Revealed Preferences and Boundedly Rational Choice Procedures: an Experiment

Paola Manzini
University of St. Andrews and IZA

Marco Mariotti*
University of St. Andrews

DRAFT - this version: February 2010

Abstract

If people are irrational, how are they irrational? And how can we describe their behavior and perform welfare analysis? We study the question experimentally. We test several boundedly rational decision models which are testable by means of simple revealed preference type of axioms. In an experiment, we first show 'menu effects' to drive irrationality more than cycles of choice. Then, by using the revealed preference methodology, we show that a version of the Categorise Then Choose model we propose performs best (in terms of the Selten score of predictive success) in a group of models.

*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 contains the experimental material which was originally in "Categorize Then Choose: Boundedly Rational Choice and Welfare", from which it was excised. It also supersedes our earlier paper "Two-stage boundedly rational choice procedures: Theory and experimental evidence". For insightful discussions and comments we are grateful to Douglas Bernheim, Gianni de Fraja, Tim Feddersen, John Hey, Nagore Iriberri, 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

If people are inconsistent in their choices, how are they inconsistent? And how can we describe their behavior and perform welfare analysis? In the textbook description of a decision maker, choice behavior results from the maximization of some binary preference relation (possibly summarized by a utility function). Yet, choice behavior is often incompatible with this model.[1] This is 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? This paper studies experimentally the general nature of choice inconsistencies which lead to violations of the utility maximization model.

By relying on a theoretical result which classifies choice 'anomalies' into two elementary distinct phenomena (*pairwise cycles* and *menu effects*), we present laboratory choice data that identify one of the two types of violation as far more relevant than the other. The bulk of experimental literature on decision 'anomalies' has focussed on testing *pairwise* preferences between alternatives. This necessarily leads to disregarding menu effects. And in general, neglecting choices from non-binary sets, which are easily elicited, amounts to discarding an enormous amount of information: out of a grand set of 4 alternatives, there are only 64 possible different observations of pairwise choice, but over 20,000 patterns of choice from subsets: thus, many individuals who look indistinguishable on the basis of their pairwise choices might be distinguished by observing their 'higher order' choices. In this way, we also show that while the majority of individuals exhibits choice inconsistencies that are incompatible with utility maximization (specifically, they violate the Weak Axiom of Revealed Preference, or WARP in short), the large majority of them nevertheless exhibits a weaker but distinctive form of choice consistency (specifically, they satisfy a weaker property than WARP, WWARP in short, introduced in Manzini and Mariotti [21]).[2] This opens the possibility of describing behavior through an enriched model of preference maximization, and brings us to the second main goal of this paper.

In addition, we test how various choice procedures proposed in the recent and expanding literature on boundedly rational choice perform with our laboratory data. We focus

[1] See e.g. Roelofsma and Read [30], Tversky [40], and Waite [43] who find evidence of *pairwise cycles* of choice. Additional discussion on cycles is contained in Cherepanov, Feddersen and Sandroni [9]. *Menu effects* are another important class of choice anomalies which is 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 [26], Masatlioglu, Nakajima and Ozbay [27], Eliaz and Spiegler [11] The evidence presented in this paper points to further violations of rational behavior.

[2] WWARP adds to WARP the clauses between brackets in the following definition: if x is directly revealed preferred to y [both in pairwise contests and in the presence of a 'menu' of other alternatives], then y cannot be directly revealed preferred to x [in the presence of a smaller menu].

only on those choice procedures that are characterized in terms of few, simple conditions on observable data on choices out of feasible sets. This permits direct, simple and non-parametric tests of the models in the spirit of the standard 'revealed preference' economic approach as pioneered by Samuelson [35].

In both the permissive version (characterized by the single property WWARP) and the restrictive version (characterized by WWARP and the condition that pairwise choices are acyclic) the CTC model performs very well in explaining our experimental data. Of course, any other model characterized by WWARP would perform just as well as the permissive CTC model, notably the recent 'Rationalization' model proposed by Cherepanov, Feddersen and Sandroni [9]. This shows that sometimes we might need even richer sets of choice data, or non-choice data, to discern which model the decision makers are actually following. Fortunately, however, we shall be able to distinguish by choice data alone the restrictive versions of the rationalization and the CTC model.

Before summarizing in more detail our experimental findings, let us dwell on the behavioral properties we test. If you pick option A over option B, option B over option C and option C over option A, you have exhibited 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.* If a choice behavior is not consistent with WARP, it either exhibits pairwise cycles or menu dependence (or both). This is a useful classification of 'irrationality', because it zeroes in on two very different aspects of it: one involving only binary comparisons, and one involving the ability to use binary comparisons to make higher order choices form larger sets.

In the experiment we test the violations of these two elementary properties, as well as the axioms characterizing alternative decision making procedures. We use as choice alternatives time sequences of monetary rewards. We elicit two entire choice functions from each subject, over the domain of all subsets of a grand set of four alternatives. This allows us to have a much richer and informative dataset than the more frequent method of eliciting only the binary choices over alternatives, which could never detect menu dependence. Our data show that *WARP is violated by a majority of subjects.* Is this due prevalently to pairwise cyclical choices or to menu dependence? In our context, more than 15% pairwise cyclical choices were observed. But, interestingly, these violations of full rationality were strongly associated with menu effects. Menu effects are largely responsible for failures of WARP. The consequence of this fact is that in this case any procedure that fails to account for menu effects will not make a significant improvement of the standard maximization model on the basis of our data.

Models which are characterized by WWARP, such as the CTC model and the Rationalization model, yield a step change in explanatory power in the present case. Indeed, the *large majority of subjects satisfies WWARP* in all their choices. Note well: this is not simply a consequence of the fact that WWARP, being a weaker axiom, is able to bring in most of the observed choices. We demonstrate this by using *Selten's Measure of Predictive Success*, which takes into account not only the 'hit rate' of a model (proportion of correctly predicted observations), but also its 'parsimony' (inversely measured by the proportion of possible cases in principle compatible with the model). Using our data, the CTC model scores higher than Full Rationality also in terms of Selten's Measure. This is true in both its permissive version (which allows pairwise cycles) and in its restricted version (which forbids pairwise cycles and demands the preference relation to be transitive).

We present the theoretical results in section 2, while the experiment is discussed in section 3. Section 4 concludes.

2 Theory

2.1 Rational choice

Let X be a finite set of alternatives. A choice set a subset of X, and the domain of choice is a collection of nonempty choice sets $\Sigma \subset 2^X \backslash \varnothing$. The decision maker's choice behavior is summarised in his choice function γ on Σ which associates to each set a single alternative from that set, i.e. $\gamma : \Sigma \to X$ and $\gamma(S) \in S$ for all $S \in \Sigma$. We assume that all pairs of alternatives are included in the domain, that is: for all distinct $x, y \in X$, $\{x, y\} \in \Sigma$.

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:

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

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,[3] in the present context the fully rational choice functions are exactly those that satisfy Samuelson's [35] Weak Axiom of Revealed Preference (WARP), defined below:

[3]See e.g. Moulin [29] or Suzumura [39]

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

In wards, WARP states that the direct revealed preference relation is asymmetric relation: if alternative x is chosen in some set when y is available, then it can never be the case that in some other set y is chosen in the presence of x.

One might fear that failures of rationality might be compatible with all sorts of choice patterns. In fact, as we showed elsewhere, failures of WARP decompose in just two very basic patterns, menu dependence and pairwise inconsistency. The latter category involves exclusively choices between *pairs* of alternatives, while the former category involves choices from larger sets. Let $\succ_\gamma$ 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 $\{x_1, ..., x_n\}$ is a is a *base cycle of* γ if $x_i \succ_\gamma x_{i+1}$ for all $i = 1, ..., n-1$ and $x_1 = x_n$. Consider the following properties:

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

Pairwise consistency: There is no base cycle of γ.

