Choice, Deferral and Consistency

Miguel Costa-Gomes, Carlos Cueva, Georgios Gerasimou and Matus Tejiscak

School of Economics and Finance Discussion Paper No. 1416
12 Dec 2014 (revised 13 Dec 2019)
JEL Classification: C91, D01, D03, D11, D12,
Keywords: Choice deferral; choice reversals; revealed preference; indecisiveness; incomplete preferences
Choice, Deferral and Consistency

Miguel Costa-Gomes‡ Carlos Cueva§ Georgios Gerasimou¶ Matúš Tejiščák∥

July 23, 2020

Abstract

We study the extent to which forcing people to make active choices has an adverse effect on choice consistency. To do so, we conduct three novel choice experiments with real goods and money lotteries where subjects in one treatment are forced to choose from every menu, while in the other treatment they are not forced to choose, and can instead incur a small cost to defer choice. With choice consistency proxied by the subjects’ degree of conformity with standard no-cycle conditions on their active choices, we find that subjects who are forced to choose behave less consistently than subjects who are not. This suggests that forced-choice designs can potentially lead researchers to overestimate the fraction of subjects that do not maximize a stable and transitive -but possibly incomplete- preference relation. We then use a new combinatorial-optimization technique -called the distance score method- that finds which model in a given collection minimizes the number of changes to a subject’s dataset for it to be perfectly consistent with that model. Using this method on our richer real-goods data, we find that approximately 67% of all non-forced choice subjects are best explained by utility maximization, while 29% by models of dominant choice with incomplete preferences and 4% by a model of overload-constrained utility maximization. Subjects on average deviate from perfect conformity with their best-matching model by a single decision.

Keywords: Choice deferral; choice reversals; revealed preference; indecisiveness/incomplete preferences; distance scores.

JEL Classification: C91, D01, D03, D11, D12
1 Introduction

In real-world consumer or election decisions people typically have the opportunity to choose none of the options they are presented with. Economic experiments on individual decision making on the other hand have traditionally required participants to choose a market alternative from the set of those that the experimenter makes available to them. A possibility that arises when choice is forced in this way is that experimental subjects may be asked to “actively” choose an alternative in situations where they would rather opt for a “choice deferral” outside option instead. Starting with Tversky and Shafir (1992), many studies in psychology have suggested that this may indeed happen when people are indecisive/unable to compare the available options, in violation of the fundamental axiom that preferences are complete.\footnote{Completeness requires that whenever an agent is presented with any two choice alternatives, either she prefers one over the other or she is indifferent between them. As such, it rules out the possibility of the agent being indecisive in the above sense. While completeness is often considered to be a rationality condition on preferences alongside transitivity, the behavioural restriction that it imposes has rendered it vulnerable to criticism on both normative and positive grounds by, among others, von Neumann and Morgenstern (1947), Savage (1954), Aumann (1962) and Gilboa, Maccheroni, Marinacci, and Schmeidler (2010).} As a case in point, Shafir, Simonson and Tversky, (1993, p. 21) cited the following real-world example:

“At the bookstore, [Thomas Schelling] was presented with two attractive encyclopedias and, finding it difficult to choose between the two, ended up buying neither – this, despite the fact that had only one encyclopaedia been available he would have happily bought it.”

One may therefore expect that when such individuals are forced to choose, they will generally be more likely to do so in a random way that does not reflect an underlying preference than those who are not forced. To the extent that this is so, a natural testable hypothesis is that such decision makers’ active choices will feature a higher degree of inconsistency. An early intuition for such an effect can be traced back to Luce and Raiffa (1957) who -possibly in response to the cyclic-preference findings reported in May (1954)- argued that “intransitivities often occur when a subject forces choices between inherently incomparable alternatives”. Despite its significance from a methodological, theoretical and empirical point of view, however, no experimental/empirical study that we are aware of has tested this hypothesis. Based on data that we collected in three incentivized experiments involving real goods and money lotteries, in this paper we provide evidence suggesting that forced choice is indeed generally more inconsistent than non-forced choice.

Intuition and experimental evidence point towards at least two main reasons why economic agents may be indecisive and hence more prone to inconsistencies when forced to choose. Sen (1997), for example, argued that indecisiveness “can arise from limited
information, or from ‘unresolved’ value conflicts’. The adverse role of such conflicts on the ability of agents to compare alternatives has also been discussed by philosophers (e.g. Levi, 1986), while evidence for them has been provided by consumer psychologists through the identification of a link between the occurrence of decision conflict among multi-attribute alternatives and the ensuing avoidant/deferring behaviour of the agents in menus that include such alternatives. As Shafir, Simonson and Tversky, (1993, p. 21) succinctly described it, “there are situations in which people (...) do not have a compelling reason for choosing among the alternatives and, as a result, defer the decision, perhaps indefinitely.” Value conflicts aside, indecisiveness is also more likely to occur in situations where the agent has inadequate information about the alternatives in question, a state that Sen (1997) referred to as “tentative” indecisiveness. Such decision makers may well be willing to acquire additional and possibly costly information about the available options before making an active choice.2

The experimental design that we propose reflects these ideas and features two treatments that are identical except that choice is forced in the control but not in the target treatment. During the main phase of the experiment subjects are sequentially presented with a collection of menus of alternatives in random order. In our experiments, these menus were derived from a set of 5 headphone sets (Experiments 1, 2; menus with varying numbers of alternatives) and a set of 6 lotteries that were pairwise-unranked by second-order stochastic dominance (Experiment 3; binary menus). In the target treatment, subjects have the option to defer choice at any menu, whereas in the control treatment they are always required to make an active choice. After this stage, one menu is randomly selected for each subject to be the payoff-relevant one. Before being asked to make a final choice from that menu for a 1/4 chance of winning the chosen alternative, subjects either receive experiential information about the alternatives in the menu (we allow them to try the relevant headsets in Experiment 1) or do not (Experiments 2 and 3). Crucially, if a subject turns out to win the chosen alternative, and her final choice differs from the one made originally at that menu (of which she was reminded before making the final choice), then she would have to pay a switching cost which was deducted from the amount of money allocated to her at the start of the experiment. On the other hand, if she had originally deferred choice at that menu, she would pay a lower deferral cost. Finally, if her original choice and final choice were the same, she would pay nothing after winning her alternative.

We find that significant proportions of subjects in the target treatment opted for costly choice deferral, both when doing so allowed them to receive more information. 

---

2In the personality psychology literature, Rassin (2007), for example, approaches indecisiveness as a phenomenon that results from decision impediments such as lack of information, valuation difficulty and outcome uncertainty. Among other things, this literature finds that individuals with high scores in indecisiveness personality questionnaires are significantly more likely to delay making a decision and to seek additional information.
about the choice alternatives (Experiment 1: 33%) and also when it did not (Experiment 2: 42%; Experiment 3: 37%). With consistency of the participants’ active choices measured separately by the number of WARP, Congruence and Binary Congruence violations they gave rise to, we also find significant differences in consistency between forced- and non-forced-choice subjects in Experiments 1 and 2. Although such a treatment effect was not found in Experiment 3, a significant difference in consistency of the subjects who were allowed to defer and did so at least once and the subjects who were allowed to defer but did not was detected in that experiment. This, in particular, points towards a novel type of overconfidence in decision making whereby subjects effectively consider their preferences to be more well-defined than they actually are, and mistakenly self-force choice upon themselves.

An important implication of our analysis is that data from forced-choice experiments entail the risk of the analyst erroneously concluding that the subjects in question are not maximizing some stable (i.e. menu- or context-independent) and transitive preference relation. Our results suggest that if subjects are given the opportunity to defer choice, then some of the inconsistency that might have been observed if they were forced to choose may in fact disappear, and a stable -but possibly incomplete-preference ordering may be elicited.³

In addition to the experimental design that we propose and the negative forced-choice effect on consistency that we document, we deploy a novel non-parametric combinatorial-optimization method for assessing the consistency of datasets that may also include deferral/outside-option choices. More specifically, taking a list of six general choice models as the primitive (including ones that predict consistent active choices alongside occasionally deferring behaviour), we aim to recover the model(s) that best explain(s) a given individual subject’s dataset by asking the following question: which model requires the smallest number of changes to the dataset to fit it perfectly? Applying this distance-score method that effectively extends the axiom-based Houtman and Maks (1985) principle⁴ in a model-based way, we assign to each subject the model with the lowest distance score for that subject.

Further to the standard model of utility maximization on a finite set of general alternatives and an extension of that model that also allows for choice of the outside option, our distance-score analysis is also informed by the models of undominated choice with incomplete preferences (Schwartz, 1976; Eliaz and Ok, 2006); maximally

³We also stress, however, that the experiments we report on below are not replications of existing experiments. We therefore make no claims that the findings in existing forced-choice studies documenting intransitivities in riskless or risky forced choice (e.g. May, 1954; Tversky, 1969; Loomes, Starmer and Sugden, 1991) are non-robust in deferral-permitting environments. Such investigations are left for future work.

(Gerasimou, 2018) and partially (Gerasimou, 2016a) dominant choice with incomplete preferences; and overload-constrained utility maximization (Gerasimou, 2018). With the exception of undominated and partially dominant choice, all these models predict WARP- and Binary Congruence-consistent active choices. Applying our method on the richer data that we collected in Experiments 1 and 2 and which include decisions in binary as well as non-binary menus, we find that 67% of all subjects in the non-forced-choice treatments are best explained by utility maximization, 27% by maximally dominant choice with incomplete preferences, 2% by partially dominant choice with incomplete preferences, and 4% by overload-constrained utility maximisation.

The model of maximally dominant choice, in particular, portrays a decision maker with incomplete preferences as choosing one of the most preferred alternatives in a menu if such an alternative exists, and as deferring at that menu if such a preference-dominant alternative does not exist. Unlike utility maximization with an outside option, this model therefore provides a suitable explanation of behaviour such as the one reported in the quote from our opening paragraph: each alternative is individually better than the deferral outside option, but the decision maker’s inability to choose in a preference-guided way between such alternatives leads her to defer choice. A rather striking result of our analysis considering the deterministic nature of preferences and choices in the models that have been included in it is that subjects in both the forced and non-forced choice treatment on average deviate from perfect conformity with their best-matching model by a single decision. Finally, our distance score analysis suggests that, conditional on a subject making active choices at all menus, her behaviour is more likely to perfectly conform with utility maximization if she operates in a non-forced choice environment, which is consistent with (and reinforces) the above-mentioned negative forced-choice effect on consistency that is obtained through different methods.

The remainder of the paper is structured as follows. Section 2 outlines the general structure of our experimental design and places it in the literature. Sections 3 and 4, respectively, elaborate on the methodological aspects of our consistency and distance-score analysis. Section 5 provides a detailed presentation and comparison of the finer details in the design and implementation of our three experiments. Sections 6 and 7 present the empirical findings from Experiments 1,2 and Experiment 3, respectively. Section 8 discusses further our main results and presents some additional ones that shed some light on the possible reasons of deferring behaviour in our data. The last section concludes.
2 Experimental Design

2.1 An Approach Towards Incomplete-Preference Elicitation in the Lab

We used a between-subjects experimental design with two treatments: Non-Forced-Choice (NFC) and Forced-Choice (FC). Our design aimed to generate incentives in the direction of enabling subjects to make decisions in a way that reveals their preferences over the underlying choice alternatives, including, possibly, their indecisiveness/incompleteness component. In both treatments, every subject was presented with a sequence of menus generated from the relevant underlying set of alternatives. The order in which menus appeared was random. It varied across sessions but was the same for all subjects within each session. It was also ensured that each item appeared top-left, middle, top-right etc. in an even manner across menus.

Subjects in the FC treatment were asked to choose an item from all menus presented to them, without being able to defer choice. Subjects in the NFC treatment had the opportunity to choose one item or to select "I'm not choosing now" in each menu. Once past a menu, subjects could not go back to review and change their choice, except in the case of the randomly selected payoff-relevant menu (see below).

In both treatments, after everyone had made a decision in all menus (including "I'm not choosing now" in the NFC treatment), one menu was randomly selected for each subject. Everyone then saw their own randomly selected menu and was reminded of the decision they had made there initially. Subjects were informed from the beginning that, at the very end of the experiment, 1 out of every 4 subjects would be randomly selected to win the item of their final choice from their randomly selected menu.\(^5\) We refer to such subjects as "winners". Participants were also told from the beginning that winners might face some costs which would be deducted from their monetary endowment, \(I\), depending on both their first and second decisions at the randomly selected menu (see Table 1).

In particular, if a subject that later became a winner had decided to choose an option other than the one she originally selected from this menu, an amount \(c_r\), \(0 < c_r < I\), was taken away from her initially allocated \(I\). In contrast, there was no deduction if the subject opted for the same headset again the second time they chose from that menu. These features were shared by both the FC and NFC treatments. Finally, subjects in the NFC treatment who originally deferred at their randomly selected menu and later became winners incurred the cost of a \(c_d > 0\) deduction from the initial allocation of \(I\),

\(^5\)The inclusion of this feature in our design was driven purely by experiment-budget constraints.
Table 1: The subjects’ payoff as a function of their first and second decisions at the randomly selected menu, respectively.

<table>
<thead>
<tr>
<th>Randomly selected menu</th>
<th>First decision</th>
<th>Second decision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case 1</strong></td>
<td>Choose $x$</td>
<td>Choose $x$</td>
</tr>
<tr>
<td></td>
<td>$\rightarrow$ Receive $x$ and earn $I$</td>
<td></td>
</tr>
<tr>
<td><strong>Case 2</strong></td>
<td>Choose $x$</td>
<td>Choose $y \neq x$</td>
</tr>
<tr>
<td></td>
<td>$\rightarrow$ Receive $y$ and earn $I - c_r$</td>
<td></td>
</tr>
<tr>
<td><strong>Case 3</strong></td>
<td>Defer</td>
<td>Choose $z$</td>
</tr>
<tr>
<td></td>
<td>$\rightarrow$ Receive $z$ and earn $I - c_d$</td>
<td></td>
</tr>
</tbody>
</table>

Note: $I - c_r < I - c_d < I$

with $c_d < c_r$. Participants were told from the beginning that if they were not selected to win a headset they would receive their endowment, $I$, irrespective of their first and second decisions at the randomly selected menu.

Intuitively, under this incentive structure a decision maker who has a most preferred alternative in the given menu and does not expect this preference to change later -e.g. following further deliberation or after acquiring more information about the alternatives- would choose it at that menu in order to: (a) win that alternative; (b) maximize her monetary payoff, $I$. On the other hand, if the decision maker has no most preferred option in that menu and is deciding between deferring and actively choosing one of the undominated feasible ones, then, considering that $c_d < c_r$, she may regard deferral as safer than making choice which may later be reversed.\textsuperscript{6} In a real-world setting, $c_r$ might correspond to the time and monetary cost from having to return a purchased item to a shop with a no-refunds policy once the decision maker realizes that it was a sub-optimal choice, whereas $c_d$ might be thought of as the time cost associated with the ongoing decision process.

We stress that subjects were informed at the beginning of all sessions that, in order to proceed to the random selection of the payoff-relevant menu per subject, every participant should have completed their decisions at all menus. Therefore, no time-saving gains were to be made by quickly deferring at every non-singleton menu in the hope of leaving the lab sooner, unless every participant who behaved in this way

\textsuperscript{6}Azrieli, Chambers, and Healy (2018) identified conditions on subjects’ complete and transitive extended preferences over the uncertain acts that map the experiment-induced state space into the set of payoff-relevant choice alternatives in order for a hybrid random incentive system to make the subjects’ truthful preference-revealing choice incentive-compatible. Our experiment lies outside the range of application of their theorems, both because our subjects’ preferences—and hence their extended preferences—may be incomplete, and also because our design allows choice of the deferral/outside option and generates the richer set of three distinct possibilities that are laid out in Table 1. A formal investigation of the conditions on the subjects’ possibly incomplete extended preferences that, through dominant active choices, would make the truthful revelation of preferences incentive-compatible is left for future work.
(arbitrarily) expected every other participant to behave in the same way, and these expectations turned out to be correct.

2.2 Alternative Approaches in the Literature

We outline three broad approaches in the existing experimental economics literature through which other scholars have attempted to understand whether subjects’ behaviour may have been influenced by preference incompleteness. Although these approaches differ, one feature they have in common—and which also distinguishes them from our approach—is that they were all implemented in environments of choice under risk and/or ambiguity. In the case of the first and third approach, this feature of their design is essential for the relevance of the incomplete-preference interpretation of the corresponding findings.

(1) Preference imprecision in binary menus of money lotteries.

The recent study of Cubitt, Navarro-Martinez, and Starmer (2015) provides a detailed overview of the various applications of this approach and offers additional results. The key feature here is that subjects are given a list of binary menus, where each row in the list presents a binary menu comprising a certain amount and a non-degenerate lottery. The certain amount between consecutive rows in the list differ by a small fixed increment. The subject is asked to state in each row whether she is: (i) sure she prefers the lottery; (ii) sure that she prefers the certain amount; or (iii) unsure about her preference between the two. A subject’s responses reveal a coherent preference-imprecision interval if they (a) coincide with (i) when the certain amount is up to some level $a$; (b) coincide with (iii) when the certain amount is in the range $(a, b]$ for some $b > a$; and (c) coincide with (ii) when the certain amount is any $c > b$. Finally, subjects with a non-degenerate imprecision interval are asked to identify a row/binary menu within that interval that corresponds to their own “best estimate” of their certainty equivalent for the given lottery, in case their payoff-relevant randomly selected menu fell within that interval, so that the experimenters can reward them with either the lottery or the certain amount depending on the subject’s choices at that row/menu. This design, of course, does not make incentive-compatible the subjects’ certainty equivalent report in their preference-imprecision interval.

