

University of  
St Andrews

School of Economics and Finance Discussion Papers

# Choice, Deferral and Consistency

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

School of Economics and Finance Discussion Paper No. 1416  
12 Dec 2014 (revised 26 Dec 2016)  
JEL Classification: C91, D01, D03, D11, D12,  
Keywords: Choice deferral; choice reversals; indecisiveness;  
unattractiveness; revealed preference; Houtman-Maks

---

# Choice, Deferral and Consistency\*

Miguel Costa-Gomes<sup>†</sup>   Carlos Cueva<sup>‡</sup>   Georgios Gerasimou<sup>§</sup>   Matúš Tejišćák<sup>¶</sup>

December 26, 2016

## Abstract

We conduct a novel experiment in which subjects are forced to choose one of several real goods in one treatment, but are not forced to choose, and can instead incur a small cost to defer choice, in the other treatment. We find that forcing subjects to choose, which is the convention in choice experiments, leads to relatively higher rates of choice reversals. This implies that standard choice experiments may lead researchers to overestimate the fraction of subjects that do not maximize a stable and transitive preference relation. We then use a new combinatorial-optimization method that detects a subject's possibly incomplete (due to indecisiveness) or truncated (due to undesirability) preferences by minimizing the number of switches to her active choices and/or choice deferrals that are necessary to generate behavior consistent with maximization of such preferences. Slightly above one half of our subjects' decisions are best explained by such preferences or by a Bayesian preference for information, whereas the rest are best explained by standard utility maximization.

Keywords: Choice deferral; choice reversals; indecisiveness; unattractiveness; revealed preference; Houtman-Maks.

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

---

\*We thank Syngjoo Choi, Jan Heufer, Larry Samuelson, conference audiences at the World Congress of the Econometric Society (2015, Montreal), Bounded Rationality in Choice (2014, Barcelona), Foundations of Utility and Risk (2014, Rotterdam) and workshop/seminar audiences at St Andrews, Technion, Aberdeen, Alicante, Cyprus, WZB Berlin, East Anglia, Bath and Durham for useful comments. Gerasimou and Costa-Gomes gratefully acknowledge financial support from the British Academy. Cueva acknowledges support from the Spanish Ministry of Economics and Competitiveness and the FEDER fund. Any errors are our own.

<sup>†</sup>School of Economics & Finance, University of St Andrews.

<sup>‡</sup>Departamento de Fundamentos del Análisis Económico, Universidad de Alicante.

<sup>§</sup>School of Economics & Finance, University of St Andrews.

<sup>¶</sup>School of Computer Science, University of St Andrews.

---

# 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.<sup>1</sup> One may therefore expect that individuals who are forced to choose are generally more likely to do so in a random way that does not reflect an underlying preference than those who are not. To the extent that this is so, a natural testable hypothesis is that such decision makers’ choices will feature a higher degree of inconsistency, which would be manifested through a higher incidence of choice reversals. For example, when consumers commit themselves to making purchases on special sales days, they may later find themselves willing to return their purchases to the store. Based on data that we collected by implementing a novel and suitably incentivized experimental protocol, in this paper we provide evidence suggesting that forced choice is indeed more inconsistent than non-forced choice.

Intuition and experimental evidence suggest 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 their occurrence in decisions among multi-attribute alternatives and the ensuing avoidant/deferring behavior of the agents in menus that include such alternatives. As Shafir, Simonson, and Tversky (1993) 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 relevant options in order to make an active choice.<sup>2</sup>

---

<sup>1</sup>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, this behavioral restriction that it imposes has rendered it vulnerable to criticism on both normative and positive grounds by, among others, von Neumann and Morgenstern (1944), Savage (1954), Aumann (1962) and Gilboa et al (2010).

<sup>2</sup>Some studies by psychologists that have lent support to this hypothesis are Ferrari and Dovidio (2001) and Rassin et al (2008). Among other things, these works showed that individuals with high scores in indecisiveness personality questionnaires were signif-

---

The experiment 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 were sequentially presented with 31 menus in random order, consisting of every possible combination from a fixed set of 5 choice alternatives (headphone sets). In the target treatment, subjects had the option to defer choice at any menu, whereas in the control treatment they were always required to make a choice. After this stage, one menu was randomly selected for each subject, who then had the opportunity to physically inspect and try the alternatives contained in that menu. After inspection, subjects were required to re-enter a final choice for that menu for a 0.25 chance of winning the chosen alternative. Crucially, if a subject turned out to win the alternative, and her final choice differed from her original choice in that menu (of which she was reminded before making her final choice), then she would have to pay a switching cost of £4 which was deducted from her show-up fee. On the other hand, if she had originally deferred choice at that menu, she would pay a smaller cost of £1. Finally, if her original choice and final choice were the same, she would pay nothing after winning her alternative.

As we explain in detail below, the features of our design make subjects' behavior compatible with a model of "rational indecisiveness" according to which a decision maker makes an active choice at a menu if and only if there is a most preferred alternative in that menu. It also makes it compatible with a Bayesian model of rational information acquisition whereby an agent with completely ordered and stable first-period preferences decides between actively choosing and deferring based on whether his belief that the additional information will actually update his preferences is relatively low or high, respectively. In the context of our experimental design this Bayesian model can also explain the behavior that is generated by the model of rational indecisiveness under some suitable choice of parameter values.