Then WARP is decomposed into Condorcet and Pairwise consistency, which we shall use in the experiment:

Proposition 1 *(Manzini and Mariotti [22])Let all triples $\{x, y, z\}$ be in Σ. Then a choice function on Σ satisfies WARP if and only if satisfies both Condorcet consistency and Pairwise consistency.*

2.2 Two stage choice procedures

We consider a number of boundedly rational choice procedure models that have been recently proposed in the literature: common to all is that they postulate that a decision maker arrives at his final choice in two stages, by first focussing on a subset of the initial choice set, and then selecting his choice from this subset.

The mechanism through which the decision maker simplifies the original problem is what distinguishes the models we consider. There are two main classes of mechanism at work: those relying on the maximisation of binary relations, and those relying on some coarser filter to reduce the original choice set. We begin by focusing on the former.

2.3 Binary relations in the first stage

Three main models fall in this category: Rational Shortlist Methods (henceforth RSM), Categorise then Choose (henceforth CTC),[4] and Rationalization.

The RSM is the simplest version of Sequentially Rationalizable Choice (introduced in Manzini and Mariotti [21]). The latter notion postulates that, in contrast to standard rational agents who maximise a single binary relation, the individual selects an alternative from a given choice set by applying sequentially a number of binary relations. This tallies with the psychologists emphasis on sequential heuristics, as opposed to one single relation, to explain choices:[5] decision makers apply sequentially a sequence of 'criteria' which they deem relevant for their choice. The simplest weakening of standard rationality is a sequentially rationalizable choice procedure with two binary relations, RSM. Formally:

Definition 2 *A choice function γ is a* ***Rational Shortlist Method (RSM)*** *if and only if there exists an ordered pair $(\succ_1, \succ_2)$ of asymmetric binary relations (rationales) such that:*

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

In that case $(\succ_1, \succ_2)$ are said to sequentially rationalize γ.

So the choice from each S can be represented as if the decision maker went through two sequential rounds of elimination of alternatives. In the first round he makes a 'shortlist' by retaining only the elements which are maximal according to rationale $\succ_1$. In the second round, he retains only the element which is maximal according to rationale $\succ_2$: that element is his choice. RSM's are characterized on the domain we are considering by two properties, Weak WARP (WWARP) which we introduced in [21], and an expansion axiom:

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 says that if e.g. you choose steamed salmon over steak tartare when they 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

[4] See Manzini and Mariotti [22].

[5] Notable in this respect are Tversky's [41] Elimination by Aspects procedure and Gigerenzer and Todd's (e.g.[14]) idea of 'fast and frugal heuristics'.

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.[6]

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 $x = \gamma(\cup_i S_i)$.

Expansion says that if steamed salmon is chosen in each of a series of menus, then it is also chosen when all the menus are merged.

The following characterization result follows easily by adapting the argument in the proof of Theorem 1 in [21].[7]

Corollary 1 *Suppose the domain Σ is closed under set union. A choice function on Σ is an RSM if and only if it satisfies WWARP and Expansion.*

The notion of RSM can be further relaxed if we allow the decision maker in the first stage to compare *sets* of alternatives: that is, the decision maker goes through some initial 'categorization' stage. This appears very natural as soon as the set of alternatives exhibits some complexity. The categorization stage is helpful in simplifying the decision process because it allows the decision maker to ignore entire categories which look inferior (in the restaurant example of the introduction, the category 'Italian restaurants' trumps all other restaurants). Formally, we assume that categories can be (partially) compared:

Definition 3 *A **shading relation** on X is an asymmetric (possibly incomplete) relation $\succ$ on $2^X \backslash \varnothing$.*

In words, when $R \succ S$, category S is ignored if R is available: R 'shades' S. The corresponding maximal set can be defined in the obvious way,[8]. Call any asymmetric and complete binary relation on X a *preference.* Then Manzini and Mariotti [22] introduce:

[6] 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.

[7] The proof in that paper does not apply word for word here because of the more general domain we are considering.

[8]

Definition 4 *Given a shading relation $\succ$ and $S \in \Sigma$, the $\succ -$maximal set on S is given by:*

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

Definition 5 *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:*

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

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

Perhaps surprisingly, CTC choice functions are characterized by the single property Weak WARP (WWARP):

Theorem 1 *(Manzini and Mariotti [22])A choice function* γ *is CTC if and only if it satisfies WWARP.*

The definition of CTC can be strengthened by requiring that the transitivity of the preference applied in the second stage - this generates a **Transitive CTC (TCTC)**, i.e. a CTC rationalized by a linear order (i.e. an asymmetric, transitive and complete binary relation) $\succ^*$.

Theorem 2 *A choice function* γ *is a Transitive CTC if and only if it satisfies WWARP and Pairwise consistency.*

Proof: For the first part of the statement, in view of theorem 1 it suffices to recall that the preference $\succ^*$ must coincide with the base relation $\succ_\gamma$, and that (as is easily checked) given the assumptions on the domain implying the completeness of $\succ_\gamma$, if $\succ_\gamma$ is acyclic it must also be transitive.

For the second part of the statement, repeat the construction of $\succ$ in the proof of theorem 1. Obviously this construction still rationalizes the choice. Suppose that $R_1 \succ R_2 \succ ... \succ R_k$. Then for each R_{i-1}, R_i, R_{i+1}, there exists $S_i, S_{i+1} \in \Sigma$ such that $R_{i-1} = Lo_\gamma(\gamma(S_i), S)$, $R_i = Up_\gamma(\gamma(S_i), S_i) = Lo_\gamma(\gamma(S_{i+1}), S_{i+1})$ and $R_{i+1} = Up_\gamma(\gamma(S_{i+1}), S_{i+1})$ (notation as before). So for some $y \in S_{i+1}$ we have $y \succ_\gamma \gamma(S_{i+1}) \succ_\gamma \gamma(S_i)$. Then if $\succ$ were cyclic we would also have a base cycle, in violation of Pairwise consistency. ■

When behavior is represented by a TCTC choice function, the only possible anomalies are 'failures of aggregation' caused by menu dependence: the binary revealed preferences is perfectly rational, but the decision maker fails to it them to determine choices from larger sets. At the experimental level, this fact highlights how in order to test the TCTC procedure it is indispensable to elicit the entire choice function, beyond binary preferences.

The last model we consider is introduced in very recent work by Cherepanov, Feddersen and Sandroni [9]. They study another plausible procedure, where now a number of possible 'motivations' are considered simultaneously in the first stage: he only retains those alternatives that come top of each motivation, and he maximises preferences over such surviving alternatives in the second stage:

Definition 6 *(Cherepanov, Feddersen and Sandroni [9]) A choice function γ is **Rationalized** if and only if there exist a set of asymmetric and transitive relations (rationales) $\{\succ_1, \succ_2, ..., \succ_K\}$ and an asymmetric relation (preference) $\succ^*$ such that:*

$$\begin{aligned} \textit{For all } S &\in \Sigma\textit{: there exists } \succ_i \textit{ such that } x = \gamma(S) \succ_i y \textit{ for all } y \in S\backslash\{x\}, \\ \textit{and } x &\succ {}^*y \textit{ for all } y \in S\backslash\{x\} \textit{ for which there exists } \succ_i \textit{ such that } y \succ_i z \\ \textit{for all } z &\in S\backslash\{y\}. \end{aligned}$$

Interestingly, this procedure generates (on the full domain) exactly the same choice data as a CTC choice function:

Corollary 2 *Let $\Sigma = 2^X \backslash \varnothing$. A choice function is Rationalized if and only if it is CTC.*

Proof: The result follows immediately from theorem 1 in Cherepanov, Feddersen and Sandroni [9], stating that a choice function is rationalized if and only if it satisfies WWARP, and from theorem 1 above. ■

In other words, CTC choice functions and rationalized choice functions cannot be distinguished on the basis of choice data alone. Nevertheless, it turns out that the two models *do* have different testable implications in their more restrictive versions, where the preference relation is required to be transitive.

Define the *nested revealed preference relation* R^N_γ by $xR^N_\gamma y$ iff there exist $S, T \in \Sigma$ with $x, y \in S \subset T$, $x = \gamma(S)$ and $y = \gamma(T)$.