(2) Commitment, flexibility and deferral in binary menus of money lotteries.

Prominent studies include Cohen, Jaffray, and Said (1987); Butler and Loomes (1988); Dubourg, Jones-Lee, and Loomes (1994, 1997); Morrison (1998); Butler and Loomes (2007, 2011). As the authors noted, this literature has produced mixed results on whether preference imprecision—as measured in those studies—contributes importantly to violations of expected-utility theory under risk such as preference reversals or disparities between willingness-to-pay and willingness-to-accept. Moreover, in their own study—which comprises the latest development in this body of work that we are aware of—they give a negative answer to this question, despite their finding that non-degenerate preference-imprecision intervals do indeed exist in their data and are also stable and coherent.
Danan and Ziegelmeyer (2006) proposed a design that interpretationally links incomplete preferences under risk with *preference for flexibility* (Kreps, 1979) in an environment of choice over menus of lotteries where these lottery menus are restricted to be binary. The utility representation of preference for flexibility in Kreps’ ordinal framework features a generally non-unique subjective state space to reflect the agent’s uncertainty about her future preferences over the alternatives in the different menus. Importantly, the preference relation in such a representation emerges as a completion of the *incomplete dominance relation* over menus. With individual lotteries being viewed as singleton menus, menu $A$ weakly dominates menu $B$ in this sense if for every lottery in $A \cup B$ there is some lottery in $A$ that is weakly preferred to it. Danan and Ziegelmeyer’s restriction to binary menus of lotteries can therefore be thought of as mapping flexibility into incompleteness between the underlying lotteries through this channel. Specifically, subjects were interpreted as revealing incompleteness if, during the first stage of the experiment, they chose the “flexibility” menu that included a risky lottery and a certain cash amount over the two singleton “commitment” menus that consisted of the lottery and the certain amount, respectively, and with both these options topped-up with a tiny extra payment compared to their appearance in the binary “flexibility” menu. In the second stage of the experiment that took place a week later, one of the subjects’ decision problems was selected at random for payment and, depending on their first-stage decision in that problem, they were rewarded with either the lottery/cash in their commitment menu or with the lottery/cash that they were then asked to choose from their flexibility menu.

A common feature between our design and Danan and Ziegelmeyer’s is that the “flexibility/choice deferral” option was both allowed and made costly in both designs, even though in these authors’ setting this took the form of a bonus for choosing the “commitment/active choice” option whereas in our setting it takes the form of an explicit cost for deferring that is taken away from some initial endowment. Among the several differences between our design and these authors’, however, and arguably the most important one, is that subjects in their experiment were *unable to reverse* their first-stage choices, whereas in our design they are allowed to do so and, crucially, this is more costly to them than deferring.

(3) *Mixing in forced choices from binary menus with money lotteries or acts.*
Cettolin and Riedl (2019) reported findings from an experiment where subjects were asked to either (i) choose one item from every binary menu in a sequence of such menus that feature a lottery and an ambiguous monetary act over the same two certain amounts, or (ii) choose the “mix” option of delegating choice between these prospects to a randomization device. They found that half of the subjects actually opted for the
mix option more than once, and explained that such behaviour in this particular sequence of menus is incompatible with models of complete preferences under ambiguity. The authors also reported findings from a subsequent experiment suggesting that, from those subjects who chose the mix option more than once, half of them were always unwilling to pay a small amount to do so, and about one third were always willing to do so. The authors concluded that the first class of these repeatedly-mixing subjects exhibit behaviour that is consistent with Bewley’s (2002) incomplete-preference model of indecisiveness in beliefs, and that the latter class are consistent with related experimental findings and models of preference for randomization in Agranov and Ortoleva (2017) and Cerreia-Vioglio, Dillenberger, Ortoleva and Riella (2019), respectively.

3 Consistency Analysis: Methodology

3.1 Choice Reversals and Binary Choice Cycles

Given a set of general choice alternatives $X$ and a collection $M$ of menus derived from $X$, the decision maker’s observable behaviour is described by a choice correspondence $C$ on $M$ that satisfies $\emptyset \subseteq C(A) \subseteq A$ for every menu $A$ in $M$. Here, $C(A) = \emptyset$ models the situation where the decision maker, by opting for the no-choice/choice-deferral outside option, has chosen none of the feasible market alternatives. The agent’s weak, strict and indirect weak revealed preference relations $\succ_R$, $\succ^R$ and $\succ^R$ are defined, respectively, by $x \succ_R y$ if there is a menu $A$ such that $x \in C(A)$ and $y \in A$; $x \succ^R y$ if there is a menu $A$ such that $x \in C(A)$ and $y \in A \setminus C(A)$; and $x \succ^R y$ if there are alternatives $x_1, \ldots, x_n$ such that $x = x_1$, $y = x_n$ and $x_i \succ^R x_{i+1}$ for all $i = 1, \ldots, n - 1$. The restrictions of these relations on binary menus will be denoted by $\succ^B$, $\succ^B$ and $\succ^B$, respectively.

Given these concepts and notation, we can now introduce compactly the following fundamental principles of choice consistency (Samuelson, 1938; Uzawa, 1956; Arrow, 1959; Richter, 1966; Sen, 1971) that lie at the heart of our analysis:

**WARP**

$x \succ^R y \implies y \not\succ^R x$.

**Congruence**

$x \succ^R y \implies y \not\succ^B x$.

---

9 See Gerasimou (2016a, 2018) for justifications of this modelling approach.
Binary Congruence

\[ x \succ^B y \implies y \not\sim^B x. \]

We compare the incidence and distribution of WARP and Binary Congruence violations in the FC and NFC treatments.\(^\text{10}\) It is well-known that if a decision maker with incomplete preferences is forced to choose, and does so by always opting for one of the \textit{undominated} alternatives, then her choices will generally be incompatible with each of the above two principles of consistency (Luce and Raiffa, 1957; Schwartz, 1976; Mandler, 2005; Eliaz and Ok, 2006). On the other hand, if she operates in a non-forced choice environment and employs the \textit{dominance} decision rule whereby a market alternative is chosen if and only if it is the most preferred feasible one, then such an agent will make fully consistent active choices but will occasionally opt for the no-choice/deferral outside option, not because it is preferred to the feasible market alternatives, but because the incompleteness of her preferences deprives her of the possibility of making an optimal choice in that sense (Gerasimou, 2018). The opposing predictions that these distinct decision rules and choice environments generate for the same individual with incomplete preferences can therefore be thought of as a theoretical motivation for the hypothesis that we are testing.

WARP and Binary Congruence are logically independent but each of them is implied by Congruence even when \( C \) is possibly empty-valued, as it is assumed to be here. WARP violations such as \( C(\{w, x, y\}) = \{x\} \) and \( C(\{x, y, z\}) = \{x, y\} \) amount to \textit{direct choice reversals} involving two market alternatives \( x \) and \( y \). Violations of Binary Congruence are \textit{binary choice cycles} such as \( C(\{x, y\}) = \{x\}, C(\{y, z\}) = \{y\} \) and \( C(\{x, z\}) = \{x, z\} \). These too are of particular interest given that choices at binary menus are intuitively the most likely to reflect a genuine preference for one alternative over the other, and also considering that many choices from lab experiments – especially with money lotteries – often come from binary menus. In addition to these two special ways in which the more general Congruence axiom can be violated, however, the latter can also happen by means of more general revealed-preference cycles such as \( C(\{a, b, c\}) = \{a\}, C(\{b, c, d\}) = \{b\} \) and \( C(\{a, c, d\}) = \{c, d\} \). Therefore, analysing the subjects’ conformity with respect to each of these principles offers complementary ways in which one can understand how consistent their active choices actually are. The results presented below pertain to the analysis of the subjects’ actual \textit{single} choices, and are thus informed by violations of the above axioms in the \textit{strict-cycle} sense. However, in Section 3.3 below and in Online Appendix 5 we explain (briefly and in more...
detail, respectively) how we also elicited and processed verbal, stated-indifference data, and in Sections 6.2 and 7.2 we present our corresponding findings for both the primitive raw as well as for the indifference-augmented choices.

3.2 Two Normalizations of Consistency Violations to Facilitate Cross-Treatment Comparisons

One might argue that NFC subjects who exercise their option to defer choice and hence end up with fewer active choices than their FC counterparts are expected to exhibit fewer violations of WARP, Congruence and Binary Congruence in general, precisely because of this discrepancy and the correspondingly fewer opportunities to make “mistakes”. In other words, the number of revealed-preference cycles might be biased downwards in NFC data. Arguing along these lines, of course, implicitly assumes that subjects should generally not be expected to make active choices from every possible menu in a utility-maximising (hence perfectly consistent) way, for otherwise no difference would be expected in their consistency levels regardless of whether they operated in a forced or non-forced choice environment. Yet, this statement is essentially the very hypothesis that we are testing in our study.

This point notwithstanding, we propose two possible ways in which an analyst might respond to this potential concern. Both these ways involve normalising the violations that are recorded from an NFC treatment in a way that makes them more directly comparable to violations from a similarly-structured FC treatment.

1. Linear normalizations: for each subject, divide the number of WARP, Congruence and Binary Congruence violations by that subject’s total number of active (and in the latter case, binary active) choices.

2. Nonlinear normalizations: for each subject, divide the number of WARP and Binary Congruence violations by the maximum number of WARP and Binary Congruence violations that is theoretically possible given the subject’s number of active (respectively, binary active) choices. Such a non-linear normalisation was not performed for Congruence violations due to computational constraints.

The first method is simpler to explain and apply: if the average subject’s consistency violations are generally expected to increase linearly in the number of active choices she makes, then dividing violations by that number removes any potential bias and makes FC and NFC data more comparable. In addition to its conceptual and computational simplicity, this method has the advantage that the only NFC subjects that must necessarily be removed in order for it to be applied are those who always deferred and who are therefore perfectly consistent in a trivial way.
Although the second method is computationally very challenging, it does not assume a linear positive relationship between consistency violations and active choices. The idea here is that the normalisation accounts for the worst that could happen to the subject’s consistency violations conditional on the particular number of menus that she chose to defer at. This maximum is also increasing in the number of active choices. Therefore, for a fixed number of violations, the corresponding normalised violations that come about by dividing the original by the maximum that are possible for a given number of deferrals are again decreasing in the number of deferrals. However, unlike the linear-normalisation method above, violations that are normalised in this way are indeed neither linearly nor strictly decreasing. Moreover, this method is even more “punitive” than the above towards NFC subjects in that it requires not only dropping perfectly consistent subjects with zero active choices but also subjects with one (cf. normalised WARP violations) or up to two (cf. normalised Binary Congruence violations). Although these additional exclusions provide yet another robustness check for our tests, we acknowledge the possibility of it potentially leading to a pro-FC consistency bias instead.

3.3 Indifference Elicitation and the Transition from Choice Functions to Choice Correspondences

Given that at most one item was actually chosen from every menu in all three experiments, in an attempt to disentangle strict preference, indifference and indecisiveness between two alternatives, in Experiments 1 and 3 we used a survey-data method that was based on the subjects’ responses to questions that appeared after their binary-menu decisions. Following processing of these indifference statements and acceptance of those that were consistent in the sense described below, we then extrapolated the constructed indifference classes from binary choices to choices from bigger menus as well.

Specifically, at each binary menu where FC and NFC subjects from these two experiments had chosen a headset, they were also asked if i) “they preferred it to the other headphone set in this menu”; ii) “they found both to be equally good, and therefore chose randomly”; iii) “other reason”. Subjects were also told from the beginning that if a menu with two headsets was randomly selected for them and they had previously stated that “they found both to be equally good”, then they would not get the chance to change their choice at that stage if they had chosen an alternative, and one of the two

---

11 The relevant graphs showing the maximum number of WARP and Binary Congruence violations as a function of the relevant number of active choices are shown in Online Appendix 9. We note that this output emerged from novel and highly non-trivial computations. The relevant computer program is available from the authors upon request.

12 This was not done in Experiment 2 in order to simplify the experimental instructions.

13 The computational process is described in more detail in Online Appendix 5.
elements of that menu would be selected for them at random by flipping a coin. Thus, subjects had no incentive to state indifference in binary menus where they chose an alternative if they were not actually indifferent, because in the event that their randomly selected menu was binary this statement would deprive them from the possibility of benefiting from making a final choice after having tried out the two relevant headsets (Experiment 1) or after thinking more about the lotteries in their menu (Experiment 3). We note that this design does not make it a dominant strategy for subjects who are actually indifferent to state so. It does, however, make such a statement an undominated strategy, provided subjects are truly indifferent, in which case they would value neither the additional information/time to think about the problem more, nor the opportunity to change their choice later.\footnote{If subjects have a preference for randomization on the other hand (Agranov and Ortoleva, 2017; Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella, 2019), then stating indifference here would be the dominant strategy for them.}

Following binary menus where the NFC subjects’ decision was the “I’m not choosing now” option, they were asked if they made this choice because i) “they could not decide which one they prefer”; ii) “they found both to be equally good”; iii) “other reason”. If such a menu was randomly selected for them and they had deferred at that menu and had also stated indifference, then in the event that they were winners, the £1 or €0.50, €0.20 deferral cost (in Experiment 1 and 3, respectively) was applied and they won the alternative that was determined by the flip of a coin. The coin-flip procedure did not apply in the other two cases.

In constructing weak preferences from this combination of choice and stated-preference data from binary menus we give priority to choices over indifference statements whenever the two disagree. More specifically, the indifference-augmentation process consists of two parts: a consistency check of indifference statements (which may remove some or all of them), and then the augmentation itself. In the consistency check, we first consider all indifference statements following active choice, and for each of them check if the stated indifference between the pair of alternatives generates a violation of weak-preference transitivity given the subject’s choices at all other binary menus and given her indifference statements. Any indifference statement that generates a cycle (e.g. choices $C(\{x, y\}) = \{y\}$, $C(\{y, z\}) = \{z\} = C(\{x, z\})$ and expression of indifference between $x$ and $z$) is ignored, in which case we view an active choice at a binary menu as an expression of strict preference between the alternatives. The augmentation that follows the consistency check transforms the original singleton choice -where applicable- at each menu into the possibly multi-valued choice that comprises every alternative in that menu that was indifferent to the chosen alternative, including the latter.

Figure 1 illustrates the indifference elicitation and augmentation process with two
Figure 1: Indifference-augmented choice: examples of (un)acceptable combinations of choice & indifference data.

(a) examples. The subjects’ active choices are the same in both and are compatible with the ordering \( w \succ x \succ y \succ z \), but their indifference statements differ slightly. In Figure 1(a), the subject has declared herself indifferent between (i) \( w \) and \( x \); (ii) \( w \) and \( y \); (iii) \( x \) and \( y \). These statements induce a transitive indifference relation that does not conflict with any of the subject’s active choices or the other (non-)indifference statements. These statements are therefore acceptable. The transition from single-valued to multi-valued, indifference-augmented choices at the three menus involved therefore become

\[
\begin{align*}
C(\{w, x\}) &= \{w\} \rightarrow C(\{w, x\}) = \{w, x\}, \quad C(\{w, y\}) = \{w\} \rightarrow C(\{w, y\}) = \{w, y\} \quad \text{and} \\
C(\{x, y\}) &= \{x\} \rightarrow C(\{x, y\}) = \{x, y\} \quad \text{for this subject. Moreover, in every bigger menu where one of these three alternatives is chosen and at least another is in the menu, all feasible alternatives that are in the same indifference class to the one that was actually chosen are now considered to be choosable. For example, } \\
C(\{w, x, z\}) &= \{w\} \rightarrow C(\{w, x, z\}) = \{w, x\} \quad \text{and} \\
C(\{w, x, y, z\}) &= \{w\} \rightarrow C(\{w, x, y, z\}) = \{w, x, y\}.
\end{align*}
\]

By contrast, in addition to (i), (ii) and (iii), the subject has also declared herself indifferent between (iv) \( w \) and \( z \) in Figure 1(b). Yet, with this additional statement the constructed indifference relation is no longer transitive because it conflicts with the absence of indifference statements between \( y \) and \( z \) and also between \( x \) and \( z \) in the other two pairs. Since there is no obvious and non-arbitrary way to retain a subset of the indifference statements or to include additional ones that were not actually made by the subject in order to break the resulting intransitivity, all indifference statements are now discarded and the augmented choices coincide with the original, single-valued

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference Stated</th>
<th>Ticked</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>Yes</td>
<td>w,x</td>
</tr>
<tr>
<td>w,y</td>
<td>w</td>
<td>Yes</td>
<td>Yes</td>
<td>w,y</td>
</tr>
<tr>
<td>w,z</td>
<td>w</td>
<td>-</td>
<td>-</td>
<td>w</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>Yes</td>
<td>x,y</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>-</td>
<td>-</td>
<td>y</td>
</tr>
<tr>
<td>x,z</td>
<td>x</td>
<td>-</td>
<td>-</td>
<td>x</td>
</tr>
</tbody>
</table>

Indifference classes: \( \{z\}, \{w, x, y\} \)

(b)
It is evident from these examples that a feature of our indifference-augmentation method is a strong tendency toward making subjects’ behaviour appear at least as consistent as their behaviour based on single-valued choices alone, and will typically (but not always) make it appear strictly more consistent than the latter. In the case of the first subject, for example, the extrapolation from binary to bigger menus gives her many “degrees of freedom” in reversing her choice from $w$ to either $x$ or $y$ without this counting as a WARP violation, which, of course, ought to be the case for someone who is indeed indifferent between these three alternatives. But even in the case of the second subject’s behaviour, this ultimately enters the analysis as fully consistent on the basis of her single-valued raw active choices despite the presence of inconsistencies in the indifference statements that were elicited from her.