In addition to the novel experimental protocol that we propose and the forced-choice effect on consistency that we document, we also develop a new combinatorial-optimization method for assessing the consistency of datasets that also include deferral observations. The main idea here is to find what is the closest model of "rational choice deferral" for each dataset. In the context of our experimental data this helps us categorize subjects on the basis of what is the most likely reason for their deferrals. The possibilities that we consider are the standard model of unconstrained utility maximization (which is particularly relevant for subjects who never deferred), the Bayesian and rational indecisiveness models, and a model where deferral is driven by utility maximization subject to unattractiveness constraints in the sense that alternatives below a certain desirability threshold are never chosen. Our subject categorization is obtained by finding the minimum number of changes that need to be made in a given dataset for it to be as if it was generated by a decision maker who conformed perfectly with each of the

---

icantly more likely to delay making a decision and to seek additional information, especially about the alternative they ultimately chose. As we explain below, our analysis also supports this conclusion.

---

unconstrained and unattractiveness-constrained utility maximization models and the Bayesian or rational indecisiveness ones, and then by assigning each subject to the model with the minimum number of required changes. This approach extends in important new directions the Houtman-Maks (1985) method which has been widely studied and applied by economists in recent years.<sup>3</sup>

Our analysis suggests that 43% and 30% of the subjects in the non-forced choice treatment are best explained by unconstrained and unattractiveness-constrained utility maximization, respectively. Moreover, for 21% of the subjects in this treatment the best matches are the rational indecisiveness and Bayesian models, with two independent sources of questionnaire data strongly suggesting that around two thirds of the members of this joint category in fact deferred due to indecisiveness. Finally, for the remaining 6% there is an overlap between two observationally distinct models. It is worth noting from the outset that we also consider deferral patterns consistent with either “choice overload” or “choice fatigue” but find no evidence for either of these effects. This, in turn, provides additional re-assurance that our experimental subjects’ decisions were indeed preference-based, as intended.

## 2 The Experiment

We used a between-subjects design with two treatments: Non-Forced-Choice (NFC) and Forced-Choice (FC). The instructions that were given to subjects are available in Online Appendices 1 and 2, respectively. The experimental interface was created with the z-Tree software (Fischbacher, 2007). The grand choice set consisted of five headsets, with their brands and models chosen so that their price was approximately the same (between £10 and £20 at the time of purchase) but with their attributes differing 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 brands (e.g. the headset with the less commonly known brand name was wireless whereas all others were not).

The menus (decision problems) consisted of images of the headsets and of a short description of their main features (see Online Appendix 3 for some examples). 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. We anticipated that this fact would be an additional source of incomparability for some subjects, and that these subjects would welcome the possibility of obtaining more information about the

---

<sup>3</sup>See, for example, Fisman et al (2007), Choi et al (2007), Choi et al (2014), Smeulders et al (2014), Heufer and Hjertstrand (2015), Apestequia and Ballester (2015), Dean and Martin (2016) and Echenique et al (2016).

---

alternatives before choosing.

The principal aim of our design was to generate incentives that would elicit the subjects' weak preferences over headsets, including, possibly, their indecisiveness component. In both treatments every subject was presented with the sequence of all 31 menus generated from the set of five headsets. The order in which menus appeared was random. It varied across sessions but was the same for all subjects within each session. This was also true for the order from left to right in which headsets appeared in each menu, while it was 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 a headset from all menus presented to them, without being able to defer choice. Subjects in the NFC treatment had the opportunity to choose one headset in each menu or to select "*I'm not choosing now*". Once past a menu, subjects could not go back to review and change their choice, with one crucial exception described in the next paragraph.

In both treatments, after everyone had made a decision (including "*I'm not choosing now*" in the NFC treatment) in all 31 menus, one menu was randomly selected for each subject. Each subject then saw their own randomly selected menu and was reminded of the decision they had made there initially. Then, in groups of four, subjects were asked 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 thus visible to the subjects), 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. After this stage subjects went back to their desks and were asked to choose one of the headsets. Thus, they could either maintain their original active choice (if one had been made in the case of NFC subjects) or change it. NFC subjects who had not chosen a headset when originally presented with that menu were required to choose one at this stage.

Everyone was 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 headset of their *final* choice from their randomly selected menu. We refer to such subjects as "*winners*". Participants were also told from the beginning that winners might face some costs which would reduce the amount of £7 initially allocated to them, depending on their first and second decisions at the randomly selected menu. In particular, if a subject that later became a winner had decided to choose a headset other than the one he or she originally selected from this menu, £4 were taken away from her initially allocated £7. In contrast, there was no deduction if the subject opted for the same headset again the second time they chose from that menu. Finally, subjects in the NFC treatment who originally deferred at their randomly selected menu and later became winners incurred the cost of a £1 deduction from the initial allocation of £7. Participants were told from the beginning that if they were *not* selected to win a headset they would receive their full £7 fee irrespective of their first and second decisions at the randomly selected menu.