Theorem 3 *(Cherepanov, Feddersen and Sandroni [9]). A choice function is Rationalized with a transitive preference (i.e. it is **Order Rationalized**) if and only if the nested revealed preference relation is acyclic.*

It follows that the restrictive versions of Rationalization and CTC can be separated, at least in principle, by the available choice evidence.

We conclude this section by highlighting some additional relations between the models presented so far. First of all, theorem 1 and corollary 1 imply that the RSM model is strictly nested in the CTC model, in the sense that any choice function which is an RSM can also be seen as being CTC, but not viceversa.

Corollary 3 *Every RSM is also a CTC choice function, and there exist CTC choice functions which are not RSMs.*

Secondly

Corollary 4 *Every TCTC choice function is also rationalized with a transitive preference, but there exist choice functions rationalized with a transitive preference that are not TCTC.*

Proof: We first show that Pairwise consistency and WWARP imply the acyclicity of R_{γ}^{N}. Suppose that there exists a $R_{\gamma}^{N}-$cycle. That is, there exist x_i, $i = 1...n$ such that $x_i R_{\gamma}^{N} x_{i+1}$ for all $i = 1, 2..., n-1$ and $x_n R_{\gamma}^{N} x_1$. So there exist pairs of sets (R_i, S_i), $i = 1...n$ such that $x_i, x_{i+1} \in R_i \subset S_i$, $x_i = \gamma(R_i)$, $x_{i+1} = \gamma(S_i)$ for $i = 1, ..., n-1$, and $x_n = \gamma(R_i)$, $x_1 = \gamma(S_i)$. If WWARP is violated we are done, so suppose it holds. Then it must be that $x_i = \gamma(\{x_i, x_{i+1}\})$ for all $i = 1, ..., n-1$ and $x_n = \gamma(\{x_1, x_n\})$, a violation of Pairwise consistency.

Next, we provide an example that satisfies acyclicity of R_{γ}^{N} (and WWARP) but fails Pairwise consistency. This is simply the basic 3-cycle, with $X = \{x, y, z\}$ and $x = \gamma(\{x, y\}) = \gamma(\{x, y, z\})$, $y = \gamma(\{y, z\})$ and $z = \gamma(\{x, z\})$. Given the pairwise choices, the acyclicity of R_{γ}^{N} implies only that $x R_{\gamma}^{N} y$ (examples involving more alternatives are also easy to find). ■

3 Experiment

3.1 Experimental Design

Our experiment consists in eliciting the choice function over a set of alternatives. The task is straightforward:[9] pick the one you prefer among a set of alternative remuneration

[9]The experiment was carried out at the Computable and Experimental Economics Laboratory at the University of Trento, in Italy. We ran a total of 13 sessions. Participants were recruited through bulletin board advertising from the student population. Male and female subjects took part in each experimental session in roughly equal proportions. The experiment was computerised, and each participant was seated at an individual computer station, using separators so that subjects could not see the choices made by other participants. Experimental sessions lasted an average of around 26 minutes, of which an average of 18 minutes of effective play, with the shortest session lasting approximately 16 minutes and the longest around 37 minutes. At the beginning of the experiment subjects read instructions on their monitor, while

plans in installments to be received staggered over a time horizon of nine months, each consisting of €48 overall. We consider two treatments, one where payment to subjects consists of a €5 showup fee only (a total 56 subjects in 4 sessions), and one with payments based on actual choice (a total of 102 subjects in 9 sessions).[10] We will refer to these two treatments as HYP (for hypothetical) and PAY (for paid), respectively.[11] More precisely, in the case of the PAY treatment it was explained that at the end of the experiment one screen would be selected at random, and the preferred plan for that screen would be delivered to the subjects.[12]

Unlike the majority of choice experiments in the literature, we elicit the entire choice functions with domain over *all* subsets for each of the two grand sets. This enables us to check whether or not the axioms discussed in section 2 hold. In particular, we can assess (i) what the main reason is for the failure of full rationality (violation of Pairwise Consistency or violation of Condorcet consistency), and (ii) the predictive success of the various models discussed in the theoretical section.

Each of the two grand sets, which differed in the number of installments, consisted of four plans each, namely an increasing (I), a decreasing (D), a constant (K) and a jump (J) series of payments, over either two or three installments, as shown below. Though in both cases payments extended over nine months, because of the different number of installments we abuse terminology and refer to 'two-*period*' and 'three-*period*' sequences rather than two/three-*installment* sequences:

Two period sequences				
	I2	D2	K2	J2
in three months	16	32	24	8
in nine months	32	16	24	40

Three period sequences				
I3	D3	K3	J3	
8	24	16	8	in three months
16	16	16	8	in six months
24	8	16	32	in nine months

Table 1: the base remuneration plans

The use of payment plans as alternatives stems from our desire to present the ex-

an experimenter read the instructions aloud to the participants (see the appendix for the translation of the original instructions). Instructions were the same in both treatments, bar for one sentence, which in the HYP treatment clarified that choices were purely hypothetical, so that the only payment to be received would be the show up fee.

[10] The show up fee alone, for an average of less than thirty minute long experimental session, was higher than the hourly pay on campus, which was €8. At the time of the experiments the exchange rate of the Euro was approximately €1=$1.2=£0.7.

[11] Distinguishing by treatment, sessions lasted an average of about 28 minutes for the PAY treatment, of which an average of just above 19 minutes of effective play; and an average of around 22 minutes for the HYP treatment, of which an average of about 16 minutes of effective play.

[12] The experimental lab has a long tradition, so there was no issue of (mis)trust in receiving delayed payments. All subjects have been paid.

perimental subjects with alternatives of a ‘richer’ description than mere money amounts (sequences can be classified by their ‘shapes’), but which at the same time are objectively expressed in money terms, so as to reduce difficult to control emotive connotations (which might happen, for example, by using food items).

Figure 3 displays two distinct sample screenshots of the choices subject had to make. The participants made their choice by clicking with their mouse on the button corresponding to the preferred remuneration plan. Once made, each choice had to be confirmed, so as to minimize the possibility of errors. Both the order in which the questions appeared on screen and the position of each option on the screen was randomized.

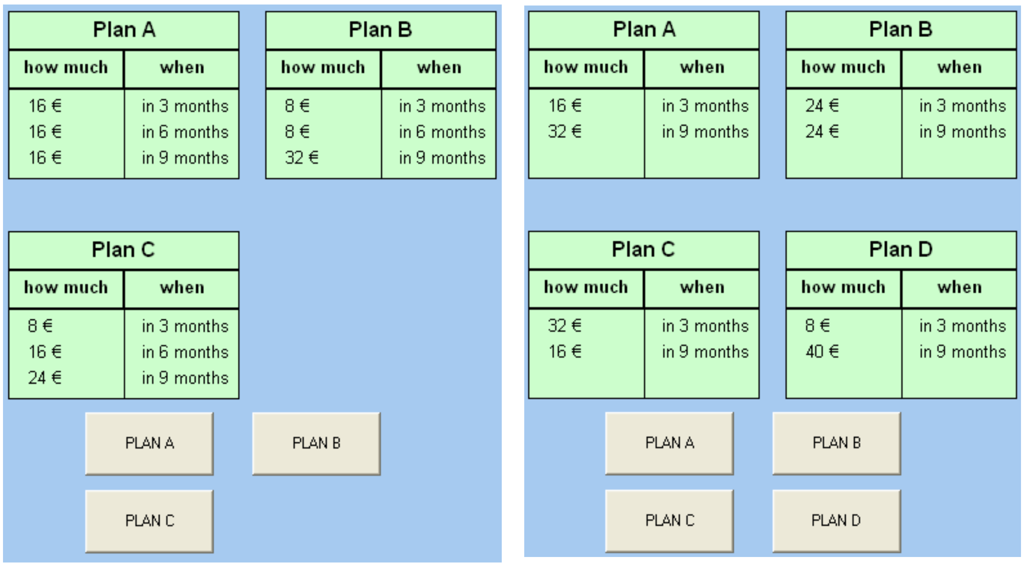

Figure 3: Sample screenshots

3.1.1 Two remarks on the experimental design

Why two treatments (PAY and HYP)? As we have discussed at length, one of the issues up for testing is consistency in choice. As observed by Johansson-Stenman and Svedsäter [18], ‘people seem to prefer to do what they say they would do’. In our context, this might imply that subjects who have revealed the preference of a over b, say, in one choice set, might favour replicating this preference in other choice sets only to avoid cognitive dissonance.[13] If experimental subjects worry about consistency in choice, it is a possibility that the administration of purely hypothetical questions might result in fewer

[13]The notion of cognitive dissonance was introduced in Festinger [12], and refers to the discomfort experience by a subject when he is made aware that he is holding two conflicting beliefs, or that he has produced contrasting choices.

violations of WARP when compared to a situation with real, incentive compatible choices. For this reason we ran a session with real payments, and one with only hypothetical ones. As we will see, although we do find that inconsistencies are uniformly less numerous in the HYP than in the PAY treatment, even in the case of purely hypothetical questions Full Rationality fails, underlying the need for an alternative explanation of observed choice patterns.