4 Goodness of Fit Analysis: Model Distance Scores

4.1 Outline of the General Method

In addition to the number of violations of some core consistency axiom that was mentioned in Section 3, various other indices that capture the degree of a choice dataset’s “rationality” have been proposed in the recent literature. Some of these indices are only applicable in choice from Walrasian budget sets, while others can in principle be used in more general choice environments as well. A well-known and widely applied general measure of the latter kind is the so-called Houtman-Maks (1985) index which identifies the degree of consistency of a choice dataset with the size of the maximal subset of this dataset that is consistent with some axiom (typically WARP, SARP or GARP). In our present framework of choice from a finite set of general alternatives, when the dataset under investigation is complete in the sense that it contains no deferral observations, an equivalent way in which standard application of the HM index on WARP can be interpreted is as counting the smallest number of changes that need to be made on such a dataset for it to be as if it was generated by the model of utility maximization. While a count of zero obviously corresponds to perfect conformity with this model, an important feature of this analysis is that positive but low HM scores here may suggest that the relevant subject does not conform perfectly with utility maximization due to a few choice errors rather than due to systematic deviations from this model.

While all existing indices of “rationality” -including the HM index itself- have been

proposed as relevant for complete datasets, an interesting but unexplored feature of the HM index is that its core idea is sufficiently flexible to be made applicable in a model-based way and, in particular, toward performing a goodness-of-fit analysis for possibly bounded-rational models that generate either forced- or non-forced choice datasets. More specifically, taking a list of general choice models (including ones that occasionally predict deferring behaviour) and all their -finitely many- possible instances\textsuperscript{16} as the primitive, our proposed method toward understanding which model best explains a given individual choice dataset, whether complete in the above sense or not, is to apply the following non-parametric, combinatorial-optimization principle: \textit{find the model that requires the smallest number of changes in the dataset for it to become perfectly explainable by (some instance of) that model.} We say that this -possibly non-unique- model has the smallest distance score for that subject, and, after suitable noisiness-based robustness checks, we categorize that subject as behaving according to that model.\textsuperscript{17}

4.2 Models of Forced and Non-Forced Choice

Before stating the models that have been included in our present distance-score analysis we first remind the reader of a few basic definitions pertaining to properties that a weak and/or strict binary preference relation might have. Specifically, a general preference relation \( R \) on a finite set of alternatives \( X \) is \textit{reflexive} if \( xRx \) holds for all \( x \in X \); \textit{transitive} if \( xRy \) and \( yRz \) implies \( xRz \); \textit{asymmetric} if \( xRy \) implies that \( yRx \) is not true; \textit{complete} (respectively, \textit{total}) if \( xRy \) or \( yRx \) holds for all (respectively, all \textit{distinct}) \( x, y \in X \); \textit{incomplete} (\textit{non-total}) if there are (\textit{distinct}) \( x, y \in X \) such that neither \( xRy \) nor \( yRx \) holds; and has a \textit{non-trivial indifference relation} if the symmetric part of \( R \) that is defined by \( xIy \) if both \( xRy \) and \( yRx \) are true is reflexive, transitive and such that \( xIy \) holds for some distinct alternatives \( x, y \in X \).

We are now in position to formally introduce and briefly discuss the six models that we use to analyse the choice data from the three experiments through the lens of the distance-score method.

1. \textit{Utility maximization (UM)}: there is a complete and transitive preference relation \( \succsim \) on \( X \) such that, for every menu \( A \),

\[
C(A) = \{ x \in A : x \succsim y \text{ for all } y \in A \}. \tag{1}
\]

\textsuperscript{16}For the purposes of this study the different instances of the various models are captured by the different (and possibly incomplete -see below) orderings on a set of 5 elements. The reader can find more details about the precise numbers of the various kinds of orderings on such a set in the following entries of The On-Line Encyclopedia of Integer Sequences: https://oeis.org/A000798, https://oeis.org/A001035, https://oeis.org/A000670 and https://oeis.org/A000142.

\textsuperscript{17}This method was first introduced in Costa-Gomes, Cueva, Gerasimou, and Tejiščák (2016) and has since been extended in many directions through the open-source Prest computer program (Gerasimou and Tejiščák, 2018) that can be downloaded from https://prestsoftware.com.
This is the baseline model of rational choice.

2. **Utility maximization with an outside option (UMOO):** there is an alternative \( x^* \in X \) and a complete and transitive preference relation \( \succeq \) on \( X \) such that, for every menu \( A \),

\[
C(A) = \begin{cases} 
\{ x \in A : x \succeq y \text{ for all } y \in A \}, & \text{if } x \succ x^* \text{ for such } x \in A \\
\emptyset, & \text{if } x \succeq y \text{ for all } y \in A \text{ implies } x^* \succeq x, 
\end{cases}
\]

(2)

and, in addition, \( x^* \succeq x \) holds for some \( x \in X \) that is distinct from \( x^* \). In the present formulation\(^{18}\) of this extension to the baseline model of utility maximization when an outside option is feasible it is understood that the agent’s outside option \( o \) enters her preference domain directly and belongs to the same indifference class as \( x^* \in X \); this option is feasible in every menu and, intuitively, captures her *desirability threshold:* when the most preferred feasible alternative(s) is (are) preferred to \( x^* \), then one of these alternatives is chosen from the menu, otherwise the agent defers (or, perhaps more appropriately in this case, *permanently avoids*) making an active choice there.\(^{19}\)

3. **Undominated choice with incomplete preferences (UC).** The strict and non-strict versions of this model (axiomatized in Schwartz, 1976 and in Eliaz and Ok, 2006, respectively; see also de Clippel and Rozen, 2018) postulate either the existence of an asymmetric, transitive and non-total preference relation \( \succ \) on \( X \), or the existence of a reflexive, transitive and incomplete preference relation \( \succeq \) on \( X \) whose asymmetric part is the same strict relation \( \succ \) as above and its symmetric part \( \sim \) is a non-trivial indifference relation.\(^{20}\) In both versions of the model the agent’s choice behaviour is such that, for every menu \( A \),

\[
C(A) = \{ x \in A : y \not\succ x \text{ for all } y \in A \}.
\]

(3)

\(^{18}\)This was formalised and analysed in Gerasimou (2018, Section 3). To our knowledge, this was also the first study that focused on the axiomatic structure -in a full-domain environment- of the model of utility maximization with an outside option.\(^{19}\)This model can be equivalently formulated with \( x^* \not\in X \) instead. Moreover, “\( C(A) = \emptyset \)” can be replaced with “\( C(A) = \{ x^* \} \)” in this model without any loss of generality. Crucially, however, this is no longer true for the other models of non-forced choice we consider here. See Gerasimou (2018, Sections 3, 6) for a more detailed discussion of these points.\(^{20}\)We note that although the non-strict version of the model has a richer structure than its strict counterpart, any dataset that is perfectly compatible with that version for some incomplete weak preference relation \( \succeq \) with strict part \( \succ \) and non-trivial (transitive) indifference part \( \sim \) is in fact perfectly compatible with the strict version of the model with the same strict preference relation \( \succ \) and with the incomparability relation with respect to \( \succ \) being a generally intransitive relation that contains all pairs of distinct alternatives that are declared equally good under the indifference relation \( \sim \) of the non-strict model. While the converse implication is not generally true if one insists on the indifference relation in the non-strict model being non-trivial (in which case the strict model is more general than the non-strict one), this implication does become true if one drops this requirement. In this case, the two versions of the model become *observationally equivalent* for datasets that are *perfectly* explainable by the undominated-choice model, and the distance scores of the two versions will generally differ in other datasets (see Table 7 below).
Undominated choice is the baseline model of forced choice with incomplete preferences. It predicts that a decision maker who does not have a most preferred alternative in a menu due to the incompleteness of her preferences, but who nevertheless has to make an active choice somehow, ultimately chooses one of the preference-undominated alternatives instead.

4. **Maximally dominant choice with incomplete preferences (MDC):** there is a reflexive, transitive and incomplete preference relation $\succeq$ on $X$ such that, for every menu $A$,

$$C(A) = \{ x \in A : x \succeq y \text{ for all } y \in A \}. \tag{4}$$

This model of dominant non-forced choice with incomplete preferences was motivated and formally analysed in Gerasimou (2018, Sections 2, 5–7). It predicts that the individual makes an active choice at a menu if and only if there is a most preferred alternative in that menu, and otherwise defers. Unlike utility maximization with an outside option, this model explains simply and, one hopes, intuitively behaviour such as the one summarized in the opening quote. In particular, the deferral outside option here does not enter the individual’s preference domain, and every alternative is sufficiently desirable to be chosen when on its own. When the deferral outside option is chosen, however, this happens not because it is preferred to the feasible market alternatives, but instead because the agent cannot decide on the best alternative(s) in the relevant menu and is reluctant to choose an undominated option that might later turn out to be inferior to some of the alternatives that it was initially incomparable to. The difference between (1) and (4) therefore lies in the fact that, due to preference incompleteness, the set of choosable alternatives $C(A)$ will occasionally be empty in (4) whereas it will always be nonempty in (1).

5. **Partially dominant choice with incomplete preferences (PDC):** there is an asymmetric, transitive and non-total preference relation $\succ$ on $X$ such that, for every menu $A$,

$$C(A) = \emptyset \iff x \not\succ y \text{ and } y \not\succ x \text{ for all } x, y \in A \tag{5a}$$

$$C(A) \neq \emptyset \iff \begin{cases} z \not\succ x \text{ for all } z \in A & \text{ and } \\
 x \succ y & \text{ for some } y \in A \end{cases} \tag{5b}$$

This is a model of non-forced and context-dependent choice but with context-independent and incomplete preferences. It is in the spirit of the reason-based narrative (Shafir, Simonson, and Tversky, 1993) and predicts that the individual
makes an active choice of a market alternative only if it is undominated and at least partially (as opposed to maximally) dominant in the menu in the sense of it being preferred to at least some other feasible alternative. When no such alternative exists in a menu because no two alternatives in it can be compared, then the agent defers. Maximally and partially dominant choice therefore coincide in their predictions in this special case for a given incomplete preference relation, but generally diverge otherwise. This model was proposed and analysed in Gerasimou (2016a).  

6. Overload-constrained utility maximization (OCUM). We consider the simple special case of the model that postulates the existence of a complete and transitive preference relation $\succeq$ on $X$ and of an individual-specific, constant menu-size complexity threshold $n < |X|$ such that, for every menu $A$,

$$C(A) = \begin{cases} 
\{ x \in A : x \succeq y \text{ for all } y \in A \}, & \text{if } |A| \leq n \\
\emptyset, & \text{if } |A| > n
\end{cases}$$

(6)

This model of non-forced choice [Gerasimou (2018, Sections 4–6)] predicts that the decision maker is cognitively constrained in the sense that she has a cut-off menu size below which she chooses from the relevant menu like a standard utility maximizer, while if the number of alternatives in the menu strictly exceed that threshold she defers choice, in the spirit of empirical findings suggesting the existence of choice overload effects.

Remark 1. Models 2–6 presented above generally include standard utility maximization (UM) with strict and/or weak preferences as a special case. In the way they were stated above, however, they have been restricted not to do so as they were explicitly required to feature an outside option/desirability threshold that lies above at least one market alternative in the agent’s ranking (cf. UMOO); non-total/incomplete preferences (cf. UC, MDC, PDC); and a menu-complexity threshold that does not exceed

---

21 The properties of its forced-choice counterpart where “$C(A) = \emptyset$” is replaced by “$C(A) = A$” on the left hand side of (5a) were analysed in Gerasimou (2016b), while additional axiomatizations of that version were given in Qin (2017) and Barokas (2017). We emphasize that when active choices are single-valued (as they are in our primitive experimental data), this forced-choice version of the model coincides with the baseline model of undominated choice and is therefore omitted from our analysis without loss of generality.

22 The general version of this model allows for menu complexity to be captured by a function on the set of menus that is increasing with respect to set inclusion. The special case that we study here where a menu’s complexity is identified with its size was also used in Gerasimou and Papi (2018), for example, to model a population of heterogeneously choice-overloaded consumers that firms in a duopolistic market were competing for. Some recent studies on deterministic or stochastic preference representation theorems where the agent’s preferences over menus are explicitly portrayed as featuring an aversion towards larger menus in a variety of cardinal environments are Sarver (2008), Ortoleva (2013), Fudenberg and Strzalecki (2015) and Buturak and Evren (2017).

23 Iyengar and Lepper (2000) is the seminal study that demonstrated in a field experiment how the increased complexity associated with making an active choice in large menus of real goods -with as many as 30 alternatives- often leads to choice deferral in such menus, a phenomenon that is commonly referred to as choice overload in the literature. Other relevant studies include, for example, Iyengar and Kamenica (2010); Iyengar, Huberman, and Jiang (2004); Scheibenhenne, Greifeneder, and Todd (2010); Chernev, Böckenholt, and Goodman (2015).
the total number of alternatives in the largest possible menu (cf. OCUM). These restrictions are imposed precisely in order to estimate these models’ added explanatory value in our data relative to the utility maximization benchmark.

5 Structure of the Three Experiments

The instructions that were given to subjects are available in Online Appendices 1, 2 and 3, respectively. The experimental interfaces were created with the z-Tree software (Fischbacher, 2007) and subjects were recruited through the ORSEE software (Greiner, 2015). This section outlines the general structure of the design that was common in the three experiments, while in later sections we provide more information about features that were particular to each of them.

The grand choice set in Experiments 1 and 2 consisted of five headsets, whereas that in Experiment 3 comprised six money lotteries. The headsets in Experiments 1 and 2 were the same and their brand names and models were chosen so that the products’ prices were approximately the same (between £10 and £20 at the time of purchase) but their attributes differed in ways that made comparisons between them non-trivial. For instance, some headsets were basic but with well-known brand names, whereas others were more sophisticated or had some superior or distinctive features but were associated with less recognizable brand names (e.g. the headset with the less commonly known brand name was wireless/bluetooth whereas all others were not). The six 3-outcome lotteries used in Experiment 3 are displayed in Table 2. They were constructed so as to have the same expected value of €20 -this was not communicated to subjects- but be pairwise-unranked by second-order stochastic dominance (SOSD). This was expected to generate trade-offs involving, for example, the maximum amount (higher in lottery $x$ than in $y$) and the most likely or smallest amount (higher in lottery $y$ than in $x$).

Table 2: The six lotteries used in Experiment 3.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$25 \frac{100}{100} \circ e2; 35 \frac{100}{100} \circ e18; 40 \frac{100}{100} \circ e33$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B$</td>
<td>$25 \frac{100}{100} \circ e2; 67 \frac{100}{100} \circ e25; 8 \frac{100}{100} \circ e34$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>$20 \frac{100}{100} \circ e2; 60 \frac{100}{100} \circ e16; 20 \frac{100}{100} \circ e50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>$20 \frac{100}{100} \circ e3; 50 \frac{100}{100} \circ e13; 30 \frac{100}{100} \circ e43$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E$</td>
<td>$30 \frac{100}{100} \circ e4; 40 \frac{100}{100} \circ e20; 30 \frac{100}{100} \circ e36$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>$10 \frac{100}{100} \circ e1; 70 \frac{100}{100} \circ e19; 20 \frac{100}{100} \circ e33$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The menus (decision problems) consisted of images of the relevant headsets and of
a short description of their main features (see Online Appendix 1 for some examples). Menus were identical in both experiments. However, although all 31 possible menus from the grant set of 5 headsets were presented in Experiment 1, only the 26 non-singleton ones were shown to subjects in Experiment 2. In order to make the decision problems as realistic as possible, the short description of each headset’s main features reproduced exactly the same information (in bullet-point form) that the large online retailer from which the headsets were purchased had chosen to provide on the relevant product’s web page. As a consequence, a direct attribute-by-attribute comparison of the various headsets was typically impossible because the information for different items revolved around different attributes. By contrast, and in line with much of the experimental literature in choice under risk that focuses on binary menus, each decision problem in Experiment 3 corresponded to one of the 15 possible pairs of lotteries shown in Table 2. Each lottery in every such pair was presented to subjects using simple pie charts (see Online Appendix 3 for an example).

To facilitate the subjects’ familiarity with -and understanding of- the experiment’s procedures, before the beginning of the main task which involved the sequential presentation of all relevant menus, subjects in Experiments 1 and 2 went through a mini (“trial”) version with three completely unrelated alternatives (MP3 players or voice recorders) that aimed to accustom them with the experimental interface. All decision problems in the trial were hypothetical and subjects were aware of this. After the trial, subjects were asked to answer a series of quiz questions that aimed to help them understand -and also test their understanding of- the instructions. Subjects had to answer all questions correctly before proceeding to the main phase of the experiment. They were given three attempts to get their answers right, and those who still had incorrect answers after the third attempt were either excluded from participating in the main phase or they were allowed to do so and their data was discarded ex post. Subjects in Experiments 2 and 3 were told from the beginning that they would receive a small extra monetary payment if they answered all quiz questions correctly by the third attempt (£2 and €3, respectively).

Participants in Experiments 1 and 2 were University of St Andrews undergraduate and postgraduate students of no particular field of specialization. Data from a total of 12 and 17 subjects in Experiments 1 and 2, respectively, were discarded due to these individuals’ failure to answer correctly -after the third attempt- all questions in the instructions understanding quiz. Data from 59 subjects in Experiment 1 were also excluded because they featured deferrals in at least one singleton menu. This,
in particular, suggested that these subjects’ seemingly correct understanding of the instructions as manifested in their accurate quiz responses failed to translate into the avoidance of such dominated decisions as deferrals at singleton menus when the main part of the experiment commenced. Finally, in the spirit of Bronars (1987), based on the distributions of WARP violations resulting from Monte-Carlo simulations of one million decision makers choosing uniform-randomly under each of the FC and NFC modes, we determined that 6 subjects in the NFC treatments in each of Experiments 1 and 2 were indistinguishable from noise at the 5% level (two-tailed test). These subjects were therefore also excluded from the analysis. Following all these exclusions there were 84 evaluable subjects in the NFC and 76 in the FC treatments of Experiment 1, while these numbers were 54 and 62 for Experiment 2, respectively.