---

It is clear that the possibility of a £4 penalty that is associated with a choice reversal in the above structure makes it incentive-compatible for all subjects who have decided to make an active choice at some menu during the main phase of the experiment to indeed choose their currently most preferred element in that menu. Moreover, under certain conditions that we identify in a separate section below, the above structure also makes it incentive-compatible for subjects to defer at those menus where the additional information is perceived as being potentially helpful in *updating* their completely ordered preferences (case of Bayesian decision makers) or in *completing* their incompletely ordered preferences and thereby in finding a most preferred feasible option (case of indecisive decision makers).

Given that at most one headset was chosen from every menu, in order to disentangle strict preference, indifference and indecisiveness between two alternatives we used a questionnaire-based method. Specifically, at each *binary* menu where subjects 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 a headset. Thus, subjects had no incentive to state indifference in binary menus where they chose a headset 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. We note that this design does not make it a dominant strategy for subjects who are indifferent to state so. It does, however, make such a statement an undominated strategy, provided subjects are truly indifferent, in which case they would not value the additional information or the opportunity to change their choice later.

Following binary menus where their decision was the “*I’m not choosing now*” option, subjects 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 deferral cost was applied and they won the headset that was determined by the *flip of a coin*. The coin-flip procedure did not apply in the other two cases.

Before the beginning of the main task of the experiment which involved the sequential presentation of all 31 menus, subjects were presented with a mini (“trial”) version with three alternatives (MP3 players) that aimed to familiarize 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 questions that tested 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 excluded from participating in the main phase.<sup>4</sup>

After the end of the main task of the experiment subjects were asked to fill in a personality questionnaire that included the *indecisiveness* items in Germeijs and Boeck (2002). The possible responses to these items 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*”.

Our experiment was carried out at the University of St Andrews Experimental Economics Lab between September 2013 and January 2014. Subjects were undergraduate and postgraduate students of no particular field of specialization who had never participated in an economics experiment before. There were 149 evaluable subjects in the NFC and 76 in the FC treatment, respectively. Six subjects in the NFC and one subject in the FC treatment were subsequently excluded on the basis of behaving randomly during the experiment.<sup>5</sup> Therefore, our empirical analysis is based on 143 NFC and 75 FC subjects, respectively.

### 3 Empirical Evidence on Choice Deferral and Choice Consistency

In this section we provide an account of the main aggregate-level findings regarding deferral decisions and the effect of forced choice on choice consistency.

#### 3.1 The Use of Deferral

In the NFC treatment subjects have the option of deferring choice. This feature of the design allows us to study how often subjects defer, the dynamics of deferral behavior, and whether deferrals relate to the number of choice alternatives available at the menu. We analyze deferrals at non-singleton and singleton menus separately, since deferrals at the latter menus are dominated decisions.<sup>6</sup>

We start with deferrals at non-singleton menus. Figure 1 displays a histogram of the number of non-singleton deferrals for all 143 subjects. On average subjects defer at 4.49 menus, i.e., in about one out of every five non-singleton menus. However, there is a lot of heterogeneity in subjects’ deferral behavior. As it can be seen in Figure 1, while 47.6% of the subjects do not ever defer at non-singleton menus, 52.4% of the subjects defer at least once at non-singleton menus. The latter subjects defer at an average of 8.56 non-singleton menus, with most subjects deferring at between one and twelve menus, and 8 subjects deferring at all or all but one or two non-singleton menus. Overall, we can thus conclude that about half of the subjects defer choice,

---

<sup>4</sup>12 out of the original 237 subjects (5%) were excluded in this way.

<sup>5</sup>The latter criterion was determined by a subject being involved in choice reversals or in a generalized inconsistency score (see Sections 3 and 4, respectively) that exceeded the 2.5% cut-off values in the distribution that resulted from the simulations of one million random-behaving subjects per treatment (see Online Appendix 6 for more details).

<sup>6</sup>However, note that the expected monetary loss from deferring at a singleton menu is  $\frac{1}{31} \cdot \frac{1}{4} \cdot \pounds 1 = 0.8$  pence.