Why two types of sequences (two and three periods)? In order to put theories to the test, we need to have as many choices as possible for each subject. The minimum cardinality of the grand set which allows for violations of either WWARP or R_γ^N acyclicity to be observed is four. Teasing out the complete choice function from each subject requires eleven questions to be asked. On the other hand, increasing the size of the grand set by just one extra element would require an additional twenty questions to be asked (as there would be 31 non empty subsets in the full domain) in a rather repetitive task. For this reason we opted for two choice functions based on a grand set of four alternatives each, that would require only 22 questions to be asked to each subject, with the important additional bonus of more variety in the display. Then, we consider a subject's choice function as satisfying a given axiom if and only if *both* of his choice functions satisfy the axiom.[14]

3.2 Experimental results: Evaluating the models

3.2.1 Choosing at Random?

We begin by noting that, whatever the subjects are doing, they are certainly not pressing buttons at random.[15] Since we are eliciting the entire choice functions from universal sets with four alternatives, with a uniform probability distribution on each choice set, the probability of observing even only two subjects with the same choice is effectively zero for all practical purposes. In fact, as there are a possible $2^6 \times 3^4 \times 4 = 20,736$ choice

[14]It could be argued that a subject might satisfy WARP, say, because both of his choice functions do, but at the same time it may be that they do so in very different ways, for instance with $D2 \succ K2 \succ J2 \succ I2$ as the linear order determining choices in the two period set, and $K3 \succ D3 \succ I3 \succ J3$ in the three period set. Indeed, it should be observed that, in spite of our own labeling, these are all distinct alternatives, and there is no reason in principle why a subject should activate the same categorizations or any other heuristic when making choices in the two domains. The analysis of the type of time preferences compatible with our data is available in a separate paper (Manzini, Mariotti and Mittone [23]).

[15]Purely random choice is an important benchmark. Within consumer choice, the idea was first advanced by Becker [4] and it is used for example as the alternative hypothesis in the popular Bronars [6] index of power for nonparametric revealed preference tests. See Andreoni and Harbaugh [2] for a recent discussion of this issue.

functions on each universal set, the probability of any given choice function being picked by two subjects is $(20,736)^{-2} = 2.3257 \times 10^{-9}$.

On the contrary for both treatments and for both universal sets $X2$ and $X3$ we find almost half of the subjects with the same modal choice function. For illustration we report the frequency distributions of the observed choice functions only graphically in Figure 4 (we omit labels for legibility).[16]

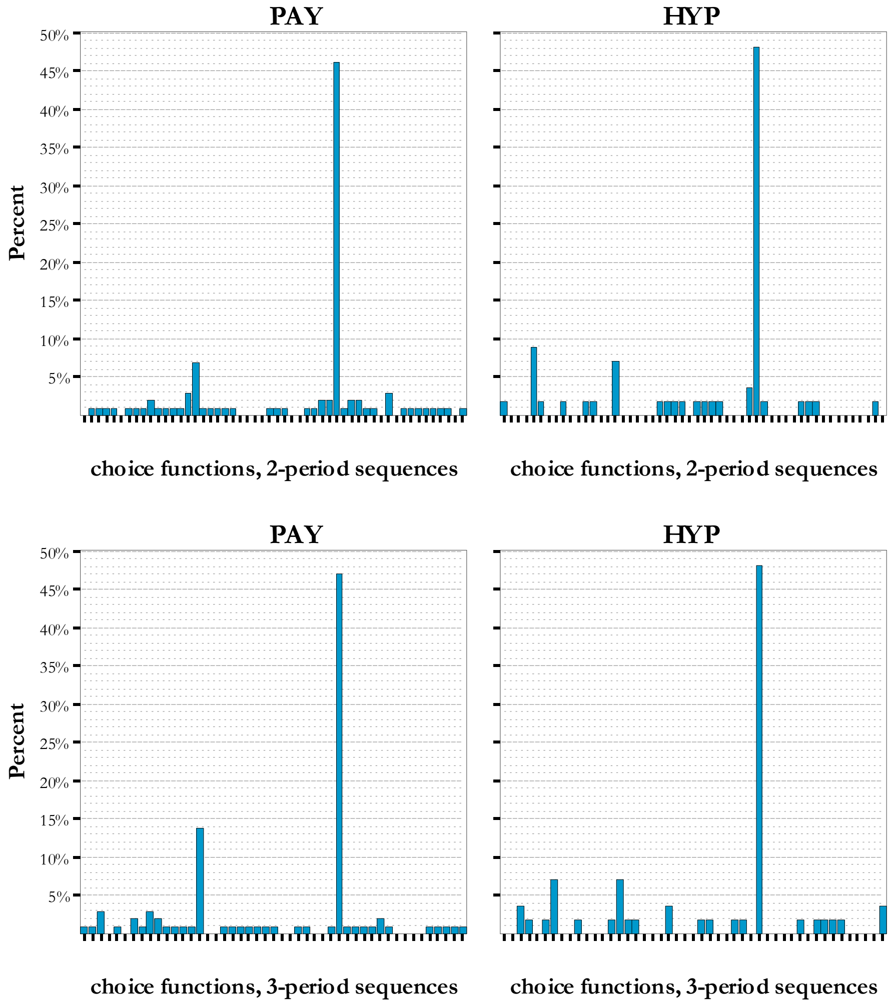

Figure 4: Frequency distributions of choice profiles by treatment and sequence length

[16]The corresponding data are reported in tables 2 and 3 in the technical appendix available online at http://webspace.qmul.ac.uk/pmanzini/twosimiltechnical_and_data_appendix.pdf.

The modal choice pattern is with the decreasing sequence preferred to the constant, which is preferred to the increasing, which is preferred to the jump sequence, and the choice each subset maximising the preference relation. Just below half of the subjects exhibit this modal choice patters (around 46% of subjects in the PAY treatment and 48.2% in the HYP treatment, irrespective of sequence length).

3.2.2 Which Type of Rationality Failure?

Before turning to the evaluation of the models,[17] we check which, if any, of the two distinct elementary failures of full rationality highlighted in Proposition 1 in section 2.1 is more prevalent. To this effect we begin by reporting aggregate data for the *violations* of Pairwise and Condorcet Consistency by experimental subject:[18]

	PAY		HYP	
	#	%	#	%
Condorcet Consistency	51	*50*	22	*39.3*
Pairwise Consistency	17	*16.7*	4	*7.1*

Table 2: Overall violations of PC and CC.

From table 2 it emerges that failures of Condorcet Consistency are substantially more frequent than violations of Pairwise Consistency. This difference is statistically significant, regardless of treatment. In fact, the McNemar test of the hypothesis that the proportions of subjects violating Condorcet Consistency is the same as the proportion of subjects violating Pairwise Consistency yields exact p-values of 0.009 in the case of the PAY treatment, and of 0.001 in the case of the HYP treatment. If we then look at the differences in the proportion of violations of each of the two axioms *across* treatments, the fall in the proportion of violations when moving from the PAY to the HYP treatment is not statistically significant: Fisher test's exact mid-p values are 0.110 for Condorcet Consistency and 0.470 for Pairwise Consistency.