Participants in Experiment 3 were University of Alicante undergraduate students of no particular field of specialization. Data from a total of 23 subjects were discarded due to quiz failures after the third attempt. There were 100 and 150 evaluable subjects in the FC and NFC treatment, respectively, after these exclusions. The distributions of Binary Congruence violations (binary cycles) that were derived from the choices of one million uniform-random decision makers on an underlying set of six alternatives (see Online Appendix 8) suggest that, under both forced and non-forced choice, a zero-cycle score lies within the 2.5% left tail of these distributions. Therefore, due to a lack of guidance in this respect, noisiness-driven exclusions were not made for this data.

---

25 Most of our results remain significant at the 5% level if these noisy and/or singleton-deferring subjects are also included (see Online Appendix 4).
Subjects in Experiments 2 and 3 were given no additional information about the choice alternatives contained in their randomly selected menus once these menus were determined and before subjects were asked to choose from these menus for the last time. By contrast, following determination of the randomly selected menus, all Experiment 1 subjects were invited to go to the desk at the front of the lab where the five headsets were on display from the start of the session, and to silently inspect the ones in their randomly selected menus and try them out while listening to the music provided by a central source. These subjects then went back to their desks and were asked to choose one of the headsets. The reason why we offered such additional information in our first experiment is motivated by the intuitive idea -already mentioned in the introduction via Sen’s (1997) reference to “tentative incompleteness” that “can arise from limited information”- that an individual is more likely to be unable to declare an alternative (weakly) preferred to another if she lacks the information necessary to make that comparison. At the same time, however, if a subject knows that she can incur a deferral cost in order to acquire more information later about the alternatives in her randomly selected menu, then in principle one might also think of her as choosing to defer or not on the basis of a rational, subjective expected utility calculation with menu-dependent priors over the probability that the arrival of new information on the items of that menu will change her ex ante (complete) preferences over that menu. We come back to this discussion in Section 8.

In addition to information (non-)acquisition following deferrals, another important difference between the three experiments -which is particularly relevant for the otherwise very similarly structured Experiments 1 and 2- were the different monetary values corresponding to the initial endowments, choice-reversal and choice-deferral costs. Compared to Experiment 1, the initial endowment, \( I \), was increased from £7 to £8 (+£2 from the correct quiz responses) to account for inflation between 2013-14 and 2018. At the same time, \( c_r \) and \( c_d \) were increased and decreased by 50% to £6 and £0.50, respectively, resulting in the \( \frac{c_r}{c_d} \) ratio rising from 4 to 12. The increased reversal cost in Experiment 2, in particular, provided stronger incentives for subjects in both treatments to make consistent choices that would not be subsequently reversed. The reduced deferral cost on the other hand reflected the fact that no new information was to be gained by deferring in this experiment. For the same reason, 82 NFC subjects in Experiment 3 faced the \( c_d = €0.50 \) deferral cost while the remaining 68 in that treatment operated under \( c_d = €0.20 \) instead. The reversal cost \( c_r \) in Experiment 3 was unchanged relative to Experiment 1, while the lower initial endowment of €5.50 was balanced out by the €3 quiz top up that was again received by all evaluable subjects.

These disparities notwithstanding, let us emphasize that our primary interest in
this paper is to test the hypothesis that forced-choice subjects are less consistent than non-forced choices ones, and that the channel through which such a forced-choice effect may occur is the subjects’ incomplete preferences that may not have been duly respected by the forced-choice design. Yet, in order for this hypothesis to be tested in a meaningful way, there is a clear need for the NFC design to provide sufficient incentives for deferral whenever subjects are in any doubt about their most preferred alternative in a menu. Taking this into account, and with the benefit of hindsight that the deferral data in Experiment 1 offered us from that environment where additional information about the alternatives was given at a later stage following deferral, we judged that keeping $c_d$ at its original level of £1 without offering more information to subjects would likely have resulted in many of them being unwilling to defer in menus where their preferences might have been of limited guidance to their decision. Having said that, we do acknowledge that the ensuing incomparability between the parametric structure of the two experiments is not optimal when it comes to disentangling the potentially rational from the potentially indecisive part in the deferral data of the two experiments. For the accuracy of such an analysis to be maximised, all aspects of the experiments except the information structure would ideally be held constant.

After the end of the main task in each of the three experiments, finally, subjects were asked to fill in a personality questionnaire that included the *indecisiveness* items in Germeijs and Boeck (2002). The possible responses to these were captured in a 8-level Likert scale and ranged between “strongly disagree” (0) and “strongly agree” (7). Two items in the questionnaire, for example, were “I find it easy to make decisions” and “I don’t hesitate much when making a decision”.

6 Experiments 1 & 2: Findings

6.1 The Use of Deferral

We first study how often subjects in the NFC treatments of Experiments 1 and 2 defer, the dynamics of deferral behaviour, and whether deferrals relate to the number of choice alternatives available at the menu. Figures 2(a-1) and (a-2) display the histograms of the number of deferrals for all NFC subjects in Experiments 1 and 2, respectively. In Experiment 1, subjects defer on average at 4.22 menus, i.e. in about one out of every six menus. Moreover, 33% of these subjects defer at least once, and do so at an average of 12.67 menus, with most subjects deferring at between one and sixteen menus, and 8 subjects deferring at either all or all but one or two menus. Overall, we can thus conclude that one third of the subjects defer choice, and that among those who do so, the rate of deferrals at non-singleton menus is almost one half. Despite the non-
provision of additional information in Experiment 2, subjects defer on average at 5.10 menus, with 41.9% of them actually deferring at least once. Similar to Experiment 1, moreover, the latter subjects defer at an average of 12.15 menus, with most subjects again deferring at between one and sixteen menus, and with 7 subjects deferring at all or all but one or two menus. Therefore, among subjects who do defer, the rate of deferrals here is again almost one half.

We now examine the extent to which deferrals at non-singleton menus depend on menu size. Figures 2(b-1) and (b-2) plot the relative frequency of deferral at menus with 2, 3, 4 and 5 alternatives. Although -as suggested by the corresponding graph- there is no evidence that deferrals are driven by choice overload in the second experiment, consistent with the increasing pattern shown in Figure 2(b-1) there is evidence to suggest the presence of this effect for some subjects in the first experiment (see logit regressions in Online Appendix 7). As will be shown in Section 6.3 below, however, this trend is mainly due to the behaviour of 5 out of the 26 subjects who deferred in this experiment.

Finally, one might wonder if the deferrals observed in these experiments are driven by choice fatigue, i.e., that as the experiment progresses subjects get progressively tired of choosing and, as a consequence, defer choice more often at later menus than at earlier menus.26 In order to examine this hypothesis, we compare the relative frequency of deferrals in the first half of the menus with that of the second half of the menus, which we display in Figures 2(c-1) and (c-2). As can be seen in those figures, the number of subjects that fall into the different bins of deferral rates does not change much from the first 13 to the last 13 non-singleton menus. In addition to this analysis, we also estimated the probability of deferral as a function of the position of the menu in the sequence of all menus using logit regressions. These regressions find that these two variables are not correlated (more details are available in Online Appendix 7). We can thus conclude that there is no evidence of choice fatigue in our data.

6.2 The Effect of Forced Choice on Consistency

Some first evidence pointing in the direction of a negative forced-choice effect on consistency is shown in Figure 3. As demonstrated there, for both Experiments 1 [Figures 3(a),(b)] and 2 [Figures 3(c),(d)], but in a visibly more pronounced way in the case of the latter experiment, the distributions of violations in the FC treatment of revealed-preference cycles in the sense of WARP and Binary Congruence first-order stochastically dominate those in the NFC one. That is, for any fixed number of violations of these axioms, the probability of a subject exhibiting a larger number of violations is

Figure 2: The use of deferral in Experiments 1 (left column) and 2.

(a) Deferral relative frequencies

(b) Average deferral relative frequencies (± standard errors) sorted according to menu size

(c) Deferral relative frequencies in 1st & 2nd half of the 26 decisions
higher in FC than in NFC, and typically strictly so.

Figure 3: Consistency violations in the forced-choice treatment first-order stochastically dominate those in the non-forced choice treatment.

Table 4 shows the proportions of subjects in the FC and NFC treatments of the two
experiments that exhibit violations of WARP, Congruence and Binary Congruence. The differences are large and significant for WARP and Congruence and less so for binary cycles, probably due to the low frequency of the latter. Once data from both experiments are aggregated and separated per treatment, however, the differences in proportions also become significant at the 10% level for such cycles too. In addition, Table 5 displays subjects’ average number of WARP, Congruence and Binary Congruence violations in each treatment, both when these are normalized (non-)linearly (middle and right columns) and when they are not (left column). In almost all of the several subcases reported in this table the differences between FC and NFC subjects are statistically significant, although less so for Binary Congruence, again probably due to low frequency of this type of cycles and the associated low power of the corresponding test.\textsuperscript{27} Taken together, the results reported in these two tables strongly suggest that subjects who were forced to choose in this experiment made significantly less consistent active choices than those who were not.

Table 4: Proportions of subjects violating WARP & Binary Congruence within/across experiments and treatments. \(p\)-values from 2-tailed Fisher exact tests.

(a) Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>FC</th>
<th>NFC</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARP</td>
<td>35/76 (46.05%)</td>
<td>24/84 (28.57%)</td>
<td>0.033</td>
</tr>
<tr>
<td>Congruence</td>
<td>35/76 (46.05%)</td>
<td>25/84 (29.76%)</td>
<td>0.049</td>
</tr>
<tr>
<td>Binary Congruence</td>
<td>6/76 (7.89%)</td>
<td>2/84 (2.38%)</td>
<td>0.152</td>
</tr>
</tbody>
</table>

(b) Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>FC</th>
<th>NFC</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARP</td>
<td>22/54 (40.75%)</td>
<td>12/62 (19.35%)</td>
<td>0.014</td>
</tr>
<tr>
<td>Congruence</td>
<td>22/54 (40.75%)</td>
<td>12/62 (19.35%)</td>
<td>0.014</td>
</tr>
<tr>
<td>Binary Congruence</td>
<td>5/54 (9.26%)</td>
<td>2/62 (3.23%)</td>
<td>0.248</td>
</tr>
</tbody>
</table>

(c) Experiments 1 & 2

<table>
<thead>
<tr>
<th></th>
<th>FC</th>
<th>NFC</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARP</td>
<td>57/130 (43.85%)</td>
<td>36/146 (24.65%)</td>
<td>0.001</td>
</tr>
<tr>
<td>Congruence</td>
<td>57/130 (43.85%)</td>
<td>37/146 (25.34%)</td>
<td>0.001</td>
</tr>
<tr>
<td>Binary Congruence</td>
<td>11/130 (8.46%)</td>
<td>4/146 (2.74%)</td>
<td>0.059</td>
</tr>
</tbody>
</table>

We note, finally, that although the proportions of WARP violators are similar in the two FC treatments (41% and 46%, respectively), a lower proportion of subjects

\textsuperscript{27}Significance is lost at the margin in the highly lenient towards the subjects' consistency environment whereby Experiment 1 choices are indifference-augmented and WARP violations are normalized to account for possible non-linearities between the size of menus in which they deferred and the maximum possible violations that could have been generated by them conditional on these deferrals. Given the preceding discussion about our indifference-elicitation and normalization methods, however, one may interpret this test as the most favourable towards subjects and one that sets an upper bound on the \(p\)-value for the existence of a forced-choice effect.
violates WARP in the NFC treatment of the second experiment compared to that of the first (19% vs 29%, respectively). This results in a bigger and more significant choice-reversal effect in that experiment. Although the deferral rates of subjects who deferred at least once are similar in the two NFC treatments (around 12.5), this bigger WARP effect might be attributed to the larger proportions of deferring subjects in Experiment 2 (42% vs 33%).

6.3 Model-Based Goodness of Fit Analysis

We now proceed to reporting the findings from the model distance-score analysis that was presented in Section 4.1 for the six models outlined in Section 4.2. Before we do so, we make three additional remarks about some important aspects of this analysis in the present setting.

Remark 2. For the sake of completeness, and in order to compare distance scores across treatments, we conduct the same distance-score analysis in both the FC and NFC treatments. However, we note that this analysis is biased in favour of UM when applied to the FC treatment because all models except UM prescribe deferral in at least some menus. Since deferral is not possible in this treatment by design, such instances are always counted as “mistakes” when fitting these models to the data.

Remark 3. Unlike the case of all the other models that were outlined above, (active) choices in the models of undominated and partially dominant choice generally violate WARP and Binary Congruence. They are nevertheless included in our analysis so that we can test if some subjects’ behaviour may be best explained by a model of incomplete preferences that also features bounded rationality and context-dependence. We emphasize, however, that our experiment is not well suited to test these models because they specify non-singleton sets of choosable alternatives in at least some of the menus (unless they collapse to the model of strict utility maximization, of course). Given this fact, we concede from the outset that our distance-score analysis of these two models is inevitably biased against them because it uses single-valued active-choice data, in which case any dataset that can be perfectly explained by either of them can also be perfectly explained by either UM or MDC.28

Remark 4. To understand if the best distance scores that we obtained using this method meaningfully reflect the accuracy of the corresponding model as a potential

28One may wonder if this bias is eliminated once we carry out our distance-score analysis with indifference-augmented choice data instead (relevant for Experiment 1 only). The answer to this question is negative because the construction of choice correspondences using our method extrapolates the indifference-augmented data from binary choices to all larger menus where one of these two alternatives is chosen. It therefore introduces multi-valuedness in choice in a way that generates a transitive indifference relation by construction. Once this restriction is imposed, however, the models of undominated (UC) and partially dominant (PDC) choice become observationally equivalent to the model of utility maximization that may also feature a non-trivial indifference relation. Having acknowledged these limitations, we note finally that the unrestricted distance-score analysis of models of multi-valued choice will be conducted in future work.
Table 5: Average consistency violations per experiment and treatment.

p-values from 2-tailed Mann-Whitney U tests.

(a) Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>WARP violations</th>
<th>Linear-normalized</th>
<th>Nonlinear-normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ Indifference</td>
<td>- Indifference</td>
<td>+ Indifference</td>
</tr>
<tr>
<td>NFC n =</td>
<td>1.27</td>
<td>1.46</td>
<td>0.051</td>
</tr>
<tr>
<td>p-value =</td>
<td>0.050</td>
<td>0.014</td>
<td>0.077</td>
</tr>
<tr>
<td>FC n =</td>
<td>3.32</td>
<td>3.64</td>
<td>0.13</td>
</tr>
<tr>
<td>p-value =</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
</tr>
</tbody>
</table>

(b) Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>WARP violations</th>
<th>Linear-normalized</th>
<th>Nonlinear-normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ Indifference</td>
<td>- Indifference</td>
<td>+ Indifference</td>
</tr>
<tr>
<td>NFC n =</td>
<td>1.10</td>
<td>0.052</td>
<td>0.003</td>
</tr>
<tr>
<td>p-value =</td>
<td>0.003</td>
<td>0.008</td>
<td>0.008</td>
</tr>
</tbody>
</table>

30
descriptor of the relevant subject’s behaviour, we carried out Monte-Carlo simulations and found the 2.5% distance score percentiles for 100,000 uniform random-behaving subjects, both when the deferral outside option was (NFC) and was not (FC) available to them (Table 6).

Table 6: The six models’ 2.5% distance score percentiles (cut-offs) derived from the distributions of 100,000 simulated uniform-random choice datasets on the 26 non-singleton menus derived from a set of 5 alternatives.

<table>
<thead>
<tr>
<th></th>
<th>Utility Maximization</th>
<th>Utility Maximization with an Outside Option</th>
<th>Undominated Choice with Incomplete Preferences</th>
<th>Maximally Dominant Choice with Incomplete Preferences</th>
<th>Partially Dominant Choice with Incomplete Preferences</th>
<th>Overload-Constrained Utility Maximization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(UM)</td>
<td>(UMOO)</td>
<td>(UC)</td>
<td>(MDC)</td>
<td>(PDC)</td>
<td>(OCUM)</td>
</tr>
<tr>
<td>FC</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>NFC</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>

These cut-off values can then be used to determine if the fact that a given subject’s best distance score that is derived from one of the six models is 2, for example, could be thought of suggesting that this model is a reasonably good approximation of that subject’s behaviour or whether this score is indistinguishable in a statistical sense from what could be achieved by a random-behaving individual.

**Remark 5.** In general, more than one model may be associated with the same minimum-distance score for a given subject. Our primary categorization breaks such ties in the following ways:

1. UM favoured over all other models.
2. All models favoured over UMOO.
3. MDC favoured over PDC and OCUM.

In the first case tie-breaking follows Occam’s razor and favours the hypothesis that the relevant subject is rational in the standard sense. In the second case it is motivated by the fact that behaving according to utility maximization with an outside option in the context of our experimental design leads to dominated decisions and is therefore non incentive-compatible. In the third case, finally, tie breaking favours the maximally dominant choice model because its prescribed behaviour is more closely aligned with the experiment’s incentives than that corresponding to the other two models. This is more so for the comparison between maximally and partially dominant choice than between the former model and overload-constrained utility maximization. More specifically, in the latter comparison it is possible that the subjects’ cognitive costs from deliberation in larger menus are sufficiently high to exceed the deferral cost. Our tie-breaking rule that favours MDC reflects related empirical findings suggesting that
choice overload is more relevant in menus that contain more than the maximum number of five alternatives that have been included in our study. In any case, however, the finer classification within the MDC category when this third tie-breaking rule is dropped is presented at the end of this section.