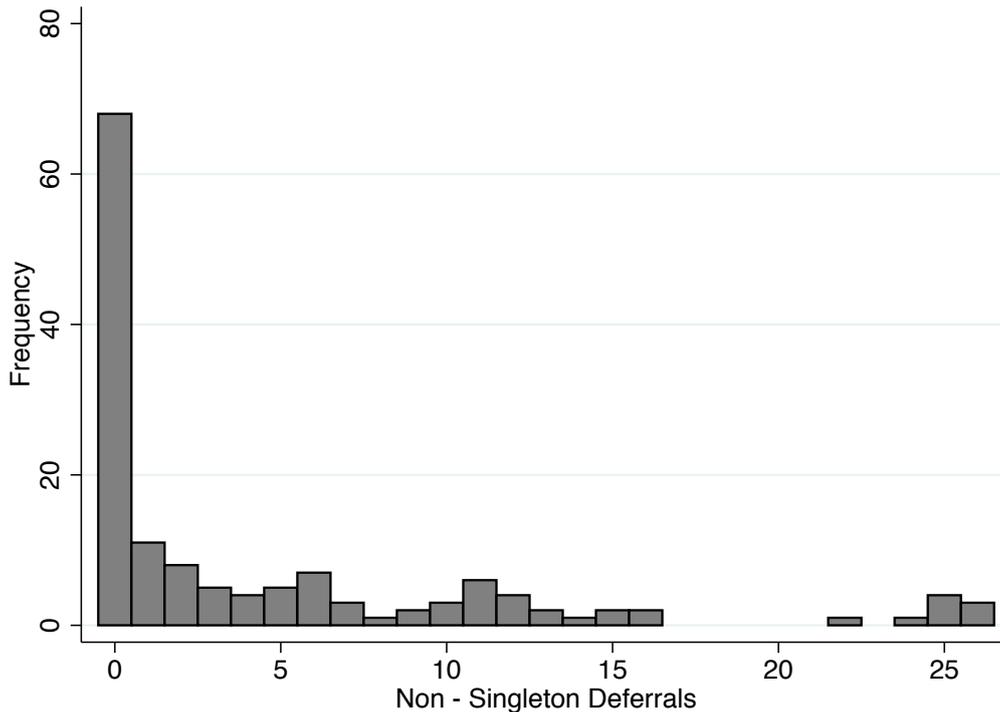


Figure 1: Number of subjects according to their number of deferrals at non-singleton menus.

and that among those who do so, the rate of deferrals at non-singleton menus is around one third.

We now examine the extent to which deferrals at non-singleton menus depend on menu size. Figure 2 plots the relative frequency of deferral at menus with 2, 3, 4 and 5 alternatives. The figure shows that the relative frequency of deferrals, if anything, is a slightly decreasing function of the number of choice alternatives at the menu. This tells us that at aggregate level there is no evidence that deferrals are driven by *choice overload*.<sup>7</sup>

We now consider deferrals at singleton menus in detail. While 84 out of the 143 subjects choose at all singleton menus, 49 subjects defer in at least one singleton menu. None of the subjects in this latter category defer at all 5 singletons. The median number of deferrals at singletons is 0, while the average number of singleton deferrals per subject is 1.07. Therefore, the relative frequency of dominated decisions at singleton menus is 21.7%. The subjects who defer at singleton menus, do so 2.59 times on average, although 2 is the number of deferrals most often observed.

One might wonder if the deferrals observed in the experiment are driven by *choice fatigue*, i.e., that as the experiment progresses subjects get progressively tired of choosing and, as a

<sup>7</sup>Iyengar and Lepper (2000) (see also the meta-analysis by Chernev et al, 2015) showed that the increased complexity associated with making an active choice in large menus of real goods often leads to choice deferral in such menus, a phenomenon that is commonly referred to as “choice overload”. However, we note that Iyengar and Lepper (2000) used menus that contained 6 to 30 alternatives, and therefore, much larger than those used in our experiment.

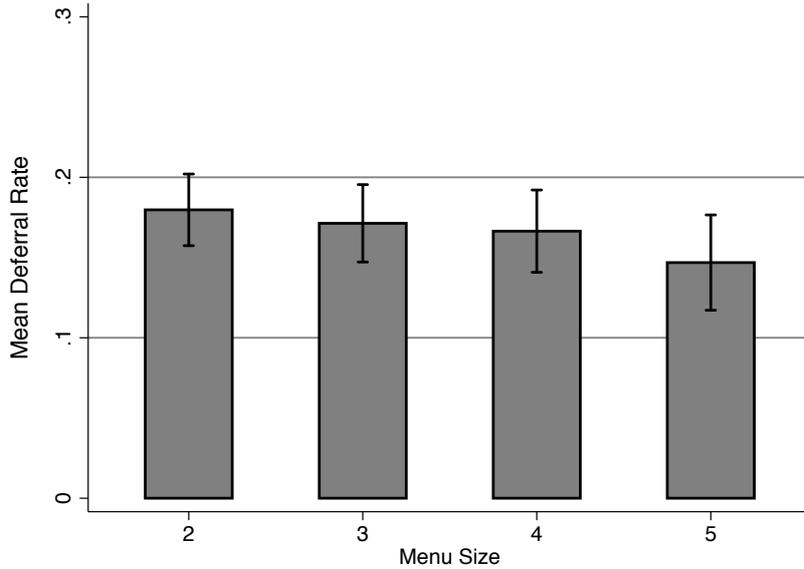


Figure 2: Relative frequency of deferrals ( $\pm$  standard errors) sorted according to menu size.

consequence, defer choice more often at later menus than at earlier menus.<sup>8</sup> In order to examine this hypothesis, we analyze 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, which we display in Figure 3. As can be seen in the figure, 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 run a logit regression of deferrals on the position of the menu in the sequence of all menus at which the deferral occurred.<sup>9</sup> That regression finds that these two variables are not correlated. We can thus conclude that there is no evidence of choice fatigue in our data, which facilitates the data analysis as it allows us to treat decisions at the start and at the end of the experimental session equally.

Given the empirical evidence that subjects defer, it is useful to understand what drives such behavior. As we noted in the introduction, an intuitive hypothesis is that choice deferrals are driven by subjects' indecisiveness. We assess this hypothesis by checking whether *psychologically indecisive* subjects, as measured by their score in the Germeijs and Boeck (2002)'s indecisive personality questionnaire that we administered at the end of the experimental session, tend to defer more often than *psychologically decisive* ones. We compare the average number of deferrals at non-singleton menus of the subjects in the top tertile of the distribution of psychological decisiveness scores with that of the subjects in the bottom third of the distribution. These figures are 3.0 and 6.6, respectively, a difference which we assess with a Mann-Whitney  $U$ -test which yields a  $p$ -value of 0.039 ( $n = 96$ ).<sup>10</sup> Further evidence that higher psychological decisiveness

<sup>8</sup> Augenblick and Nicholson (2016) provided evidence for choice fatigue in the context of actual voting decisions.