In short, then, as conjectured above, it is the case that in the HYP treatment violations are less than in the PAY treatment; nevertheless, they are sizeable even when subjects are arguably more concerned with being consistent. As we will see below, this pattern is confirmed in all other tests, that is although violations are consistently higher in the PAY treatment, they are substantial in the HYP treatment, too.

[17] All the exact statistical analysis has been carried out usting *StatXact*, v.7. For a comprehensive treatment of exact and other methods in categorical data analysis see Agresti [1].

[18] Since we are mainly interested in the decisions of individual subjects over *all* their choices, throughout the paper we report data of axiom violations *by experimental subject*. The reader interested in the breakdown by choice functions separately (i.e. the choice functions from 2^{X2} and from 2^{X3} for each treatment) can consult the extended technical and data appendix available online at http://webspace.qmul.ac.uk/pmanzini/twosimiltechnical_and_data_appendix.pdf.

3.2.3 How Well Do the Models Describe the Data?

Next, we turn to the models examined in section 2.3, and we study the violations of the axioms which characterize those models.[19] It is convenient to recall in a table the set of axioms characterizing each model

	Axioms				
	WARP	Weak WARP	Pairwise Cons.	Expansion	R_γ^N acyclicity
Full Rationality	✓				
Rational Shortlist Method		✓		✓	
Categorize Then Choose		✓			
Rationalization		✓			
Transitive CTC		✓	✓		
Order Rationalization					✓

Table 3: Theories and their caracterization.

Table 4 reports violations of the remaining axioms. As before, the proportion of violators falls when moving from the PAY to the HYP treatment. Of these differences, those concerning WARP and Weak WARP are statistically significant, while for Expansion this is not the case (Fisher test's exact mid-p values are 0.110 for Expansion, 0.042 for WARP and 0.022 for Weak WARP). Table 4 also suggests similar rates of violation for Expansion and WARP (50% and 52.9%), and substantially lower rates for Weak WARP compared to either of the other axioms (28.4%). Within treatment, however, the only meaningful comparison in the difference of proportions is between failures of Expansion and Weak WARP, which are the only two independent axioms.[20] Here the hypothesis of equality in the proportion of subjects violating the two axiom is rejected (McNemar's exact p-value is 0.002 in the PAY treatment and 0.041 in the HYP treatment).

Table 4 confirms that WARP, i.e. the full rationality model, does not describe the data well, especially in the PAY treatment where *less than half* of the subjects fit the model.

[19] In our experiment we use a small universal set of alternatives. Evidence for choice from budget sets includes Fevrier and Visser [13], Mattei [28] and especially Sippel [38], who find substantial violations of the Generalized Axiom of Revealed Preferences in choices out of budget sets. However, Andreoni and Harbaugh [2] argue that most of these violations are 'small' on the basis of Afriat's efficiency index. Indeed Harbaugh, Krause and Berry [16] and especially Andreoni and Miller [3] find that subjects have choices consistent with GARP in experiments with budget sets.

[20] For comparisons between the proportion of violations of other pairs of axioms it is not possible to rely on a McNemar test, as violations of either Expansion or Weak WARP imply violations of WARP (i.e. the relevant contingency table would have structural zeroes). We defer tackling of this issue to our discussion of the relative performance of alternative theories further below.

	PAY #	PAY %	HYP #	HYP %
Expansion	51	*50*	22	*39.3*
Weak WARP	29	*28.4*	8	*14.3*
WARP	54	*52.9*	22	*39.3*
R^N_γ acyclicity	29	*28.4*	8	*14.3*

Table 4: Overall axiom violations.

Consider now RSMs. The crosstabulation of violations of the two axioms characterizing it is reported in table 5. Interestingly, in both treatments, no individual who satisfies Expansion violates Weak WARP. That is, the (large) number of Expansion violators is not joined by another separate group of Weak WARP violators in order to determine the RSM violators. The RSM violators are simply counted by Expansion violators (of which some are also Weak WARP violators). So although Weak WARP and Expansion are logically independent properties, in our experimental sample they are not statistically independent. This also shows that, quite remarkably, although RSM is a much weaker notion than full rationality, in our experimental sample it does not do a substantially better job than the fully rational model at explaining the data.

		PAY Expansion × #	PAY Expansion × %	PAY Expansion ✓ #	PAY Expansion ✓ %	HYP Expansion × #	HYP Expansion × %	HYP Expansion ✓ #	HYP Expansion ✓ %		
Weak WARP	×	29	*28.4*	0	*0*	8	*14.3*	0	*0*	×	Weak WARP
	✓	22	*21.6*	51	*50*	14	*25*	34	*60.7*	✓	

Table 5: Violations of Weak WARP and Expansion

RSMs improve only marginally on order maximization in their ability to explain the data for the PAY treatment (decreasing the violations from 52.9% in the case of WARP to 50% in the case of RSM), and they are as bad in the HYP treatment. This is due to the fact that binary cycles, which violate full rationality but not RSM, are not a very relevant phenomenon here, unlike menu effects (in the sense of violations of Condorcet Consistency), which cannot accommodated by either of these two theories.

Turning now to Transitive CTCs, the crosstabulation of violations of the two axioms characterizing is in table 5.

		PAY Pairwise Consistency				HYP Pairwise Consistency					
		×		✓		×		✓			
		#	%	#	%	#	%	#	%		
Weak WARP	×	11	*10.8*	18	*17.6*	2	*3.6*	6	*10.7*	×	Weak WARP
	✓	6	*5.9*	67	*65.7*	2	*3.6*	46	*82.1*	✓	

Table 6: Violations of Weak WARP and Pairwise Consistency

Transitive CTCs provide a considerable improvement on RSMs in terms of accommodating the choice of a substantial majority of subjects in both treatments, for the same reason as RSMs do not improve much on standard order maximization, namely the paucity of binary cycles observed, and the abundance of menu dependent choice.

Finally, we turn to Weak WARP and the models it characterizes. From Table 4 we can see that Weak WARP is satisfied by just below 72% of the subjects in the PAY treatment and just below 86% of the subjects in the HYP treatment.

In summary then:

	PAY		HYP	
	#	%	#	%
Full rationality	48	*47.1*	34	*60.7*
Rational Shortlist Method	51	*50*	34	*60.7*
Categorize then choose/Rationalization	73	*71.6*	48	*85.7*
Transitive CTC	67	*65.7*	46	*82.1*
Order Rationalization	73	*71.6*	48	*85.7*

Table 7: Explanatory power of competing theories.

The three models are nested, that is

Full Rationality⇒RSM⇒CTC/Rationalization⇐Order Rationalization

Full Rationality⇒Transitive CTC⇒CTC/Rationalization⇐Order Rationalization

In order to compare the incremental 'explanatory' power in each more general theory in a meaningful way, we have to take into account this nestedness. We do so in the next section.