Table 7: Average model distance scores and average best distance scores within/across experiments and treatments.

<table>
<thead>
<tr>
<th></th>
<th>UM</th>
<th>UMOO</th>
<th>UC</th>
<th>MDC</th>
<th>PDC</th>
<th>OCUM</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 - All (160)</td>
<td>2.86</td>
<td>2.71</td>
<td>3.74</td>
<td>1.83</td>
<td>2.15</td>
<td>2.03</td>
<td>0.90</td>
</tr>
<tr>
<td>Experiment 1 - FC (76)</td>
<td>0.89</td>
<td>1.89</td>
<td>1.89</td>
<td>1.89</td>
<td>1.89</td>
<td>1.83</td>
<td>0.89</td>
</tr>
<tr>
<td>Experiment 1 - NFC (84)</td>
<td>4.64</td>
<td>3.44</td>
<td>5.40</td>
<td>1.77</td>
<td>2.38</td>
<td>2.21</td>
<td>0.90</td>
</tr>
<tr>
<td>Experiment 2 - All (116)</td>
<td>3.41</td>
<td>1.24</td>
<td>4.23</td>
<td>1.93</td>
<td>1.97</td>
<td>2.72</td>
<td>1.06</td>
</tr>
<tr>
<td>Experiment 2 - FC (54)</td>
<td>1.13</td>
<td>1.13</td>
<td>2.13</td>
<td>2.13</td>
<td>2.13</td>
<td>2.09</td>
<td>1.13</td>
</tr>
<tr>
<td>Experiment 2 - NFC (62)</td>
<td>5.40</td>
<td>1.34</td>
<td>6.06</td>
<td>1.76</td>
<td>1.82</td>
<td>3.27</td>
<td>1.00</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - All</td>
<td>3.09</td>
<td>2.09</td>
<td>3.95</td>
<td>1.87</td>
<td>3.59</td>
<td>2.32</td>
<td>0.97</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - FC</td>
<td>0.99</td>
<td>1.58</td>
<td>1.99</td>
<td>1.99</td>
<td>1.99</td>
<td>1.94</td>
<td>0.99</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - NFC</td>
<td>4.97</td>
<td>2.55</td>
<td>5.68</td>
<td>1.77</td>
<td>5.02</td>
<td>2.66</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 8: Proportions of subjects that are best explained by one of the six models within/across experiments and treatments with maximal tie-breaking (proportion of perfect model fits in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>UM</th>
<th>MDC</th>
<th>PDC</th>
<th>OCUM</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 - FC</td>
<td>100% (53.95%)</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>100% (53.95%)</td>
</tr>
<tr>
<td>Experiment 1 - NFC</td>
<td>70.24% (45.24%)</td>
<td>21.43% (10.71%)</td>
<td>2.38%</td>
<td>5.95% (0.00%)</td>
<td>100% (55.95%)</td>
</tr>
<tr>
<td>Experiment 1 - All</td>
<td>84.38% (49.37%)</td>
<td>11.25% (5.62%)</td>
<td>1.25%</td>
<td>3.12% (0.00%)</td>
<td>100% (55.00%)</td>
</tr>
<tr>
<td>Experiment 2 - FC</td>
<td>100% (59.26%)</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>100% (59.25%)</td>
</tr>
<tr>
<td>Experiment 2 - NFC</td>
<td>62.90% (28.53%)</td>
<td>30.65% (11.29%)</td>
<td>4.84% (0.00%)</td>
<td>1.61% (0.00%)</td>
<td>100% (58.06%)</td>
</tr>
<tr>
<td>Experiment 2 - All</td>
<td>80.17% (38.79%)</td>
<td>16.38% (19.82%)</td>
<td>2.59% (0.00%)</td>
<td>0.86% (0.00%)</td>
<td>100% (58.62%)</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - FC</td>
<td>100% (56.15%)</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>100% (56.15%)</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - NFC</td>
<td>67.12% (45.89%)</td>
<td>25.34% (10.96%)</td>
<td>3.42% (0.00%)</td>
<td>4.11% (0.00%)</td>
<td>100% (56.84%)</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - All</td>
<td>82.61% (50.72%)</td>
<td>14.13% (5.80%)</td>
<td>1.09% (0.00%)</td>
<td>2.17% (0.00%)</td>
<td>100% (56.52%)</td>
</tr>
</tbody>
</table>

Table 9: Average distance score for each best model within/across experiments and treatments.

<table>
<thead>
<tr>
<th></th>
<th>UM</th>
<th>MDC</th>
<th>PDC</th>
<th>OCUM</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 - FC</td>
<td>0.89</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.89</td>
</tr>
<tr>
<td>Experiment 1 - NFC</td>
<td>0.56</td>
<td>1.39</td>
<td>1.50</td>
<td>3.00</td>
<td>0.90</td>
</tr>
<tr>
<td>Experiment 2 - FC</td>
<td>1.13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.13</td>
</tr>
<tr>
<td>Experiment 2 - NFC</td>
<td>0.38</td>
<td>1.37</td>
<td>5.33</td>
<td>5</td>
<td>1.00</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - FC</td>
<td>0.99</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0.99</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - NFC</td>
<td>0.49</td>
<td>1.38</td>
<td>3.80</td>
<td>3.33</td>
<td>0.95</td>
</tr>
</tbody>
</table>

The distance-score computational results for the data from Experiments 1 and 2 are reported in Tables 7, 8 and 9. Table 7 shows the average distance score for each of the six models, both for the two experiments separately and when their data are...
pooled. This information is presented per treatment and across treatments for each of these cases in a separate row. The last entry in each such row shows the average best score that this six model-strong analysis achieves in the collection of choice datasets that corresponds to that row once every subject is assigned to the distance score of her best model. Table 8 reproduces the general structure of Table 7 but reports the proportions of subjects in the relevant sample/row that are best explained by each model, and focuses only on the four models with non-zero explanatory power. Finally, Table 9 reproduces the general structure of Table 8 and goes back to the models’ distance scores, this time reporting the average best distance score of each model, defined as the average distance score of the subjects for whom this model was declared best.

Several interesting results emerge from this analysis. First, the combined force of two out of the six models of deterministic preference and choice that have been included in these computations (namely, UM and MDC, and to a much lesser extent PDC and OCUM) perform remarkably well and explain the subjects’ behaviour in each of the two treatments with an average of just a single deviation from perfect conformity with their best-matching model. Second, standard UM is the best model for 70% and 63% of NFC subjects in Experiment 1 and Experiment 2, respectively. So overall, two thirds of NFC subjects are best explained by standard UM. As noted in Remark 2, the FC treatment by design does not allow non-UM models to offer a good fit to the data, which explains the fact that all FC subjects are categorized as standard UM. Third, MDC is the best model for 24% of all NFC subjects in Experiment 1 (average score: 1.50) and 31% of all such subjects in Experiment 2 (average score: 1.37), hence for 27% (average score: 1.44) of all NFC subjects overall. Given that both UM and MDC are models of WARP and Binary Congruence-consistent active choices, the latter two findings at one level attest to the effectiveness of our experimental design to motivate consistent and preference-guided behaviour for the majority of subjects in these two experiments. At the same time, they also point towards the existence of a non-trivial proportion of subjects with incomplete preferences over the five headsets that comprised the domain of choice in these two experiments. In addition, they highlight the descriptive relevance of the model of maximally dominant choice with incomplete preferences to explain these subjects’ behaviour, and the potentially very helpful role that non-forced choice experimental designs could play in the effort to elicit such incompleteness. Finally, consistent with Figure 2(b-1) and the relevant discussion, 5 out of the 84 (1 out of 62) NFC subjects in Experiment 1 (2) are best explained by OCUM, and 3 out of the 62 NFC subjects in Experiment 2 are best explained by PDC (2 out of 84 in Experiment 1). In both these cases the average distance scores are higher than for the other models.
but nevertheless admissible for all subjects.

Elaborating further on these results, it is of particular interest that the relative majority of subjects who are best explained by UM (across experiments and treatments) are in fact perfect utility maximisers (cf. 51% vs 32%). It is similarly interesting that almost half of all NFC subjects that are best explained by the MDC model also conform with this model perfectly (cf. 11% vs 16%).

Table 10: Proportions of NFC subjects that are best explained by MDC—including jointly with other models—after minimal tie-breaking (proportion of perfect model fits in parenthesis).

<table>
<thead>
<tr>
<th></th>
<th>MDC</th>
<th>MDC</th>
<th>MDC</th>
<th>MDC</th>
<th>MDC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PDC</td>
<td>PDC</td>
<td>UMOO</td>
<td>PDC</td>
<td>UMOO</td>
</tr>
<tr>
<td>Experiment 1 - NFC</td>
<td>11.90% (5.95%)</td>
<td>3.57% (1.19%)</td>
<td>0.00% (0.00%)</td>
<td>5.95% (3.57%)</td>
<td>0.00% (0.00%)</td>
</tr>
<tr>
<td>Experiment 2 - NFC</td>
<td>9.68% (1.19%)</td>
<td>3.23% (0.00%)</td>
<td>11.29% (3.23%)</td>
<td>0.00% (0.00%)</td>
<td>6.45% (6.45%)</td>
</tr>
<tr>
<td>Experiments 1 &amp; 2 - NFC</td>
<td>10.96% (4.11%)</td>
<td>1.37% (0.00%)</td>
<td>6.85% (2.05%)</td>
<td>3.42% (2.05%)</td>
<td>2.74% (2.74%)</td>
</tr>
</tbody>
</table>

Relaxing now the third tie-breaking rule allows for a detailed decomposition of subjects who were classified as MDC decision makers under the above categorization. In particular, while 11% of all NFC subjects in the two experiments are uniquely best explained by that model (and 4% perfectly so), the remaining 14% are jointly explained by a combination of MDC and other non-UM models. It is worth pointing out that the relatively large proportion of ties is not so much a reflection of generally poor separation between these models but rather a consequence of the somewhat extreme behaviour of most of these subjects. In particular, subjects which tied to three or more models made at most two active choices in the experiment, which, inevitably, makes model identification difficult.

Finally, we observe that the average distance score of NFC subjects who are best described as utility maximisers is much lower than that of their FC counterparts: 0.56 vs 0.89 in Experiment 1, and 0.38 vs 1.13 in Experiment 2. Noting first that all subjects who were declared as utility maximisers by this analysis were actually ones who never deferred (whether they belonged to the FC treatment or not), this sharp contrast between those averages reinforces the forced-choice findings that we reported in Section 6.2. Specifically, they suggest that conditional on a subject making active choices at all menus, her behaviour is more likely to be closer to the UM model if she is in the NFC rather than in the FC treatment. Put differently, subjects who self-force themselves to always choose a market alternative are more likely to reveal a complete and transitive preference ordering from their choices than ones who are exogenously forced to always choose.
7 Experiment 3: Findings

7.1 The Use of Deferral

Figure 4(a) presents a histogram with the number of deferrals for all 150 NFC subjects in Experiment 3. Although no additional information about the six lotteries was provided in the interim stage between the main part of the experiment and when subjects had to make their payoff-relevant choice from their randomly selected menu, subjects on average defer at 1.44 menus (9.6%), with 36.7% of them deferring at least once. Subjects who defer do so at an average of 3.93 menus, with most subjects deferring at between one and eight menus, and with 6 subjects deferring at between eleven and all fifteen menus. Therefore, among subjects who do defer, the rate of deferrals is just above one quarter.

Once again, we investigate choice fatigue by comparing the relative frequency of deferrals in the first half of the menus with that of the second half of the menus, as displayed in Figure 4(b). As can be seen in the figure, the number of subjects that fall into the different bins of deferral rates again does not change much from the first 7 to the last 7 non-singleton menus (the median menus were excluded from this analysis). Logit regression estimates of the probability of deferral as a function of the position of the menu in the sequence of all menus again found that these variables are not correlated (more details are available in Online Appendix 7).

7.2 The Effect of (Self-)Forced Choice on Consistency

Since datasets here comprise choices from binary menus only, WARP violations cannot be made. We therefore focus on violations of Binary Congruence. First, in Figure 5 we present the cumulative densities of the distributions of Binary Congruence violations in the FC and NFC treatment. Unlike Experiments 1 and 2, the former distribution no longer first-order stochastically dominates the latter, and neither does dominance exist in the opposite direction. Moreover, as is also visible by contrasting Figures 3(b),(d) and Figure 5, a very important difference between the two real-goods experiments and this lottery-based one is that the vast majority of subjects in both treatments exhibit binary cycles (with some in each treatment having more than 25), and that half of all subjects have more than 3 cycles. Interestingly, FC violations do first-order stochastically dominate NFC ones if the range of binary cycles is restricted to not exceed 8 but this is actually reversed for the range 16 – 28. All but two NFC subjects with binary cycles in this range never deferred, and the two who did defer only did so once.
Figure 4: The use of deferral in Experiment 3

(a) Deferral relative frequencies

(b) Deferral relative frequencies in the first & second half of the 15 decisions
Furthermore, the first part of Table 11 shows the proportions of subjects in the FC and NFC treatments that exhibit binary choice cycles, while the second and third parts, respectively, present these proportions for subjects within the NFC treatment who did and did not defer, and for FC subjects and deferring NFC ones. Although the inconsistent subjects are indeed relatively more frequent in the FC treatment, this difference is not significant. Interestingly, however, unique to this experiment is the finding of a large and highly significant difference in the proportion of Binary Congruence violators between deferring and non-deferring NFC subjects. Similarly, there are significantly more inconsistent subjects in the FC treatment than in the subset of NFC subjects who did make use of deferral at least once.

Table 11: Proportions of subjects violating Binary Congruence across: (1) FC and NFC subjects; (2) (non-)deferring NFC subjects; and (3) FC and deferring NFC subjects in Experiment 3. p-values from 2-tailed Fisher exact tests.

<table>
<thead>
<tr>
<th></th>
<th>FC</th>
<th>NFC</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>79/100 (79%)</td>
<td>110/150 (73.33%)</td>
<td>0.368</td>
</tr>
<tr>
<td>NFC: non-deferring</td>
<td>NFC: deferring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76/95 (80%)</td>
<td>34/55 (61.81%)</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FC</td>
<td>NFC: deferring</td>
<td></td>
</tr>
<tr>
<td>79/100 (79%)</td>
<td>34/55 (61.81%)</td>
<td>0.025</td>
<td></td>
</tr>
</tbody>
</table>

The first part of Table 12 moreover shows subjects’ average and median (normalized) Binary Congruence violations in each treatment, both for single-valued raw and for possibly multi-valued indifference-augmented data. The second and third parts do
so for deferring and non-deferring NFC subjects, and for FC and non-deferring NFC subjects, respectively. Consistent with the above findings, the differences between FC and NFC subjects - although almost invariably in the expected direction - are not statistically significant, in contrast to the large and highly significant differences between deferring and non-deferring NFC subjects, and also between FC and non-deferring NFC subjects.\footnote{Differences in the proportions of violators as well as in the mean and median numbers of violations between deferring and non-deferring NFC subjects are generally in the same direction also in Experiments 1 and 2 for both measures of consistency analysed there. In Experiment 1, where the sample size is larger, some of these differences are also significant (see Online Appendix 6 for more details).}

Despite the absence of a forced-choice treatment effect in Experiment 3, the analysis of the NFC data provides a different insight on the mitigating effect of deferral on active-choice inconsistency by contrasting the behaviour of subjects who made use of that option and those who, despite also being able to do so, opted for self-forced choice instead. A possible interpretation of the latter group’s behaviour is that they are overconfident about their capacity for consistent decision making, or, equivalently, that they are at least partially unaware of their limitations in this regard.\footnote{For a representation theorem linking incompleteness of a decision maker’s preferences over uncertain acts with confidence in her beliefs over the relevant state space we refer the reader to Hill (2016). Hill (2015), moreover, studies axiomatically the behaviour of such a decision maker that operates in an environment of choice under uncertainty and who may opt for costly deferral when her confidence in the relevant beliefs is low. Although not unrelated, the kind of}
in an environment with incentives for subjects to either choose market alternatives in a consistent, preference-guided manner or defer choice when that’s not possible, the noisy behaviour of non-deferring subjects might be thought of as pointing towards a mistaken belief that their preferences under risk were sufficiently well-defined to prevent them from making cyclic choices. Put differently, to the extent that these subjects believed that deferring would be useless to them because they were capable of making consistent preference-guided choices at all menus, this belief is contradicted by their actual behaviour.\textsuperscript{31}

In addition, although a direct forced-choice treatment effect is not found in this data, focusing on the comparison between FC subjects and those NFC ones who deferred does reveal a significant difference in binary-choice consistency, both in terms of the proportions of inconsistent subjects and also in terms of the distribution of binary cycles. We emphasize, however, that this should not be interpreted as evidence of a treatment effect because deferring NFC subjects are a selected subsample. Nevertheless, this finding is relevant because it suggests that, in our data, subjects who are forced to choose are expected to be significantly less consistent than subjects who are not, conditional on actually choosing to defer at least once.\textsuperscript{32}

7.3 Model-Based Goodness of Fit Analysis

We note from the outset that the scope of the distance-score method is much more limited in datasets such as those derived from our Experiment 3 where only choices from binary menus are available. In particular, Monte-Carlos simulations for the UM, UC and MDC models that are relevant in such datasets\textsuperscript{33} suggest that any subject in this experiment whose data generate a positive distance score for any of these models is statistically indistinguishable from a random-behaving decision maker in such a binary-choice environment.\textsuperscript{34} This limitation notwithstanding, the method remains applicable and informative, especially through the detection of those subjects whose behaviour conforms perfectly with one of these models.