<sup>9</sup> Singleton menus were excluded from this regression but their inclusion does not affect the conclusion.

<sup>10</sup> When we restrict ourselves to subjects who never deferred at singletons, the corresponding numbers are 1.7 and 7.4, and the corresponding test statistic is ( $p = 0.006, n = 57$ ).

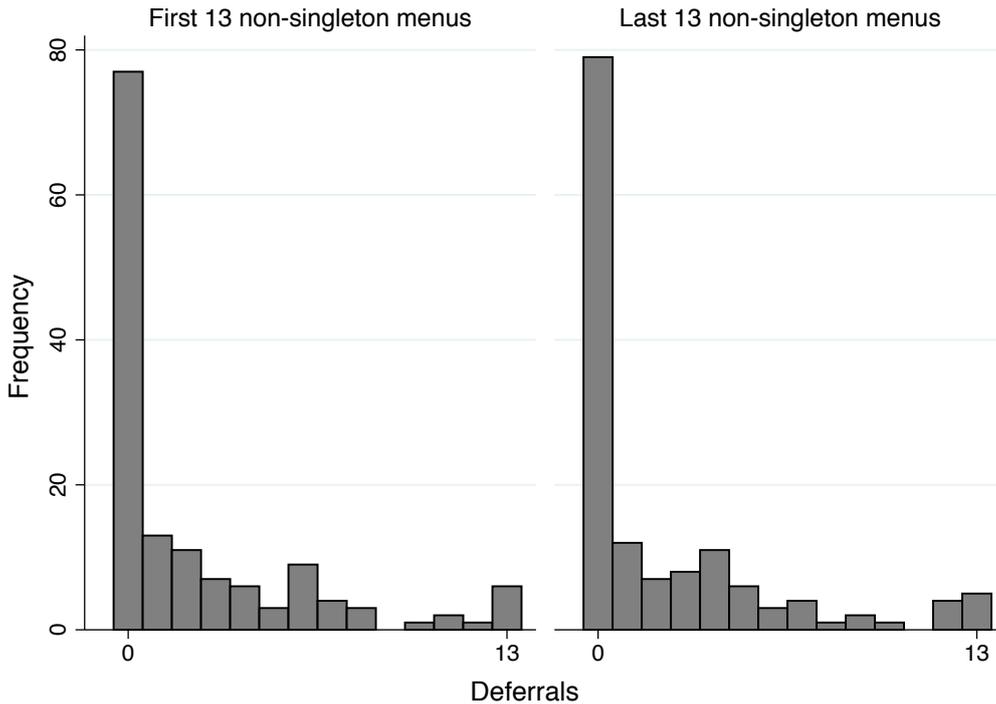


Figure 3: Number of subjects for different intervals of deferral rates.

scores are negatively correlated with the number of deferrals emerges from Spearman’s rank correlation test, which yields  $\rho = -0.161$  ( $p = 0.055$ ,  $n = 143$ ). This correlation gets stronger when we focus on the 84 subjects who never defer at singletons, as the Spearman’s  $\rho$  now becomes  $-0.309$  ( $p = 0.004$ ,  $n = 84$ ). Overall, we can thus conclude that, in our experiment, psychological indecisiveness is positively correlated with deferral behavior.

### 3.2 The Effects of Forced Choice on Consistency: Reversals and Cycles

Our first pass at analysing choice inconsistency in our data focuses on *choice reversals* as identified by the total number as well as the relative frequency of WARP violations.<sup>11</sup> We compare subjects’ WARP violations in the FC and NFC treatments. We conduct the analysis in two different ways. One of the ways treats choice as the expression of strict preference. That is the typical approach that other choice experiments follow, and therefore it makes our study more easily comparable to other studies. The other way treats choices as the expression of preferences that might be weak. To do the former, we proceed as usual in the literature: whichever alternative is chosen is interpreted as strictly preferred to all the other alternatives in the menu. To do the latter we take into account subjects’ statements after choices at binary menus in the FC and NFC treatments. As discussed in the previous section, after choosing an alternative at a binary menu, subjects were required to state whether they preferred it to the other alternative, were indifferent, or whether they chose it for some “other reason”. As mentioned earlier, stating

<sup>11</sup>See also Famulari (1995) and Gross (1995).

indifference removes the possibility of the subject being able to revise her choice at a latter stage, which is instead randomly determined for her. Stating indifference is, therefore, an undominated option for a truly indifferent subject who thinks that inspecting the alternatives on that menu, if it were to be randomly selected, would not lead her to break the indifference between the pair of alternatives in consideration. Nevertheless, stating indifference is also clearly never a strictly dominant option, because a truly indifferent subject would not be better off by stating her indifference.

