Tversky, Amos (1969) “Intransitivity of Preferences”, *Psychological Review*, 76: 31-48.

- [49] Tyson, Christopher J. “Cognitive constraints, contraction consistency, and the satisficing criterion”, *Journal of Economic Theory*, 138: 51 – 70
- [50] Waite, Thomas A. (2001) “Intransitive preferences in hoarding gray jays (*Perisoreus canadensis*)”, *Behavioral Ecology and Sociobiology*, 50: 116 - 121.