Tables 13–15 are analogous to Tables 7–9 in Section 6.3. The first feature of our overconfidence that we are postulating here is distinct from the notions of confidence in beliefs studied in Hill’s work because the decision maker in those models cannot exhibit cyclic active choices.

\textsuperscript{31}Although we have not elicited the subjects’ beliefs about how well-defined preferences they have and how consistently they can choose in their corresponding experimental environments that would have allowed for this possible interpretation to be empirically tested, we think this is a potentially interesting problem for future work.

\textsuperscript{32}Such effects are also present in Experiments 1 and 2, which is not surprising given that we do have treatment effects in that data.

\textsuperscript{33}Specifically, we have a coincidence of the distance scores between UM and UM00 and also between MDC and PDC with such data in the sense that whenever one model in each of these pairs has the minimum distance score, so does the other. Moreover, the OCUM model is implausible in this binary environment.

\textsuperscript{34}The 2.5% distance score percentiles (cut-offs) derived from the distributions of 100,000 simulated uniform-random choices on the 15 binary menus derived from a set of 6 alternatives are as follows: UM (FC): 1; UM (NFC): 3; UC (FC): 2; UC (NFC): 3; MDC (FC): 2; MDC (NFC): 1
results worth highlighting is that, as with binary cycles, average distance scores are substantially larger in Experiment 3 than in the previous two experiments. As Table 13 shows, the best distance score achieved by subjects averages 1.46, while this number was 0.96 for Experiments 1 and 2 together. Bearing in mind that these experiments involved 26 non-singleton menus with up to 5 alternatives while Experiment 3 only involves 15 binary menus suggests that subjects had greater difficulty making consistent choices in the latter experiment. A natural interpretation of this is that subjects find it easier to make decisions about familiar consumer goods (such as headsets) than about abstract money lotteries. This observation notwithstanding, an average distance score of 1.46 implies that subjects on average make between 1 and 2 mistakes out of 15 decisions, which is still relatively little.

Table 13: Average model distance scores and average best distance scores within/across treatments.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>UM</th>
<th>UC</th>
<th>MDC</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 3 - FC</td>
<td>1.32</td>
<td></td>
<td></td>
<td>1.32</td>
</tr>
<tr>
<td>Experiment 3 - NFC (all)</td>
<td>2.67</td>
<td>3.38</td>
<td>2.21</td>
<td>1.56</td>
</tr>
<tr>
<td>Experiment 3 - NFC (deferring, n = 55)</td>
<td>4.73</td>
<td>4.93</td>
<td>1.75</td>
<td>1.69</td>
</tr>
<tr>
<td>Experiment 3 - NFC (non-deferring, n = 95)</td>
<td>1.48</td>
<td>2.48</td>
<td>2.48</td>
<td>1.48</td>
</tr>
<tr>
<td>Experiment 3 - All</td>
<td>2.13</td>
<td>2.95</td>
<td>2.25</td>
<td>1.46</td>
</tr>
</tbody>
</table>

Table 14: Proportions of subjects that are best explained by a model across treatments and within the NFC treatment (proportion of perfect model fits in parenthesis).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>UM</th>
<th>MDC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 3 - FC</td>
<td>100% (21.00%)</td>
<td>0.00%</td>
<td>100% (21.00%)</td>
</tr>
<tr>
<td>Experiment 3 - NFC (all)</td>
<td>68.67% (12.67%)</td>
<td>31.33% (7.33%)</td>
<td>100% (20.00%)</td>
</tr>
<tr>
<td>Experiment 3 - NFC (deferring)</td>
<td>14.55% (0.00%)</td>
<td>85.45% (20.00%)</td>
<td>100% (20.00%)</td>
</tr>
<tr>
<td>Experiment 3 - NFC (non-deferring)</td>
<td>100% (20.00%)</td>
<td>0.00%</td>
<td>100% (20.00%)</td>
</tr>
<tr>
<td>Experiment 3 - All</td>
<td>81.20% (16.00%)</td>
<td>18.80% (4.40%)</td>
<td>100% (20.40%)</td>
</tr>
</tbody>
</table>

Table 15: Average distance score for each best model across treatments and within the NFC treatment.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>UM</th>
<th>MDC</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 3 - FC</td>
<td>1.31</td>
<td></td>
<td>1.31</td>
</tr>
<tr>
<td>Experiment 3 - NFC (all)</td>
<td>1.57</td>
<td>1.53</td>
<td>1.56</td>
</tr>
<tr>
<td>Experiment 3 - NFC (deferring)</td>
<td>2.63</td>
<td>1.53</td>
<td>1.69</td>
</tr>
<tr>
<td>Experiment 3 - NFC (non-deferring)</td>
<td>1.48</td>
<td></td>
<td>1.48</td>
</tr>
</tbody>
</table>

Turning to our primary concern with this analysis, we find that 69% and 31% of NFC subjects are categorized as UM and MDC, respectively, while as many as 85% of deferring subjects are categorized as MDC. As anticipated by our preceding discussion, the proportion of subjects perfectly explained by any of the models is relatively low (20%) compared to Experiments 1 and 2 (57%). Nevertheless, it is
worth noting that, among deferring subjects, 23% are perfectly explained by MDC. This again highlights the descriptive relevance of the model of maximally dominant choice with incomplete preferences to explain the behaviour of a non-trivial proportion of subjects in Experiment 3.

8 Discussion

8.1 On the Nature of the Forced-Choice Effect

Our analysis of the data from all three experiments generally suggests that experimental participants who are (possibly self-)forced to choose are significantly more likely to make inconsistent active choices than those who are not. In particular, the comparison of deferring behaviour (Figures 2, 4 and Table 16) and consistency (Tables 4, 5, 11, 12) across the three experiments indicates that, when an FC-NFC treatment effect occurs (cf. Experiments 1, 2), it is associated with significantly higher deferral frequencies in the NFC treatment than when it does not (cf. Experiment 3). At the same time, our findings from Experiment 3 show that subjects who seize the opportunity to delay choice make significantly more consistent active choices than those who are forced to choose, or those who, despite given the opportunity to delay choice, are self-forced to choose. All these findings support the mediating role of deferral in the occurrence of forced-choice effects.

Table 16: Deferring subjects and average deferrals across the three experiments.

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>28/84 (33.33%)</td>
<td>26/62 (41.94%)</td>
<td>55/150 (36.66%)</td>
</tr>
<tr>
<td>Deferrals</td>
<td>4.23/26 (16.25%)</td>
<td>5.10/26 (19.61%)</td>
<td>1.44/15 (9.6%)</td>
</tr>
</tbody>
</table>

This effect notwithstanding, we also stress that being able to defer choice opens up the possibility for a decision maker’s behaviour to be inconsistent in ways that are qualitatively different from active-choice cycles. In the marketing research literature, for example, Dhar and Simonson (2003) found that the well-known asymmetric dominance/attraction/decoy effect is strengthened if choice is not forced because, even though the vast majority of people still choose the target option \( y \) from the menu \( \{x, x', y\} \), a significant minority defer at the menu \( \{x, y\}\).35 This behaviour can be captured deterministically by

\[
C(\{x, x', y\}) = \{y\} \quad \text{and} \quad C(\{x, y\}) = \emptyset.
\]  

(7)

35Here, the alternatives \( x, x', y \) are multi-attribute and such that \( x' \) is dominated by \( x \) in all dimensions while \( x, y \) and \( x', y \) cannot be ranked by the standard dominance relation. The non-forced choice version of the PDC model that was presented in Section 4 provides an incomplete-preference deterministic explanation for this finding.
Importantly, however, although (7) features a violation of Independence of Irrelevant Alternatives, it is not a WARP violation as it does not involve a choice reversal between any market alternatives. It is nevertheless associated with a milder form of inconsistency in the decision maker’s revealed preferences whereby, by observing the decisions at \( \{x, x', y\} \) and \( \{x, y\} \), the analyst concludes that \( x \succ_R y \) and \( x \not\succ_R y \), respectively. This is in sharp contrast to the more severe inconsistency \( x \succ_R y \) and \( y \succ_R x \) that would have been inferred from a standard forced-choice attraction effect. It is also worth noting that, unlike active-choice reversals which in theory render a decision maker vulnerable to exploitations of a money-pump type, such exploitation is not possible for this deferral-specific class of non-reversal inconsistencies.

Similarly, choice-overload effects in deferral-permitting environments are manifested in the pattern

\[
x \in C(A) \cap C(B) \quad \text{and} \quad C(A \cup B) = \emptyset.
\]

Here too, although revealed preferences are inconsistent because the decisions at \( A \) and \( B \) suggest \( x \succ_R y \) for each \( y \) in \( A \) or \( B \) and the decision at \( A \cup B \) suggests \( x \not\succ_R y \) instead, this behaviour is also not associated with any active-choice reversals. Therefore, since our analysis is limited to the comparison of consistency in people’s active choices (indeed, (7) and (8) are impossible in a forced-choice environment), it is important to emphasize that our claim about the presence of a negative forced-choice effect on consistency is in that sense only, and hence does not contradict the existence of these distinct types of behavioural inconsistencies.\(^{36}\)

We note, finally, that although the prediction of a negative forced choice effect on consistency such as the one we documented here is implicit in Luce and Raiffa (1957) and Gerasimou (2018), our paper appears to be the first in the literature to test and find evidence for it. On the other hand, studies that have used non-experimental data on health insurance, pension savings and voting decisions and which have shown, among other things, that decision makers often choose their non-market outside option include Samuelson and Zeckhauser (1988), Madrian and Shea (2001), Iyengar, Huberman, and Jiang (2004), Carroll, Choi, Laibson, Madrian, and Metrick (2009) and Augenblick and Nicholson (2016). In some of these studies this behaviour is induced by the relatively large number of options that were available to decision makers, which is known to be associated with overload-driven deferral and which, as previously shown, is not relevant for most of our data. In addition, findings from non-incentivized experiments with non-forced choice in “small” menus of multi-attribute, tradeoff-generating alternatives where significant proportions of respondents— in the 20% – 50% range— opted for

\(^{36}\)We further note that such non-reversal inconsistencies as those captured by (7) and (8) also help explain why the “noisiness” cut-off model distance scores that are reported in Table 6 are generally higher in the non-forced choice simulations.
the deferral outside option are also reported in Tversky and Shafir (1992), Redelmeier and Shafir (1995), Dhar (1997), Luce (1998) and Dhar and Simonson (2003) (see also Anderson (2003) and Broniarczyk and Griffin (2014) for relevant surveys).

8.2 On the Nature of Deferral

With regard to the nature of deferral in our experiments, we again acknowledge that, in addition to incomplete preferences, it is theoretically possible that such decisions are ultimately made via a cost-inclusive subjective expected utility calculation whereby a rational agent holds some beliefs about the probability that her current completely ordered preferences over the alternatives in a given menu will change at a later stage. While it is possible that some subjects -even ones that were categorized as MDC- in our experiments deferred on such a basis, it is important to emphasize that, unless the subjects’ internal signals lead to substantial preference-updating across decisions, no significant forced-choice effect would arise if the majority of subjects were of this type. Indeed, assuming that the FC and NFC samples are drawn randomly from the same population, if most subjects in that population are “Bayesians” in the above sense, then those who are assigned to the FC treatment should make utility-maximizing -hence perfectly consistent- choices according to their current, stable preferences. Nevertheless, if one insists on the possibility of substantial preference-updating across decisions, then our approach of categorizing these subjects as incomplete-preference maximizers may be regarded as a convenient “as if” simplification which allows us to extract the stable part of these subjects’ preferences.

An additional source of information which sheds light on the nature of deferral comes from the indecisive personality questionnaire that was mentioned in Section 5. Specifically, focusing on the combined data from the three experiments’ NFC treatments, Table 17 reports the average psychological decisiveness score for subjects in those treatments who did and did not defer, respectively, and shows that the difference is highly significant. Thus, deferring NFC subjects are more indecisive in this sense than non-deferring ones, which is consistent with the incomplete-preferences hypothesis.

Finally, to get a flavour of whether deferrals were driven by decision conflict/difficulty rather than simply by time-saving motivations, it is instructive to look at response times. A natural hypothesis is that subjects spent more time on menus in which they

---

37Eighteen subjects with contradictory responses in the questionnaire were excluded. Specifically, the indecisiveness questionnaire consists of 11 pairs of questions. Within each pair, the questions are similar but phrased in the opposite way so that a pro-decisiveness answer in one question would be 7 and in the other it would be 0. If the answers in a pair differ by more than 2 points, this pair is considered inconsistent, so we do not count it for calculating our decisiveness score for that subject. Finally, we exclude any subject that has less than 7 consistent pairs out of 11.
Table 17: Decisive personality scores (“psychological decisiveness”) between deferring and non-deferring subjects across the three experiments. *p*-value from 2-tailed Mann-Whitney *U* test.

<table>
<thead>
<tr>
<th>NFC: deferring subjects</th>
<th>Average psychological decisiveness score</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>n</em> = 101</td>
<td>0.614</td>
</tr>
<tr>
<td>NFC: non-deferring subjects</td>
<td>0.685</td>
</tr>
<tr>
<td><em>n</em> = 178</td>
<td></td>
</tr>
<tr>
<td><em>p</em>-value</td>
<td>0.007</td>
</tr>
</tbody>
</table>

deferred than on menus in which they made an active choice, because they found the former more difficult and thus required longer deliberation. Although, unfortunately, response time data was not recorded in Experiment 1, it is available for Experiments 2, 3 and is summarized in Table 18.

Table 18: Average response times (in seconds) per subject per menu in the NFC treatments of Experiments 2 and 3.

<table>
<thead>
<tr>
<th>NFC: deferrals</th>
<th>Average response time</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>n</em> = 74</td>
<td>21.66</td>
</tr>
<tr>
<td>NFC: active choices</td>
<td>14.90</td>
</tr>
<tr>
<td><em>n</em> = 74</td>
<td></td>
</tr>
<tr>
<td><em>p</em>-value from one-tailed paired <em>t</em> test (log data)</td>
<td>0.008</td>
</tr>
<tr>
<td><em>p</em>-value from one-tailed Wilcoxon signed-rank test</td>
<td>0.017</td>
</tr>
</tbody>
</table>

In particular, Table 18 shows that average response times were longer at menus in which a deferral took place than at menus in which an active choice was made (approximately 22 vs 15 seconds) and, based on paired *t* and signed-rank tests, the difference is indeed statistically significant. This is consistent with intuition and related recent experimental findings (e.g. Agranov and Ortoleva, 2017) suggesting a higher response time when subjects find a decision problem to be hard. Our finding, in particular, shows that subjects who deferred in these two experiments did not do so “strategically” in the sense of trying to go through the main part of the experiment quickly and effortlessly and then wait until their randomly selected menu was determined in order to finally make a careful choice.

We note that mean response times for deferring and non-deferring NFC subjects are similar across the two experiments and the reported finding is in the same direction for each experiment separately, although significant only for Experiment 3. Analogously to the results presented in Table 4(c), those reported in

---

38Since the distribution of response times is heavily skewed, the *t*-test is performed on the log-transformed data in order to approximately conform to normality.

39Specifically, here we are comparing the mean response time of a subject in menus where she chose with the mean response time in menus where she deferred. These tests therefore only use subjects that deferred at least once and who chose at least once. There are only 81 subjects who ever deferred in the last two experiments, among which 7 always deferred. This makes a total of 74 usable subjects for this paired test.

40We also tested if this might be the case specifically for subjects who always deferred but it turns out that their average response times do not differ significantly from the average response times of the other NFC subjects.
Table 18 are based on the pooled data from both experiments to increase the statistical power of the relevant tests.

9 Concluding Remarks

Existing designs in individual decision making experiments and surveys typically force subjects to always choose one of the market alternatives that the analyst makes available to them. In this study we raised and attempted to answer the question of whether the act of forcing choice in experiments leads to behaviour that is more inconsistent than what it would have been if choice was not forced and subjects were instead allowed to choose the deferral outside option.

Our first contribution is the proposal of a novel non-forced choice experimental design, the key feature of which is that deferral is costly for subjects but choice reversal -which, contrary to other approaches in the literature, is allowed here- is even more costly for them. Following implementation of this design in three experiments with real goods and money lotteries that featured both a forced- and a non-forced choice treatment, our second contribution is to establish -for the first time- that the conjectured negative forced-choice effect on consistency -as proxied by the incidence and distribution of choice reversals and binary choice cycles- generally does exist and is especially pronounced in the case of choice reversals/WARP violations.

Motivated by relevant intuition in Luce and Raiffa (1957) as well as by the formal non-forced choice-theoretic analysis of consistent maximally dominant choice with incomplete preferences that was provided in Gerasimou (2018) and related experimental findings from data collected through different methods, we then studied the hypothesis that the observed forced-choice effect in our data could at least in part be caused by subjects in our sample who had incomplete preferences. Our third contribution is to find evidence consistent with this hypothesis using a novel combinatorial-optimization method -called the distance-score method- that generalized the axiom-based idea of the well-known Houtman and Maks (1985) index in a model-based way. Applying this method on the richer real-goods experimental data, we found that 75% of all subjects are best explained by standard utility maximization (with 45% of them perfectly so), and that 41% of all non-forced choice subjects are best explained by the model of maximally dominant choice with incomplete preferences (with 22% of them perfectly so).

Finally, analysing the additional data from the subjects’ response times and psychological indecisiveness questionnaire responses, we find that deferrals are associated with longer decision times and that subjects who defer are also more psychologically...
indecisive. Together, these findings lend further support to the view that deferrals in our data are driven by preference incompleteness.

References


Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2019):


“Choice, Deferral and Consistency,” working paper, University of St Andrews.