3.2.4 Comparing Theories with Selten's Measure of Predictive Success

In order to compare the models we use Selten's Measure of Predictive Success (Selten [37]). This measure was specifically designed to evaluate 'area theories' like the ones considered in this paper, namely theories that exclude deterministically a subset of the possible outcomes. The measure takes into account not only the 'descriptive power' of the model (measured by the proportion of 'hits', the observed outcomes consistent with the model), but also its 'parsimony'. The lower the proportion of theoretically possible outcomes consistent with the model, the more parsimonious the model. In our specific case, one possible criticism of our 'revealed preference' type of tests might be that the experiment does not have enough power to reject Weak WARP even if it happened to be the wrong hypothesis. For example, if each universal set consisted of only three, instead than of four, alternatives, Weak WARP could never be violated. But the observed 100% hit rate should be interpreted as a failure of the experimental design rather than a validation of the model of RSM by categorization. By introducing the 'parsimony' element, Selten's measure would pick up this problem. More precisely, the measure, denoted s, is expressed as

$$s = r - a$$

where r is the descriptive power (number of actually observed outcomes compatible with the model divided by the number of possible outcomes) and a is the 'relative area' of the model, namely the number of outcomes in principle compatible with the model divided by the number of all possible outcomes. In the hypothetical example of a universal set of three alternatives, $s = 0$ for the model of RSM by categorization

In our experiment, we observed two choice functions for each subject. So the number of all logically possible observations of choice behavior for each subject was $(20,736)^2$. A 'hit' consists of the subject not violating the axioms (characterizing a specific model) in either of the two choice functions. Thus the values of r for the various models are in Table 7. In order to compute a for a model we have to compute the proportion of choice functions compatible with the set of axioms characterizing that model. We begin with Full Rationality. For each universal set of alternatives, the number of choice functions satisfying WARP is simply the number of strict orderings on a set of cardinality four, that is $4! = 24$. Therefore for each individual there are only 24^2 possible patterns of behavior in the experiment compatible with the Full Rationality model. The area of the

Full Rationality model is:

$$a^{FR} = \left(\frac{24}{20,736}\right)^2 \approx 0$$

The Full Rationality model is thus a beautifully parsimonious model whose Selten's Measure of Predictive Success is entirely determined by its descriptive power. From Table 7 above we thus have, for the PAY and HYP treatments, the following values of Selten's Measure for this model:

$$\begin{aligned} s^{FR}_{PAY} &\approx 0.471 \\ s^{FR}_{HYP} &\approx 0.607 \end{aligned}$$

The RSM model, as we have seen, does not improve significantly in this experiment on the Full Rationality model even in terms of sheer descriptive power. So a fortiori it is not an interesting competitor to the latter in terms of overall predictive success. Therefore we move to the computation of the values for the CTC model.

We need to compute first the number of choice functions that satisfy Weak WARP. This can be done by considering four possible exhaustive and mutually exclusive configurations, depending on whether the choice from the grand set, $\gamma(X)$, is selected in pairwise choice over *all three* other alternatives (*Configuration A*), over exactly *two* other alternatives (*Configuration B*), over exactly *one* other alternative (*Configuration C*), or over *none* of the other alternatives (*Configuration D*).

How many possible cases of Configuration A are there in which WWARP is satisfied? The choices for all sets of cardinality three which include $\gamma(S)$ are forced, as $\gamma(S)$ must be chosen from them in order not to violate WWARP. So, the four possible choices of $\gamma(S)$ can be combined with the three possible choices from the set of cardinality three not including $\gamma(S)$, and with the 2^3 possible choice combinations from the three binary sets that do not include $\gamma(S)$. All in all, we have $4 \times 3 \times 2^3 = 96$ possible choice functions compatible with WWARP in this configuration.

For Configuration B, we have first of all that the four possible choices for $\gamma(S)$ can be combined with three possible choices for the alternative which is chosen over $\gamma(S)$ in binary comparison. For each of these combinations, there are $3^2 \times 2^2$ choice combinations from the sets of cardinality three (since only two choices compatible with WWARP can be made from the two sets of cardinality three that include $\gamma(S)$ and the alternative that 'beats' it in binary comparison), together with the 2^3 choice combinations from the binary sets that do not include $\gamma(S)$. All in all, we have $4 \times 3 \times (3 \times 2^2) \times 2^3 = 1,152$ possible choice functions compatible with WWARP in this configuration.

Reasoning along similar lines leads to the count of $4\times3\times(3^2\times2^2)\times2^3=5,184$ possible choice functions compatible with WWARP in configuration C, and to $4\times3^4\times2^3=2,592$ in Configuration D.

Adding up, in total there are $9,024$ possible choice functions compatible with WWARP when the grand set has cardinality four, and thus $(9,024)^2$ possible types of choice behavior compatible with WWARP in our experiment. This leads to the area value

$$a^{WWARP}=\left(\frac{9,024}{20,736}\right)^2=0.189$$

So both the CTC and the Rationalization models are as expected far less parsimonious than the Full Rationality model. However, they do improve on the full rationality model:

$$\begin{aligned} s_{PAY}^{WWARP} &= 0.716-0.189=0.527 \\ s_{HYP}^{WWARP} &= 0.857-0.189=0.668 \end{aligned}$$

We now turn to the restricted versions of these two models. Consider now Cherepanov, Feddersen and Sandroni's Order Rationalization model. In order to compute the area of this theory, recall from the proof of corollary 4 that violations of WWARP imply violations of R_γ^N acyclicity. However, there can be choice functions that satisfy WWARP but are not Order Rationalizable - that is, in order to compute the area of this theory, it suffices to subtract from the area satisfying WWARP those cases that fail R_γ^N acyclicity. Recall that xR^Ny if and only if there exist two sets T and T' with $x,y\in T\subset T'$ such that $x=\gamma(T)$ and $y=\gamma(T')$. For convenience, we will refer to T as a 'small set' in what follows. For ease of notation, let $X=\{w,x,y,z\}$ denote our grand set of four alternatives, and fix $\gamma(X)=x$. Note that $x=\gamma(X)$ cannot be part of any R_γ^N cycle. Suppose it is, and that xR^Ny. This is impossible, since then x must be chosen in a 'small set': but if $x=\gamma(xy)$ and $y=\gamma(S)$ we have a violation of WWARP (if $S\subset X$) or (if $S=X$) a contradiction; and if $x=\gamma(xyz)$, for some z then for xR_γ^Ny it must be $y=\gamma(X)$, a contradiction. So there cannot be any R_γ^N cycle of length 4. An R_γ^N cycle of length 2 is just a violation of WWARP. Thus we look for R_γ^N cycles of length 3.

There cannot be any R_γ^N cycle involving any 'small set' of cardinality 3 or more. For, a 'small set' of cardinality 3 forces uniquely the choice from X by the definition of R_γ^N. Let $x=\gamma(X)$ and yR_γ^Nw with a small set of cardinality 3. This means that $w=\gamma(X)$, contradiction. So we look for R^N_γ cycles where the small sets are all pairs. Fix $y = \gamma(ywz)$ and let the cycle order be $yR^N_\gamma wR^N_\gamma zR^N_\gamma y$. For $yR^N_\gamma w$ it must be $\gamma(xyw) = w$ and for $wR^N_\gamma z$ it must be $z = \gamma(wzx)$. These two equalities and WWARP imply that $\gamma(xw) = w$ and $\gamma(xz) = z$. Moreover, we can exclude the case $x = \gamma(xy)$, $y = \gamma(xyz)$ by WWARP. So the possible combinations are given by remaining the combinations of choices from $\{x, y\}$ and from $\{x, y, z\}$, that is either $y = \gamma(xyz) = \gamma(xy)$; or $x = \gamma(xyz) = \gamma(xy)$; or $x = \gamma(xyz)$, $y = \gamma(xy)$. These three possibilities can arise in any of the $4 \times 3 \times 2$ possible configurations (four ways to pick $\gamma(X)$, three ways to pick the choice from the remaining alternatives, and two directions of cycle. So in all 72 possibilities, yielding an area

$$a^{ORat} = \left(\frac{9,024 - 72}{20,736}\right)^2 = 0.186$$

which is only slightly lower than the area for just WWARP, so that the corresponding Selten's indices in each treatment are

$$\begin{aligned} s^{ORat}_{PAY} &= 0.716 - 0.186 = 0.53 \\ s^{ORat}_{HYP} &= 0.857 - 0.186 = 0.671 \end{aligned}$$

Finally, we turn to the TCTC model, again distinguishing by the four possible exhaustive and mutually exclusive configurations A, B, C and D, depending on whether the choice from the grand set, $\gamma(X)$, is selected in pairwise choice over *all three* other alternatives, over exactly *two* other alternatives, over exactly *one* other alternative, or over *none* of the other alternatives (as we did when computing the area for WWARP). Indeed, we do have to proceed as for WWARP, but this time making sure we eliminate pairwise cyclical choices. For Case A, there are $4 \times 3 \times 6 = 72$ cases (where the last number in the multiplication is explained by there being 2^3 pairwise choices in all in the $3-$sets, but we deduct the two cycles). For case B, in the $3-$set that does not include $\gamma(X)$, the pairwise choice is forced for two binary comparisons - otherwise a cycle obtains. This leaves only two possible binary choices, the ranking of the two alternatives beaten by $\gamma(X)$, resulting in $4 \times 3 \times 3 \times 4 \times 2 = 288$ cases. For case C, in the $3-$set that does not include $\gamma(X)$, the pairwise choice is forced for two binary comparisons - otherwise a cycle obtains. This leaves only two possible binary choices, the ranking between the two alternatives that beat $\gamma(X)$, resulting in $4 \times 3 \times 9 \times 4 \times 2 = 864$ cases. Finally, case D is similar to case A, noting that now WWARP imposes no restriction on the choice from any set of cardinality three, resulting in $4 \times 3^4 \times 6 = 1,944$ cases.