In allowing for weak preferences, we give priority to actual choice data over indifference statements when the two disagree. We augment choice data with indifference as follows: we first consider all indifference statements following choice and for each of them check if the stated indifference between the pair of alternatives generates a violation of weak-preference transitivity, given (i) the subject’s choices at all other binary menus, (ii) the subject’s indifference statements at binary menus, as well. Any indifference statement that generates a cycle is ignored, in which case we view an active choice at a binary menu as an expression of strict preference between the alternatives.<sup>12</sup> As it turns out, this procedure produces weakly lower WARP violations for all but one subjects in our sample. A complete explanation of the way in which our choice datasets were augmented with indifference statements is provided in Online Appendix 4.

For similar reasons, we ignore indifference statements following deferral. Such statements are dominated as they lead to a loss of  $\mathcal{L}1$  to the subject if that menu is randomly selected, with her abdicating the ability to choose one of the alternatives (indeed, stating indifference here leads to one of the alternatives being chosen for her randomly).<sup>13</sup> Indifference statements are incorporated in the sense that if a subject states indifference between, say  $x$  and  $y$ , then whenever she chooses either  $x$  or  $y$  from a menu containing both alternatives (i.e., not just the binary menu with  $x$  and  $y$ ), we “augment” her set of choosable options from that menu to  $\{x, y\}$ .

It is worthwhile noting that the choice data augmented with indifference is not very different from the starting choice as on average subjects state indifference just in one out of ten binary menus they face, and only around half of those statements are used in augmenting the choice data. Table 1 gives precise figures of this data.

Table 1: Subjects’ average number of indifference statements (standard errors in parenthesis).

NFC		FC	
Stated	Included	Stated	Included
1.140	0.329	0.973	0.600
(0.164)	(0.107)	(0.234)	(0.199)
$n = 143$	$n = 143$	$n = 75$	$n = 75$

Our first results on choice consistency are presented in Table 2. This table displays subjects’

<sup>12</sup>For instance, consider a subject who chooses  $y$  at menu  $\{x, y\}$ ,  $z$  at menu  $\{y, z\}$  and  $x$  also at menu  $\{x, z\}$ , but who then states that she is indifferent between  $x$  and  $z$ . Such an indifference statement would result in a cycle and is therefore disregarded.

<sup>13</sup>Another possible approach is to instead ignore deferral when followed by an indifference statement, replacing such decision with the augmented choice of both alternatives. All our results continue to hold when we treat indifference following deferrals in this way.

average number of WARP violations in each treatment, when indifference statements are taken into account (left column) and when they are disregarded (right column). Mann-Whitney  $U$  tests provide support for the hypothesis that non-forced choice settings lead to less inconsistency in the sense of fewer choice reversals.

Table 2: Average number of WARP violations in each treatment. Standard errors in parenthesis and  $p$ -values from Mann-Whitney  $U$  tests.

	<b>WARP</b> (w/ indifference)	<b>WARP</b> (w/o indifference)
<b>NFC</b>	1.189 (0.228)	1.308 (0.351)
$n =$	143	143
<b>FC</b>	2.84 (0.627)	3.2 (0.627)
$n =$	75	75
$p$ -value =	0.023	0.003

However, it can be argued that, since subjects can and do defer in the NFC treatment, they end up making fewer active choices in the NFC treatment, which in turn might bias the number of WARP violations downward. To address this concern, we redo the analysis in Table 2 normalizing the number of WARP violations of each subject by the number of choices (non-deferrals) at non-singleton menus. We present the results in Table 3. They provide further support for the hypothesis that non-forced choice settings lead to fewer choice reversals.

Table 3: Average number of WARP violations in each treatment normalized by the total number of active choices made by each subject. Standard errors in parenthesis and  $p$ -values from Mann-Whitney  $U$  tests.

	<b>WARP</b> (w/ indifference)	<b>WARP</b> (w/o indifference)
<b>NFC</b>	0.049 (0.010)	0.055 (0.010)
$n =$	140	140
<b>FC</b>	0.109 (0.024)	0.123 (0.024)
$n =$	75	75
$p$ -value =	0.036	0.005

A different and more restricted measure of consistency that we also consider is the number of *binary choice cycles* in our data. An example of such a cycle would be the choice pattern whereby  $x$  is chosen over  $y$  in menu  $\{x, y\}$ ,  $y$  is chosen over  $z$  in menu  $\{y, z\}$  and  $z$  is chosen over  $x$  in menu  $\{x, z\}$ . Interestingly, very few subjects display binary choice cycles: only 3 subjects out of 143 in NFC and 5 out of 75 in FC. The average number of cycles is 0.02 in NFC and 0.15 in FC. Despite the large relative difference in these averages, we cannot reject the null hypothesis of no difference at the 5% level given the very low number of cycles in our sample ( $p = 0.083$ , Mann-Whitney  $U$ -test).

### 3.3 Forced Choice and Psychological Indecisiveness

Our previous results raise the possibility that treatment effects might be strongest for *psychologically indecisive* subjects. We therefore categorize subjects with psychological decisiveness scores in the top and bottom tertiles of the distribution as *psychologically decisive* and *psychologically indecisive*, respectively, and analyze treatment effects separately for these two subgroups. Tables 4 and 5 present the results of this analysis for the case of non-normalized and normalized numbers of choice reversals, respectively.