Online Appendix 1

Experiment 1 – Non-Forced Choice Treatment Instructions

General procedure
This experiment aims to study people’s choice behaviour. The choice objects will be 5 headphone sets (HSs). At the start of the experiment you will be allocated £7. You will then be presented with a sequence of 31 menus of HSs (a menu is simply a collection of HSs). Each menu may have 1 to 5 HSs. When a menu appears on your screen you will have the opportunity to look at the image of each HS in that menu and also to read a short description of its main features. You will then be able to choose one of the available HSs, or to select the option “I’m not choosing now”.

You may spend as much time as you want at each menu before deciding what to do. You will see each menu once, and when you proceed to the next menu you will not be able to go back.

After you have seen all 31 menus, one of them will be picked at random (each menu has a 1/31 chance of being selected). You will be reminded of your original decision in this menu (henceforth menu R). You will then get to examine the actual HSs contained in menu R and to try them while listening to a song. Lastly, you will be asked to make a final choice from R (not choosing a HS is not possible at this stage). One in every four participants will be randomly selected to win the HS of their final choice from their randomly selected menu R.

Payment rules (Please also look at examples in separate sheet)
If you have not been selected to win a HS, you will be paid the £7 initially allocated to you.

If you have been selected to win a HS, the following rules apply regarding your payment:

A) Suppose that when you first saw menu R you had chosen some HS from it. If you chose the same the second time, you will receive the £7 initially allocated to you.

B) Suppose that when you first saw menu R you had chosen some HS from it. If you chose a different one the second time, you will receive £3 of the £7 initially allocated to you.

C) Suppose that when you first saw menu R you had chosen “I’m not choosing now”. Then, independent of what you chose from that menu the second time, you will receive £6 of the £7 initially allocated to you.

Special remarks about menus with two HSs (Please also look at examples in separate sheet)

During the phase when you are presented with the 31 menus, whenever a menu of exactly two HSs comes up and you have chosen one of them, a short follow-up question will ask you to state if you preferred the chosen HS over the non-chosen one, or if the non-chosen one was equally good to the one you chose (and therefore you chose randomly between them). If you have chosen “I’m not choosing now” in such a menu, the question will ask you if this was because both HSs were equally good or because you could not decide which one you preferred, or due to some other reason.

If your randomly selected menu R contains two HSs and you had previously stated that both were equally good, then:

1) If you had chosen a headset from R initially, you will not be able to change your decision at this stage. One of the two HS will be randomly selected and you will win this HS if you are picked as a winner.

2) If you had chosen “I’m not choosing now” at R initially, then one of the two HSs will be randomly selected and you will win this HS if you are picked as a winner.
Example 1:
Randomly selected menu with only one headset

Main phase:
- Headset "T"
  - Choose "T"
  - Choose "I'm not choosing now"

Final phase (after inspection):
- Headset "T"
  - Choose "T"
    - £7 + "T" if a winner
    - £7 if not a winner
- Headset "T"
  - Choose "T"
    - £6 + "T" if a winner
    - £7 if not a winner

Example 2:
Randomly selected menu with three headsets

Main phase:
- Choose one of the three ("T", say)
  - Choose different ("U", say)
  - Choose any one of the three ("V", say)

Final phase (after inspection):
- Choose the same ("T")
  - £7 + "T" if a winner
  - £7 if not a winner
- Choose different ("U") if a winner
  - £3 + "U" if a winner
  - £7 if not a winner
- Choose any one of the three ("V")
  - £6 + "V" if a winner
  - £7 if not a winner
Example 3:
Two-headset randomly selected menu where a choice had been made initially

Main phase

Choose any one of the two ("U", say) & state "Both were equally good (...)"

Choose any one of the two ("U", say) & state "I preferred U over V"

Final phase

(Inspection voluntary)

Choose the same ("U")

Choose different ("V")

£7 + "U" if a winner
£7 if not a winner

£7 + "U" or "V" (selected at random) if a winner
£7 if not a winner

Example 4:
Two-headset randomly selected menu where a choice had not been made initially

Main phase

Choose "I'm not choosing now" & state "Both were equally good (...)"

Choose "I'm not choosing now" & state "I could not decide (...)" or "Other reason"

Final phase

(Inspection voluntary)

Choose any one of the two ("U", say)

£6 + "U" if a winner
£7 if not a winner

£6 + "U" or "V" (selected at random) if a winner
£7 if not a winner
Experiment 1 – Forced Choice Treatment Instructions

General procedure

This experiment aims to study people’s choice behaviour. The choice objects will be 5 headphone sets (HSs).

At the start of the experiment you will be allocated £7. You will then be presented with a sequence of 31 menus of HSs (a menu is simply a collection of HSs). Each menu may have 1 to 5 HSs. When a menu appears on your screen you will have the opportunity to look at the image of each HS in that menu and also to read a short description of its main features. You will then be asked to choose one of the available HSs.

You may spend as much time as you want at each menu before deciding what to do. You will see each menu once, and when you proceed to the next menu you will not be able to go back.

After you have seen all 31 menus, one of them will be picked at random (each menu has a 1/31 chance of being selected). You will be reminded of your original decision in this menu (henceforth menu R).

You will then get to examine the actual HSs contained in menu R and to try them while listening to a song. Lastly, you will be asked to make a final choice from R. One in every four participants will be randomly selected to win the HS of their final choice from their randomly selected menu R.

Payment rules (Please also look at examples in separate sheet)

If you have not been selected to win a HS, you will be paid the £7 initially allocated to you.

If you have been selected to win a HS, the following rules apply regarding your payment:

A) Suppose that when you first saw menu R you had chosen some HS from it. If you chose the same HS the second time, you will receive the £7 initially allocated to you.

B) Suppose that when you first saw menu R you had chosen some HS from it. If you chose a different one the second time, you will receive £3 of the £7 initially allocated to you.

Special remarks about menus with two HSs (Please also look at example in separate sheet)

During the phase when you are presented with the 31 menus, whenever a menu of exactly two HSs comes up and you have chosen one of them, a short follow-up question will ask you to state if you preferred the chosen HS over the non-chosen one, or if the non-chosen one was equally good to the one you chose (and therefore you chose randomly between them).

If your randomly selected menu R contains two HSs and you had previously stated that both were equally good, then you will not be able to change your decision at this stage. One of the two HS will be randomly selected and you will win this HS if you are picked as a winner.
Example 1:
Randomly selected menu with three headsets

Main phase

Choose one of the three
("T", say)

Final phase (after inspection)

Choose the same
("T")

£7 + "T" if a winner
£7 if not a winner

Choose different
("U", say)

£3 + "U" if a winner
£7 if not a winner

Example 2:
Randomly selected menu with two headsets

Main phase

Choose any one of the two ("U", say)
& state "Both were equally good (…)"

Choose any one of the two ("U", say)
& state "I preferred U over V"

Final phase

(inspection voluntary)

£7 + "U" or "V" (selected at random) if a winner
£7 if not a winner

Choose the same
("U")

£7 + "U" if a winner
£7 if not a winner

Choose different
("V")

£3 + "V" if a winner
£7 if not a winner
Experiment 1 (& 2) – Sample Screenshots

Please look at the images of the headphone set(s) on this screen

- Carbon fiber integrated headphones
- Light weight
- Strong
- Stainless steel headband
- Carbon fiber housing and large carbon diaphragm

- Lightweight and comfortable to wear
- High-quality leatherette ear pads
- Adjustable noise reduction
- Cordless (1/4 inch (6.35 mm) stereo jack adapter

You now have the option to choose one headphone set or to choose none

- Carbon fiber integrated headphones
- Light weight
- Strong
- Stainless steel headband
- Carbon fiber housing and large carbon diaphragm

- Lightweight and comfortable to wear
- High-quality leatherette ear pads
- Adjustable noise reduction
- Cordless (1/4 inch (6.35 mm) stereo jack adapter

You now have the option to choose one headphone set or to choose none

- Carbon fiber integrated headphones
- Light weight
- Strong
- Stainless steel headband
- Carbon fiber housing and large carbon diaphragm

- Lightweight and comfortable to wear
- High-quality leatherette ear pads
- Adjustable noise reduction
- Cordless (1/4 inch (6.35 mm) stereo jack adapter
Why did you choose this headphone set?

- I chose this headphone set because I preferred it to the other headphones on the menu
- Headphone quality was similar (please specify)
- Other reason

You did not choose any headphone set from this menu. Which of these options best reflects your reason for not choosing?

- I could not decide which one to pick
- I couldn’t decide; good
- Other reason

The choice part of the experiment is now over. We will now determine which menu will be selected for your potential reward.

Please wait for the experimenter to come to you.

He will ask you to pick randomly a numbered ball from a bag and then enter the details below for you.

Experimenter authorization code:

Name/ID:

OK
Online Appendix 2

Experiment 2: Non-Forced Choice Treatment Instructions

General procedure

This experiment aims to study people’s choice behaviour.

At the start of the experiment you will be allocated £6. An additional £2 will be added to this amount if you answer correctly a few questions in a computer quiz before the experiment begins. The aim of this quiz is to help you understand the experiment’s instructions. You must answer all questions correctly by the third attempt in order to receive the additional £2.

The choice objects will be 5 headphone sets (HSs).

Once the experiment begins, you will be presented with a sequence of 26 menus of HSs (a menu is simply a collection of HSs). Each menu may have 2 to 5 HSs.

When a menu appears on your screen you will have the opportunity to look at the image of each HS in that menu and read a short description of its main features. You will then be able to choose one of the available HSs, or to select the option “I’m not choosing now”.

You may spend as much time as you want at each menu before deciding what to do. You will see each menu once, and when you proceed to the next menu you will not be able to go back.

After you have seen all 26 menus, one of them will be picked at random (each menu has a 1/26 chance of being selected). You will be reminded of your original decision in this menu (henceforth menu R). You will then be asked to make a final choice from R (not choosing a HS is not possible at this stage).

One in every four participants will be randomly selected to win the HS of their final choice from their randomly selected menu R.

Payment rules (please also look at the example in a separate sheet)

If you have not been selected to win a HS, you will be paid the £6 + £2 = £8 initially allocated to you.

If you have been selected to win a HS, the following rules apply regarding your payment:

A) Suppose that when you first saw menu R you had chosen some HS from it. If you chose the same HS the second time, you will receive the £8 initially allocated to you.

B) Suppose that when you first saw menu R you had chosen some HS from it. If you chose a different HS the second time, you will receive £2 of the £8 initially allocated to you.

C) Suppose that when you first saw menu R you had chosen “I’m not choosing now”. Then, independently of which HS you chose from that menu the second time, you will receive £7.50 of the £8 initially allocated to you.
Example:

Randomly selected menu with three headsets
(the stated amounts assume that all quiz questions were answered correctly by the 3rd attempt)
Experiment 2: Forced Choice Treatment Instructions

General procedure

This experiment aims to study people’s choice behaviour.

At the start of the experiment you will be allocated £6. An additional £2 will be added to this amount if you answer correctly a few questions in a computer quiz before the experiment begins. The aim of this quiz is to help you understand the experiment’s instructions. You must answer all questions correctly by the third attempt in order to receive the additional £2.

The choice objects will be 5 headphone sets (HSs).

Once the experiment begins, you will be presented with a sequence of 26 menus of HSs (a menu is simply a collection of HSs). Each menu may have 2 to 5 HSs.

When a menu appears on your screen you will have the opportunity to look at the image of each HS in that menu and read a short description of its main features. You will then be asked to choose one of the available HSs.

You may spend as much time as you want at each menu before deciding what to do. You will see each menu once, and when you proceed to the next menu you will not be able to go back.

After you have seen all 26 menus, one of them will be picked at random (each menu has a 1/26 chance of being selected). You will be reminded of your original decision in this menu (henceforth menu \( R \)). You will then be asked to make a final choice from \( R \).

One in every four participants will be randomly selected to win the HS of their final choice from their randomly selected menu \( R \).

Payment rules (please also look at the example in a separate sheet)

If you have not been selected to win a HS, you will be paid the £6 + £2 = £8 initially allocated to you.

If you have been selected to win a HS, the following rules apply regarding your payment:

A) Suppose that when you first saw menu \( R \) you had chosen some HS from it. If you chose the same HS the second time, you will receive the £8 initially allocated to you.

B) Suppose that when you first saw menu \( R \) you had chosen some HS from it. If you chose a different HS the second time, you will receive £2 of the £8 initially allocated to you.
Example:

Randomly selected menu with three headsets

(the stated amounts assume that all quiz questions were answered correctly by the 3rd attempt)
Online Appendix 3

Experiment 3 – Non-Forced Choice Treatment Instructions (backward-translated)

General procedure

This experiment’s objective is to study people’s decision making. The objects on which decisions will be taken will be 6 lotteries. A lottery is a collection of payments with different probabilities associated to each payment. You can find two examples of lotteries overleaf.

At the beginning of the experiment you will be assigned €5.50. After this, you will be presented with a sequence of 15 lotteries menus. Each menu is simply a pair of lotteries. When a menu appears on your screen you will have the opportunity to see images of the two lotteries in that menu. You will then have the opportunity to choose one of the two lotteries, or to choose the option “I’m not choosing now”.

You may take all the time you wish in each menu before deciding what to do. You will see each menu once, and when you move to the next menu you will not be able to go back.

After having seen the 15 menus, one of them will be selected randomly (each menu has one chance in 15 of being selected). You will be reminded of your original decision on that menu (call this menu R).

Finally, you will have to make a final decision at menu R (not choosing a lottery will not be possible at this time). One in four participants will be randomly selected to win the lottery chosen in her final decision at menu R. The examples overleaf explain what “winning a lottery” means.

Payment rules (please, look also at the examples on the separate sheet)

If you have not been selected to win a lottery you will be paid the €5.50 initially allocated to you.

If you have been selected to win a lottery, you will be paid according to the following rules:

A) Suppose that when you saw menu R for the first time you picked a lottery from it. If you choose the same lottery the second time, you will receive the €5.50 initially allocated to you.

B) Suppose that when you saw menu R for the first time you chose a lottery from it. If you choose a different lottery the second time, you will receive €1.50 of the €5.50 initially allocated to you.

C) Suppose that when you saw menu R for the first time you selected “I’m not choosing now”. Then, independently of what you choose from this menu the second time, you will receive €5 of the €5.50 initially allocated to you.

Questions between “menu and menu” (translation note: awkward in English but works in Spanish)

During the phase in which the 15 menus are presented to you, you will be asked a brief question between menu and menu.

If you have chosen one of the lotteries in a menu, we will ask you to select one of the three following reasons:

a) I prefer the lottery I have chosen over the other lottery;

b) I have chosen purely at random because I find both lotteries exactly equally good;

c) Other reason.

If you have selected the option “I’m not choosing now”, we will ask you to select one of the three following reasons:

a) I could not decide which one I prefer;
b) I have chosen purely at random because I find both lotteries exactly equally good;
c) Other reason.

**Important**: if you choose “I have chosen purely at random” in a menu and this menu is selected for the final decision, we will choose one of the two lotteries randomly for you and you will not have the opportunity to make any final decision. Therefore, you should only select this option if you really find both lotteries exactly equally good.

These are two lottery examples.

Lottery A pays €10 with a probability of 30%, €15 with a probability of 40% and €25 with a probability of 30%.

Lottery B pays €12 with a probability of 20%, €22 with a probability of 65% and €33 with a probability of 15%.

What does “winning a lottery” mean?

If you win a lottery, you will receive one of the payments detailed in the lottery randomly, according to the probabilities stated in the lottery.

For this, we will ask you to draw a ball at random from a dark bag that will contain 100 balls enumerated from 1 to 100. For example:

Imagine you have won lottery A: if the number of the ball you draw is between 1 and 30, you will obtain €10; if it is between 31 and 70, you will obtain €15; if it is between 71 and 100, you will obtain €25.

Imagine you have won lottery B: if the number of the ball you draw is between 1 and 20, you will obtain €12; if it is between 21 and 85, you will obtain €22; if it is between 86 and 100, you will obtain €33.
Ejemplo 1: si NO indicas “He elegido puramente al azar”

Ejemplo 2: si indicas “He elegido puramente al azar”
Experiment 3 – Sample Screenshots
Experiment 3 – Forced Choice Treatment Instructions (backward-translated)

General procedure

This experiment’s objective is to study people’s decision making. The objects on which decisions will be taken will be 6 lotteries. A lottery is a collection of payments with different probabilities associated to each payment. You can find two examples of lotteries overleaf.

At the beginning of the experiment you will be assigned €5.50. After this, you will be presented with a sequence of 15 lotteries menus. Each menu is simply a pair of lotteries. When a menu appears on your screen you will have the opportunity to see images of the two lotteries in that menu. You will then be asked to choose one of the two lotteries.

You may take all the time you wish in each menu before deciding what to do. You will see each menu once, and when you move to the next menu you will not be able to go back.

After having seen the 15 menus, one of them will be selected randomly (each menu has one chance in 15 of being selected). You will be reminded of your original decision on that menu (call this menu $R$).

Finally, you will have to make a final decision at menu $R$. One in four participants will be randomly selected to win the lottery chosen in her final decision at menu $R$. The examples overleaf explain what “winning a lottery” means.

Payment rules (please, look also at the examples on the separate sheet)

If you have not been selected to win a lottery you will be paid the €5.50 initially allocated to you.

If you have been selected to win a lottery, you will be paid according to the following rules:

A) Suppose that when you saw menu $R$ for the first time you picked a lottery from it. If you choose the same lottery the second time, you will receive the €5.50 initially allocated to you.

B) Suppose that when you saw menu $R$ for the first time you chose a lottery from it. If you choose a different lottery the second time, you will receive €1.50 of the €5.50 initially allocated to you.