In total, then, we have $3,168$ cases, yielding an area

$$a^{TCTC} = \left(\frac{3168}{20736}\right)^2 = 0.023$$

so that

$$\begin{aligned} s_{PAY}^{TCTC} &= 0.657 - 0.023 = 0.634 \\ s_{HYP}^{TCTC} &= 0.821 - 0.023 = 0.798 \end{aligned}$$

3.2.5 Summary and Comment.

The general indication we draw from the data is that a model addressing the lack of full rationality in the choice function must be able to explain menu-effects in the form of Condorcet inconsistency. Pairwise cycles of choice appear to be less crucial. This suggests that, while it is not difficult to induce cyclical behavior in the laboratory (e.g. Roelofsma and Read [30], Tversky [40], and Waite [43]), such behavior may apply mainly to a well identified class of circumstances, where 'adjacent' alternatives in the cycle are similar in some dimension (Tversky [40], Rubinstein [31], Leland [19] all noted the importance of this aspect in decision making). The typical cycles observed result from having the subject compare multidimensional objects, say simple gambles of the form (x, p), where p is the probability of winning the amount x. If x is close to y, it is likely that the choice between (x, p) and (y, q) is dictated by the probability dimension alone. A sufficiently long chain of such choices may however eventually break the similarity in outcome, turning the outcome into the decisive criterion. The indication of our experiment is that outside of such circumstances, menu effects tend to dominate.

This indication is confirmed in the analysis of the models we have studied in this paper. Neither the full rationality nor the RSM model are compatible with menu effects of the Condorcet inconsistency type, and indeed they both fail at explaining the data. The RSM model performs only marginally better than the full rationality model. The proportion of successes in explaining behavior is not increased significantly when weakening WARP to the combination of Expansion and WWARP.

The models of Categorize Then Choose and Rationalizability are compatible with Condorcet inconsistency, and are successful in the experiment. There is a significant leap in the proportion of successes in explaining behavior when weakening WARP to WWARP. The resulting models can explain 50% more data compared to the other two models, namely over 70% in one treatment and over 85% in the other treatment.

In addition, the loss of parsimony of the models of CTC and Rationalization is much smaller than the increase in descriptive power, as evidenced by their superiority on the basis of Selten's Measure of Predictive Success.

When the CTC and the Rationalization models are strengthened, as is the case for Transitive CTC and Order Rationalization, the latter has a worse Selten's index than its more permissive version, since it does not have a better hit rate, while at the same time it has a similar area. To the contrary, the TCTC model loses a few data points, but it decreases its area substantially. This results in the best Selten's index of predictive success among all of the theories under scrutiny.

To summarize, the Selten index ranks the new models and the textbook model in the following decreasing order of success:

TCTC; Order Rationalization; WWARP theories; Full Rationality.

As a final consideration, it is difficult to provide an economic explanation of the differences between the PAY and HYP treatment. Indeed, we might have expected the opposite result. The following are arguably reasonable working hypotheses:

- rational behaviour imposes a higher cognitive load on subjects;
- higher cognitive loads require more effort to be dealt with, and
- higher effort is associated with performance related payment.

Under these assumptions, we would expect more violations in the HYP rather than in the PAY treatment. In this respect, then, our results can be seen as counterintuitive, appearing compatible with a mere propensity for consistency. As found recently in Segal [34], however, a better performance in low stakes test appears to depend positively on test taking motivation. The latter is found to be driven by personality traits, rather than cognitive skills. One has to be careful to note that, however, the decision problems that confronted our subjects did not have a 'right' or 'wrong' answer as is typical of the test used in e.g. Segal [34]. Nevertheless, she finds that female participants seem to be less negatively affected by the lack of monetary incentives than male participants. In our experiment, too, controlling for sex, the differences in the proportions of axiom violations in the HYP and the PAY treatment are less pronounced for female participants than they are for male participants.[21] This could be seen as reinforcing the evidence that motivation of female participants seems less affected by the presence of monetary incentives.

[21] Due to space limitations, we do not report here the results controlling by sex, as this is somewhat less central to our analysis. Besides sex, our data suggest a plethora of ad-

4 Concluding remarks

We hope to have shown with a concrete experiment how the standard (nonparametric) revealed preference methodology, using choice data alone, can be successfully used to test directly 'psychological' choice procedures and compare their predictive success.[22]

We are not claiming that *only* direct choice data are relevant for economics (as argued for example by Gul and Pesendorfer [15]), but that 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. As observed, for example, by Bernheim [5], revealed preference tests may have zero power against plausible alternatives. In this paper this is demonstrated, for example, by the use of WWARP alone, which fails to distinguish between the permissive versions of the Rationalization and the CTC models. Nevertheless, one should not underestimate the high significance of revealed preference tests: our data draw, for instance, a very clear-cut and a priori unexpected line between psychologically plausible models such as RSM and CTC or Rationalization.

In addition, we hope that the use and computation of Selten's Measure of Predictive Success for various 'area theories' can be of independent methodological interest, in view of the relatively rare use of this index.[23] The adoption of the index permits in our case finer discriminations between alternative theories.

Whether or not one believes the models we have studied, the experiment also establishes quite clearly two series of facts that are model-independent. First, it highlights the importance of a precise type of menu effect in choice. This fact could only be discovered because we elicited entire choice functions from the subjects, rather than merely asking them to express binary preferences (the more usual procedure).

Secondly, the experiment shows that while one step in weakening the Weak Axiom of Revealed Preference (to the combination of Weak WARP and Expansion) does not capture a significant new portion of observed behaviors, another step in this weakening

ditional considerations - due to space limitations we cannot analyze them all in this paper, and limit ourselves to highlighting just a few in a separate appendix, available online at http://webspace.qmul.ac.uk/pmanzini/twosimiltechnical_and_data_appendix.pdf.

[22] Beside those already mentioned in the paper, theoretical applications or discussions of this methodology are Rubinstein and Salant [32], Salant and Rubinstein [33], Caplin [7], Masatlioglu and Ok ([24], [25]) Masatlioglu and Nakajima [26], Masatlioglu, Nakajima and Ozbay [27], Tyson [42], among others. The paper by Masatlioglu, Nakajima and Ozbay [27] is particularly relevant, because it characterises a very general two-stage procedure in which first a 'consideration set' is somehow selected, and then a preference is maximised. Eliaz and Spiegler [11] independently study an economic application of this idea.

[23] See e.g. Hey and Lee [17] for another recent application.

(to Weak WARP alone, or Weak WARP and Pairwise consistency) is dramatically more effective.

As a consequence of the facts above, the experiment suggests what a good model has to do in order to explain deviations from utility maximization. The utility maximization model and the RSM model ([21]) unfortunately cannot explain well the data elicited in our experiment because they can't address menu effects. Both the CTC model proposed in this paper and the Rationalization model, in both their restrictive and permissive versions, instead perform definitely better (not only in terms of brute number of observations explained but also in terms of Selten's [37] Measure of Predictive Success). The Transitive CTC model has the highest Selten score. It is likely nevertheless that in situations where cycles play a more important role, the Rationalization model will score better. This an example where contextual evidence will prove to be crucial in producing finer discriminations between theories than choice data alone.