Table 4: Average WARP violations in each treatment for psychologically decisive and indecisive subgroups. Standard errors in parenthesis and p-values from Mann-Whitney U tests.

	<i>Psychologically decisive</i>		<i>Psychologically indecisive</i>	
	<b>WARP</b> (w/ indifference)	<b>WARP</b> (w/o indifference)	<b>WARP</b> (w/ indifference)	<b>WARP</b> (w/o indifference)
<b>NFC</b>	1.46 (0.459)	1.56 (0.477)	0.727 (0.263)	0.864 (0.271)
<b>FC</b>	0.833 (0.389)	1.125 (0.464)	3.786 (1.293)	4.214 (1.275)
<i>n</i> =	74	74	72	72
<i>p-value</i> =	0.860	0.959	0.031	0.005

Table 5: Average WARP violations normalized by the number of active choices in each treatment for psychologically decisive and indecisive subgroups. Standard errors in parenthesis and p-values from Mann-Whitney U tests.

	<i>Psychologically decisive</i>		<i>Psychologically indecisive</i>	
	<b>WARP</b> (w/ indifference)	<b>WARP</b> (w/o indifference)	<b>WARP</b> (w/ indifference)	<b>WARP</b> (w/o indifference)
<b>NFC</b>	0.062 (0.020)	0.066 (0.021)	0.031 (0.011)	0.036 (0.011)
<b>FC</b>	0.032 (0.015)	0.043 (0.018)	0.146 (0.050)	0.162 (0.049)
<i>n</i> =	73	73	70	70
<i>p-value</i> =	0.798	0.976	0.043	0.008

The data shows that while psychologically decisive subjects violate WARP equally often in the FC and NFC treatments, psychologically indecisive ones violate WARP more often when they are in a forced choice environment. The finding does not depend on whether we consider choice data with or without indifference augmentation and on whether we consider original or normalized data. Thus, consistent with our interpretation of indecisiveness as a driver of choice deferral, we find that a forced choice environment is detrimental to choice consistency *only* for subjects that are psychologically indecisive.

## 4 Rational Choice Deferral: Theory and Evidence

In this section we first outline three models of “rational” choice deferral in the sense that, under certain conditions which we clearly state, the decision maker’s active choices are always based on maximization of a stable preference relation and are therefore WARP-consistent. Then, we take

these models to our experimental data and provide a categorization of all subjects in the non-forced-choice treatment on the basis of their proximity to each model. To do so we develop a new combinatorial-optimization technique that generalizes and extends in new directions the well-known and increasingly applied measure of choice consistency that was proposed by Houtman and Maks (1985).

## 4.1 Models Compatible with the Experiment’s Incentives

In this subsection we lay out two models where the agent’s decision to actively choose a feasible option or to defer can be thought of as being driven by the incentives that are built in our experimental design. The first is a simple Bayesian model of rational costly information acquisition. The second is a model of rational indecisiveness for an agent with incomplete preferences. To this end, we let  $X$  be the set of all possible choice alternatives and assume it is finite. We also let  $\mathcal{M}$  denote the collection of all subsets of  $X$ . A *menu* is a nonempty set in this collection. A choice correspondence  $C : \mathcal{M} \rightarrow \mathcal{M}$  is a possibly multi-valued and empty-valued mapping that satisfies  $C(B) \subseteq B$  for all  $B \in \mathcal{M}$  and associates every menu with the alternatives that the decision maker might choose from it. We interpret  $C(B) = B$  and  $C(B) = \emptyset$  as capturing the situations in which all and none of the alternatives are choosable, respectively. In the latter case we understand that the agent defers choice. What the agent chooses after she has deferred at some menu is unimportant in our context. Our aim instead is to understand the mechanisms that may lead to deferral in the first place.

### 4.1.1 A Bayesian Model of Rational Information Acquisition

We first consider the case where the agent is portrayed as having stable and completely ordered preferences over the outcomes in  $X$  and also as deciding between the actions of choice,  $c$ , or deferral,  $d$  whenever presented with a menu while at the same time facing uncertainty about whether the additional information to be obtained will lead to a “meaningful” update of his preferences—in the sense that the new information will change the set of best elements in the given menu— or not. When presented with menu  $B$ , the agent here is also assumed to have a menu-dependent prior belief  $q_B$  concerning the event that the new information will *not* change his preferences in the above sense, and an additional menu-dependent prior  $\alpha_B$  that, conditional on her ex post preferences having changed in the above sense, he will choose according to these updated preferences. The state space in this model is therefore  $S = \{n, y_1, y_2\}$ , where  $n$  stands for “no change”,  $y_1$  for “change and conformity with the updated preferences” and “ $y_2$  for change and conformity with the prior preferences”. The reason why the latter state may be relevant for the decision maker is that a choice reversal is costly. The action set is  $A = \{c, d\}$  and the agent’s preferences over action-state pairs are represented by a state-dependent and menu-independent utility function  $v : A \times S \rightarrow \mathbb{R}$ .