Questions between “menu and menu” (translation note: awkward in English but works in Spanish)

During the phase in which the 15 menus are presented to you, you will be asked a brief question between menu and menu. Specifically, after your choice, you will be asked to select one of the following three reasons:

a) I clearly prefer the lottery I have chosen over the other lottery;
b) I slightly prefer the lottery I have chosen over the other lottery;
c) I have chosen purely at random because I find both lotteries exactly equally good.

Important: if you choose “I have chosen purely at random” in a menu and this menu is selected for the final decision, we will choose one of the two lotteries randomly for you and you will not have the opportunity to make any final decision. Therefore, you should only select this option if you really find both lotteries exactly equally good.
Ejemplo 1: si NO indicas “He elegido puramente al azar”

**Fase principal**

- a) Prefiero la lotería que he elegido
- b) He elegido puramente al azar
- c) Otra razón

**Decisión final**

- a) Prefiero la lotería que he elegido
- b) He elegido puramente al azar
- c) Otra razón

<table>
<thead>
<tr>
<th>U</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,50€ + U si ganas</td>
<td>1,50€ + V si ganas</td>
</tr>
<tr>
<td>5,50€ si no ganas</td>
<td>5,50€ si no ganas</td>
</tr>
</tbody>
</table>

Ejemplo 2: si indicas “He elegido puramente al azar”

**Fase principal**

- a) Prefiero la lotería que he elegido
- b) He elegido puramente al azar
- c) Otra razón

**Decisión final**

<table>
<thead>
<tr>
<th>U</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,50€ + V o U (elegido al azar) si ganas</td>
<td>1,50€ + V o U (elegido al azar) si ganas</td>
</tr>
<tr>
<td>5,50€ si no ganas</td>
<td>5,50€ si no ganas</td>
</tr>
</tbody>
</table>
Online Appendix 4

Consistency Analysis without Noisiness or Singleton-Deferral Exclusions in Experiments 1 & 2
Table 19: Average (median) consistency violations per experiment and treatment. 
*p-values from 2-tailed Mann-Whitney U tests.

(a) Experiment 1

<table>
<thead>
<tr>
<th>Treatment</th>
<th>WARP</th>
<th>Linear-normalized</th>
<th>Nonlinear-normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ Indifference</td>
<td>Indifference</td>
<td>+ Indifference</td>
</tr>
<tr>
<td>NFC</td>
<td>1.87 (0)</td>
<td>1.96 (0)</td>
<td>0.08 (0)</td>
</tr>
<tr>
<td>n</td>
<td>149</td>
<td>149</td>
<td>146</td>
</tr>
<tr>
<td>FC</td>
<td>3.32 (0)</td>
<td>3.64 (0)</td>
<td>0.13 (0)</td>
</tr>
<tr>
<td>n</td>
<td>76</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>p-value</td>
<td>0.084</td>
<td>0.015</td>
<td>0.131</td>
</tr>
</tbody>
</table>

(b) Experiment 2

<table>
<thead>
<tr>
<th>Treatment</th>
<th>WARP</th>
<th>Linear-normalized</th>
<th>Nonlinear-normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+ Indifference</td>
<td>Indifference</td>
<td>+ Indifference</td>
</tr>
<tr>
<td>NFC</td>
<td>3.16 (0)</td>
<td>0.14 (0)</td>
<td>0.05 (0)</td>
</tr>
<tr>
<td>n</td>
<td>68</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>FC</td>
<td>4.83 (0)</td>
<td>0.19 (0)</td>
<td>0.07 (0)</td>
</tr>
<tr>
<td>n</td>
<td>54</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>p-value</td>
<td>0.086</td>
<td>0.161</td>
<td>0.157</td>
</tr>
</tbody>
</table>

70
Online Appendix 5

Indifference-Augmentation of Single-Valued Choices

Consistency check

The input to the consistency check are indifference statements represented as a set of binary menus where the subject has expressed indifference, named \( \mathcal{I} \), together with the choice function \( C \) giving the choice set \( C(M) \) of the subject for each menu \( M \).

**Indifference after deferrals** First, we determine the set \( \mathcal{D} \) of menus where the subject has deferred while stating indifference:

\[
\mathcal{D} := \{ M \mid M \in \mathcal{I}, C(M) = \emptyset \}
\]  

We will disregard these statements as being in conflict with the experiment’s incentives toward indifference elicitation.

**Preference graphs** Then we construct two preference graphs with labelled edges, one for indifference, one for strict preference. In both graphs, each vertex represents one of the alternatives available in some menu, while the presence of an edge indicates an indifference or strict preference relation between the two alternatives. Each edge is labelled with the set of menus that are involved in the corresponding comparison (see below). Since the set of vertices is the same for both graphs, we identify the graphs with their edge sets. Finally, let us call the set of all vertices \( V \).

The edge set of a graph is given as a function from pairs of vertices (i.e. alternatives) to the set of menus that are involved in the comparison. For example, if a subject chooses \( A \) over \( B \) from menu \( \{A, B\} \), then the menu \( \{A, B\} \) is involved in the comparison \( A \succ B \). If furthermore menu \( \{B, C\} \) is involved in the comparison \( B \succ C \), we can conclude that at least menus \( \{\{A, B\}, \{B, C\}\} \) are involved in the comparison \( A \succ C \).

The two graphs are defined as follows. For convenience, we will take the empty set of involved menus to indicate that there is no edge. We start with edges directly following from explicit choices of the subject, and then we compute their transitive closure iteratively.

- **Graph \( E \)** is an undirected graph corresponding to the relation of indifference.
  
  \[
  E_0(u, v) = \{ M \mid M \in \mathcal{I} \setminus \mathcal{D}, \{u, v\} \subseteq M \}
  \]

  This graph starts out with edges between alternatives offered in menus where
indifference has been stated after choice.

- Graph $L$ is a directed graph corresponding to the relation of strict preference.

$$L_0(u, v) = \{ M \mid u \in M \setminus C(M), v \in C(M) \}$$

This graph starts out with edges between alternatives $(u, v)$, where $v$ has been chosen over $u$ in some menu.

**Transitive closure**  Then we compute the transitive closure of both edge relations. The computation step is slightly different for each graph.

For graph $E$, we compute just its own transitive closure $E^*$.

$$E_{i+1}(u, v) := E_i(u, v) \cup \{ E_i(u, x) \cup E_i(x, v) \mid x \in V \}$$

$$E^* := E_k \text{ for } k \text{ such that } E_k = E_{k+1}$$

For graph $L$, there are further requirements on transitivity – for example, if $A \succ B$ and $B \sim C$, then we should have $A \succ C$.

$$L_{i+1}(u, v) := L_i(u, v)$$

$$\cup \{ L_i(u, x) \cup L_i(x, v) \mid x \in V \}$$

$$\cup \{ E_i(u, x) \cup L_i(x, v) \mid x \in V \}$$

$$\cup \{ L_i(u, x) \cup E_i(x, v) \mid x \in V \}$$

$$L^* := L_k \text{ for } k \text{ such that } L_k = L_{k+1}$$

**Removal of conflicting indifference statements**  The final step of the consistency check is based on the fact that the (transitive) comparisons we have just built may be absent in the set of directly stated indifferences or strict preferences, because the subject either chose differently or deferred. Each such case represents an intransitivity. The set of conflicting menus $\mathcal{C}$ collects all menus that are involved in intransitivities.

$$\mathcal{C}_E := \bigcup_{E_0(u,v)=\emptyset} E^*(u,v) \quad \mathcal{C}_L := \bigcup_{L_0(u,v)=\emptyset} L^*(u,v)$$

$$\mathcal{C} := \mathcal{C}_E \cup \mathcal{C}_L$$

Although these intransitivities may be resolved by removing only some of the indifference statements, we remove all indifference statements involved in any intransitivity because there is no clear rule to dictate which indifference statements we should preferentially keep.
Therefore, the updated set of indifference statements is:

\[ I' := I \setminus (D \cup C) \] (10)

**Repeat**  By removing some indifference statements, we may however create new conflicts because some indifferences become strict preferences. Therefore, we repeat the above process of consistency check and indifference statement removal until \( I' = I \).

**Augmentation**

Given a (consistent) set of menus where indifference has been stated, \( I' \), we calculate the augmented choice set \( C'(M) \) as the union of all indifferent menus having common elements with the original choice set \( C(M) \). Furthermore, the set \( C'(M) \) should not contain any alternatives that are not present in \( M \).

\[ C'(M) := M \cap \bigcup \{ K \mid K \in I', K \cup C(M) \neq \emptyset \} \] (11)

In all further processing, we then regard \( C' \) as the choice correspondence of the subject.

**Illustrative Examples**

**Example: 1**

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stated</td>
<td>Accepted</td>
</tr>
<tr>
<td>w.x</td>
<td>w</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>w.y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w.z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x.y</td>
<td>x</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>y.z</td>
<td>y</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>x.z</td>
<td>z</td>
<td>Yes</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \( \{ y \}, \{ z \}, \{ x \}, \{ w \} \)

**Example: 2**

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stated</td>
<td>Accepted</td>
</tr>
<tr>
<td>w.x</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w.y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w.z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x.y</td>
<td>x</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>y.z</td>
<td>y</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>x.z</td>
<td>z</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Indifference classes: \( \{ x,y,z \}, \{ w \} \)
Example: 3

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference Stated</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>w</td>
</tr>
<tr>
<td>w,y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w,z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,z</td>
<td>z</td>
<td>Yes</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \{y\}, \{z\}, \{x\}, \{w\}

Example: 4

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference Stated</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>w,x</td>
</tr>
<tr>
<td>w,y</td>
<td>y</td>
<td>Yes</td>
<td>w,y</td>
</tr>
<tr>
<td>w,z</td>
<td>z</td>
<td>Yes</td>
<td>w,z</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>x,y</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>Yes</td>
<td>y,z</td>
</tr>
<tr>
<td>x,z</td>
<td>z</td>
<td>Yes</td>
<td>x,z</td>
</tr>
</tbody>
</table>

Indifference classes: \{w,x,y,z\}

Example: 5

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference Stated</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>w</td>
</tr>
<tr>
<td>w,y</td>
<td>y</td>
<td>Yes</td>
<td>y</td>
</tr>
<tr>
<td>w,z</td>
<td>z</td>
<td>-</td>
<td>z</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>x</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>Yes</td>
<td>y</td>
</tr>
<tr>
<td>x,z</td>
<td>z</td>
<td>Yes</td>
<td>z</td>
</tr>
</tbody>
</table>

Indifference classes: \{y\}, \{z\}, \{x\}, \{w\}

Example: 6

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference Stated</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>w</td>
</tr>
<tr>
<td>w,y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w,z</td>
<td>z</td>
<td>-</td>
<td>z</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>-</td>
<td>x</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>Yes</td>
<td>y</td>
</tr>
<tr>
<td>x,z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \{y\}, \{z\}, \{x\}, \{w\}
### Example: 7

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stated</td>
<td>Accepted</td>
</tr>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>w,y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w,z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>x,z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \{w,x\}, \{y,z\}

### Example: 8

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stated</td>
<td>Accepted</td>
</tr>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>w,y</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>w,z</td>
<td>z</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>x,z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \{x\}, \{y\}, \{w\}, \{z\}

### Example: 9

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stated</td>
<td>Accepted</td>
</tr>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>w,y</td>
<td>w</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>w,z</td>
<td>w</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,z</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \{x\}, \{y\}, \{w\}, \{z\}

### Example: 10

<table>
<thead>
<tr>
<th>Menu</th>
<th>Raw choice</th>
<th>Indifference</th>
<th>Augmented choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stated</td>
<td>Accepted</td>
</tr>
<tr>
<td>w,x</td>
<td>w</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>w,y</td>
<td>w</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>w,z</td>
<td>w</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,y</td>
<td>x</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>y,z</td>
<td>y</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>x,z</td>
<td>x</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Indifference classes: \{w,x,y\}, \{z\}
Online Appendix 6

Consistency within the NFC Treatments in Experiments 1 & 2

Table 20: Proportions of (non-)deferring NFC subjects violating WARP & Binary Congruence in Experiments 1 & 2. 
*p*-values from 2-tailed Fisher exact tests.

(a) Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>NFC: non-deferring</th>
<th>NFC: deferring</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARP</td>
<td>18/56 (32.14%)</td>
<td>6/28 (21.43%)</td>
<td>0.443</td>
</tr>
<tr>
<td>Binary Congruence</td>
<td>1/56 (1.79%)</td>
<td>1/28 (3.57%)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

(b) Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>NFC: non-deferring</th>
<th>NFC: deferring</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WARP</td>
<td>7/36 (19.44%)</td>
<td>5/26 (19.23%)</td>
<td>1.00</td>
</tr>
<tr>
<td>Binary Congruence</td>
<td>1/36 (2.78%)</td>
<td>1/26 (3.8%)</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 21: Average (median) violations of (non-)deferring NFC subjects in Experiments 1 & 2.
\( p \)-values from Mann-Whitney \( U \) tests.

(a) Experiment 1

<table>
<thead>
<tr>
<th>NFC: non-deferring</th>
<th>WARP</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 56</td>
<td>1.80 (0)</td>
<td>0.069 (0)</td>
<td>0.026 (0)</td>
</tr>
<tr>
<td>N = 28</td>
<td>0.79 (0)</td>
<td>0.038 (0)</td>
<td>0.016 (0)</td>
</tr>
<tr>
<td>p-value = 0.196</td>
<td>0.347</td>
<td>0.582</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NFC: deferring</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 25</td>
<td>0.038 (0)</td>
<td>0.016 (0)</td>
</tr>
<tr>
<td>p-value = 0.196</td>
<td>0.347</td>
<td>0.582</td>
</tr>
</tbody>
</table>

(b) Experiment 2

<table>
<thead>
<tr>
<th>NFC: non-deferring</th>
<th>WARP</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 36</td>
<td>1.19 (0)</td>
<td>0.05 (0)</td>
<td>0.017 (0)</td>
</tr>
<tr>
<td>N = 26</td>
<td>0.96 (0)</td>
<td>0.06 (0)</td>
<td>0.022 (0)</td>
</tr>
<tr>
<td>p-value = 0.849</td>
<td>0.799</td>
<td>0.799</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NFC: deferring</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 22</td>
<td>0.05 (0)</td>
<td>0.022 (0)</td>
</tr>
<tr>
<td>p-value = 0.849</td>
<td>0.799</td>
<td>0.799</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NFC: non-deferring</th>
<th>WARP</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 36</td>
<td>1.61 (0)</td>
<td>0.06 (0)</td>
<td>–</td>
</tr>
<tr>
<td>N = 26</td>
<td>0.04 (0)</td>
<td>0.005 (0)</td>
<td>0.005 (0)</td>
</tr>
<tr>
<td>p-value = 1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NFC: deferring</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 25</td>
<td>0.05 (0)</td>
<td>0.022 (0)</td>
</tr>
<tr>
<td>p-value = 0.849</td>
<td>0.799</td>
<td>0.799</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NFC: non-deferring</th>
<th>WARP</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 36</td>
<td>1.61 (0)</td>
<td>0.06 (0)</td>
<td>–</td>
</tr>
<tr>
<td>N = 26</td>
<td>0.04 (0)</td>
<td>0.005 (0)</td>
<td>0.005 (0)</td>
</tr>
<tr>
<td>p-value = 1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NFC: deferring</th>
<th>Linear-normalised</th>
<th>Nonlinear-normalised</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 25</td>
<td>0.05 (0)</td>
<td>0.022 (0)</td>
</tr>
<tr>
<td>p-value = 0.849</td>
<td>0.799</td>
<td>0.799</td>
</tr>
</tbody>
</table>
Online Appendix 7

Conditional Fixed-Effects Logit Regressions

Table 22: Conditional fixed-effects logistic regressions. Dependent variable = 1 if a subject defers choice, = 0 otherwise.

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>menu number</td>
<td>0.0012</td>
<td>-0.0164</td>
<td>-0.0218</td>
</tr>
<tr>
<td></td>
<td>(0.0148)</td>
<td>(0.0146)</td>
<td>(0.0213)</td>
</tr>
<tr>
<td>menu size = 3</td>
<td>1.171***</td>
<td>-0.484**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.267)</td>
<td>(0.235)</td>
<td></td>
</tr>
<tr>
<td>menu size = 4</td>
<td>1.3197***</td>
<td>-1.281***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.318)</td>
<td>(0.321)</td>
<td></td>
</tr>
<tr>
<td>menu size = 5</td>
<td>1.489**</td>
<td>-0.951</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.578)</td>
<td>(0.619)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>650</td>
<td>572</td>
<td>780</td>
</tr>
<tr>
<td>Number of id</td>
<td>25</td>
<td>22</td>
<td>52</td>
</tr>
</tbody>
</table>

Note: menu number takes values from 1 to 26 in Experiment 1 and 2, and 1 to 15 in Experiment 3; menu size is captured with 3 dummy variables, with binary menus as the base category; number of id corresponds to the number of subjects used in the regression (subjects always deferring or who never defer drop out from the regression).

Standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1.
Online Appendix 8

Simulations Output

WARP violations (5 alternatives − 26 menus)

Binary Congruence violations (FC − 5 alternatives − 10 menus)

Binary Congruence violations (NFC − 5 alternatives − 10 menus)
Binary Congruence violations (FC – 6 alternatives – 15 menus)

Frequency

0 5 10 15 20 25 30 35

0 20000 40000 60000 80000 100000

Binary Congruence violations (NFC – 6 alternatives – 15 menus)

Frequency

0 5 10 15 20 25 30 35

0 100000 200000 300000 400000 500000

80
Online Appendix 9

Nonlinear Normalizations of Axiom Violations

The maximum number of WARP violations as a function of active choices (5 alternatives)

The maximum number of Binary Congruence violations as a function of active binary choices (5 alternatives)

The maximum number of Binary Congruence violations as a function of active binary choices (6 alternatives)