To conclude, we believe that the CTC model has three main attractions. First, its easy testability by means of choice data. Second, its ability to capture in a direct way menu effects: in this model preferences are simply revealed (uniquely) by binary choices. Third, it offers one possible solution to the hard problem of welfare analysis in the context of boundedly rational choice.

References

[1] Agresti, A. (2002) *Categorical Data Analysis*, Wiley Series in Probability and Statistics, Second Edition, John Wiley and Sons.

[2] Andreoni, J. and W. T. Harbaugh (2006) "Power Indices for Revealed Preferences Tests", mimeo.

[3] Andreoni, J. and J. Miller (2002) "Giving According to GARP: An Experimental Tests of the Consistency of Preferences for Altruism", *Econometrica*, 70: 737-753.

[4] Becker, G. S. (1962) "Irrational Behavior in Economic Theory", *Journal of Political Economy*, 70: 1-13.

[5] Bernheim, B. D. (2009) "Behavioral Welfare Economics" (2009), *Journal of the European Economic Association*, 7: 267-319.

[6] Bronars, S. G. (1987) "The Power of Nonparametric Tests of Preference Maximization", *Econometrica*, 55: 693-698.

[7] Caplin, A. (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.

[8] Chambers, C. P. and T. Hayashi (2008) "Choice and Individual Welfare", mimeo, HSS California Institute of Techology WP 1286.

[9] Cherepanov, V., T. Feddersen and A. Sandroni (2008) "Rationalization", mimeo, Nothwestern University.

[10] Ehlers, L. and Y. Sprumont "Weakened WARP and Top-Cycle Choice Rules", *Journal of Mathematical Economics*, 44: 87-94 2008.

[11] Eliaz, K. and R. Spiegler (2007) "Consideration Sets and Competitive Marketing", mimeo, University College London.

[12] Festinger, L. (1957) *A Theory of Cognitive Dissonance*, Stanford University Press.

[13] Fevrier, P. and M. Visser (2004) "A Study of Consumer Behavior Using Laboratory Data", *Experimental Economics*, 7: 93-114.

[14] Gigerenzer, G. and P. M. Todd. (1999) "Fast and Frugal Heuristics: The Adaptive Toolbox." In *Simple heuristics that make us smart*, ed. G. Gigerenzer, P. M. Todd, and the ABC Research Group, 3-34. New York: Oxford University Press.

[15] Gul, F. and W. Pesendorfer (2005) "Mindless economics", mimeo Princeton University.

[16] Harbaugh, W. T., K. Krause and T. Berry (2001) "GARP for Kids: On the Development of Rational Choice Behavior", *American Economic Review*, 91: 1539-1545.

[17] Hey, J. D. and J. Lee (2005) "Do Subjects Separate (or Are They Sophisticated)?", *Experimental Economics*, 8:233-265.

[18] Johansson-Stenman, and H. Svedsäter (2007) "Hypothetical bias in choice experiments: Within versus between sample tests",Goteborg University, Department of Economics, Working Papers in Economics: No. 252, 2007.

[19] Leland, J. W. (1994) "Generalized Similarity Judgments: An Alternative Explanation for Choice Anomalies", *Journal of Risk and Uncertainty*, 9: 151-72.

[20] Mandler, M., P. Manzini and M. Mariotti (2008) "One MiIlion Answers to Twenty Questions: Choosing by Checklist", mimeo, Queen Mary, University of London.

[21] Manzini, P. and M. Mariotti (2007) "Sequentially Rationalizable Choice", *American Economic Review*, 97: 1824-1839.

[22] Manzini, P. and M. Mariotti (2010) "Categorize Then Choose: Boundedly Rational Choice and Welfare", mimeo, University of St. Andrews.

[23] Manzini, P., M. Mariotti and L. Mittone (2007) "Choosing monetary sequences: theory and experimental evidence", CEEL Working paper 1-06.

[24] Masatlioglu, Y. and E. Ok (2005) "Rational Choice with Status-quo Bias", *Journal of Economic Theory* 121: 1-29.

[25] Masatlioglu, Y. and E. Ok (2006) "Reference-Dependent Procedural Decision Making", working paper, NYU.

[26] Masatlioglu, Y. and D. Nakajima (2008) "Choice by Iterative Search', mimeo, University of Michigan.

[27] Masatlioglu, Y., D. Nakajima and E.Y. Ozbay (2009) "Revealed Attention", mimeo, University of Michigan.

[28] Mattei, A. (2000) "Full-Scale Real Tests of Consumer Behavior using Expenditure Data", *Journal of Economic Behavior and Organization*, 43: 487-497.

[29] Moulin, H. (1985) "Choice Functions Over a Finite Set: A Summary" *Social Choice and Welfare*, 2, 147-160.

[30] Roelofsma, P. H. and D. Read, 2000 "Intransitive Intertemporal Choice", *Journal of Behavioral Decision Making*, 13: 161-177.

[31] Rubinstein, A., 1988, "Similarity and Decision Making Under Risk (Is there a utility theory resolution to the Allais Paradox?)", *Journal of Economic Theory*, 46: 145-153.

[32] Rubinstein, A. and Y. Salant (2008) "Some thoughts on the principle of revealed preference", in: *The Foundations of Positive and Normative Economics: A Handbook*, A. Caplin and Andrew Schotter (eds.), Oxford University Press.

[33] Salant, Y and A. Rubinstein (2008) "(A, f): Choice with Frames", *Review of Economic Studies*, 75: 1287-1296.

[34] Segal, C. (2008) "Motivation, test scores, and economic success", Working Paper 1124, Department of Economics, Universitat Pompeu Fabra.

[35] Samuelson, P. (1938) "A Note on the Pure Theory of Consumer's Behavior", *Economica* 5 (1): 61-71.

[36] Salant, Y. (2007) "Procedural Analysis of Choice Rules with Applications to Bounded Rationality", mimeo, Sanford Graduate School of Business.

[37] Selten, R. (1991) "Properties of a measure of predictive success", *Mathematical Social Sciences*, 21: 153-167.

[38] Sippel, R. (1997) "An Experiment on the Pure Theory of Consumer's Behavior", *Economic Journal*, 107: 1431-1444.

[39] Suzumura, K. (1983) *Rational Choice, Collective Decisions, and Social Welfare*, Cambridge University Press, Cambridge U.K.

[40] Tversky, A. (1969) "Intransitivity of Preferences", *Psychological Review*, 76: 31-48.

[41] Tversky, A. (1972) "Elimination by aspects: A theory of choice", *Psychological Review*, 79: 281–299.

[42] Tyson, C.J. "Cognitive constraints, contraction consistency, and the satisficing criterion", *Journal of Economic Theory*, 138: 51 – 70

[43] Waite, T. (2001) "Intransitive preferences in hoarding gray jays (Perisoreus canadensis)", *Behavioral Ecology and Sociobiology*, 50: 116 - 121.

A Appendix: Instructions

Please note: you are not allowed to communicate with the other participants for the entire duration of the experiment.

The instructions are the same for all you. You are taking part in an experiment to study intertemporal preferences. The project is financed by the ESRC.

Shortly you will see on your screen a series of displays. Each display contains various remuneration plans worth the same total amount of €48 each, staggered in three, six and nine months installments. For every display you will have to select the plan that you prefer, clicking on the button with the letter corresponding to the chosen plan. (HYP: These remuneration plans are purely hypothetical. At the end of the experiment you'll be given a participation fee of €5) (PAY: At the end of the experiment one of the displays will be drawn at random and your remuneration will be made according to the plan you have chosen in that display).

In order to familiarize yourself with the way the plans will be presented onscreen, we shall now give you a completely hypothetical example, based on a €7 total remuneration.

Plan A

How much	When
€3	in one year
€1	in two years
€1	in three years
€2	in four years

Plan B

How much	When
€1	in one year
€2	in two years
€3	in three years
€1	in four years

In this example plan A yields €7 in total in installments of €3, €1, €1 and €2 in a year, two years, three years and four years from now, respectively, while plan B yields €7 in total in installments of €1, €2, €3 and €1 in a year, two years, three years and four years from now, respectively.