When deciding between choosing and deferring at menu  $B$  in the context of our experimental design, a risk-neutral Bayesian agent's expected utility from  $c$  and  $d$ , respectively, is given by

$$U(c; B) = q_B \cdot v(c, n) + (1 - q_B) \cdot [\alpha_B \cdot (v(c, y_1) - p_r) + (1 - \alpha_B) \cdot v(c, y_2)] \quad (1)$$

$$U(d; B) = q_B \cdot v(d, n) + (1 - q_B) \cdot v(d, y_1) - p_d \quad (2)$$

where  $p_d$ ,  $p_r$  denote the fixed costs of deferral and choice reversal, respectively, and  $\alpha_B$  is the probability with which the agent will actually reverse his prior choice at menu  $B$  after a change in his preferences following the acquisition of additional information. Consistent with our experimental design where  $p_r = \pounds 4$  and  $p_d = \pounds 1$ , we assume  $p_r > p_d > 0$ . It also follows from our risk-neutrality assumption that  $p_r = 4p_d$ , with the precise values of  $p_r$  and  $p_d$  depending on the chosen utility scale, which in turn is unique up to a positive affine transformation.

Since all sources of disutility for the agent are captured by the costs  $p_r$  and  $p_d$ , and since, by construction, our experimental design rules out any considerations of payoff discounting, it is reasonable to assume

$$v(d, y_1) = v(d, n) = v(c, n) = v(c, y_1) > v(c, y_2) = v(d, y_2).$$

The first three equations are justified on the basis that in each of the four relevant state-action pairs the agent ends up with his most preferred alternative, either according to her unchanged prior preferences or according to her modified posterior preferences. Similarly, the equality between  $v(c, y_2)$  and  $v(d, y_2)$  reflects the fact that in both situations the agent ends up with her best option according to her prior preferences. The inequality, finally, between the first four and the last two utility levels is due to the fact that in the first case the agent chooses an alternative that is strictly better according to her posterior preferences.

Since there are only two distinct utility values and  $v$  is unique up to a positive affine transformation, we can normalize it so that

$$v(d, y_1) = 1 \quad \text{and} \quad v(c, y_2) = 0.$$

It now follows from the above that

$$U(c; B) > U(d; B) \iff p_d - a_B(1 - q_B)p_r > (1 - q_B)(1 - a_B). \quad (3)$$

Thus, the Bayesian agent makes an active choice at menu  $B$  if and only if the difference between the cost of deferral and the expected cost of a choice reversal exceeds a certain threshold, which, in this particular utility scale, coincides with the probability that he will be in the state  $y_3$  where his preferences are updated but his first-period active choice at  $B$  is not modified accordingly.

In light of the above, if the agent's ex ante preferences over the elements in  $X$  are represented

by the utility function  $u : X \rightarrow \mathbb{R}$ , then, for every menu  $B$ ,

$$C(B) = \begin{cases} \arg \max_{x \in B} u(x), & \text{if } p_d - a_B(1 - q_B)p_r > (1 - q_B)(1 - a_B) \\ \emptyset, & \text{otherwise} \end{cases}, \quad (4)$$

where we assume that the agent is never indifferent between deferring at menu  $B$  and actively choosing an element from this menu.

The first prediction of (4) is that whenever the agent does choose, his choices will be compatible with WARP, and hence will exhibit no weak choice reversals across menus for any two alternatives  $x$  and  $y$ . Formally,

$$x \in C(B), \quad y \in B \setminus C(B) \quad \text{and} \quad y \in C(D) \implies x \notin D. \quad (5)$$

Indeed, whenever the agent makes an active choice, he does so by maximizing the menu-independent utility function  $u$ . It is obviously true that if this agent was forced to choose from all menus, he would be a standard utility maximizer and his choices would satisfy WARP. Consequently, any subset of this choice dataset is also compatible with WARP.

Another prediction derived by the model is that the agent will never defer at singleton menus. Formally,

$$x \in X \implies C(\{x\}) = \{x\}. \quad (6)$$

The reason for such behavior is the irrelevance of additional information in menus where there is just one feasible option. We will refer to (6) as the *Desirability* condition because it suggests that every alternative is considered to be sufficiently good for the decision maker to choose it.

If the agent chooses  $x$  at some menu  $B$ , then his prior belief that the additional information will result in some other alternative in  $B$  becoming strictly preferred to  $x$  is very low. Therefore, rationality requires that such an agent will also find it unlikely that such a preference change can occur at any submenu  $D \subset B$  in which  $x$  remains feasible. Therefore, we will also assume that this Bayesian decision maker's beliefs are such that his active choices satisfy the following:

$$x \in C(B) \quad \text{and} \quad x \in D \subset B \implies C(D) \neq \emptyset \quad (7)$$

This is a mild weakening of the well-known *Contraction Consistency* or *Property  $\alpha$*  axiom and requires that whenever an alternative  $x$  that is choosable in a given menu  $B$  is also feasible in a submenu  $D$  of  $B$ , then the agent does not defer at  $D$ . Together with WARP, (7) implies that under this condition the alternative  $x$  is indeed choosable at  $D$ .

Now suppose the agent chooses  $x$  from some menu  $B$  where  $y$  is also feasible, and also chooses  $y$  from some other menu  $D$ . As above, the agent's priors that his preferences will change following information acquisition are very low at both menus. Therefore, since his ex ante preferences

are represented by the utility function  $u$  and are therefore transitive, it follows from  $x$  being weakly preferred to every alternative in  $A$  (including  $y$ ) and  $y$  being weakly preferred to every alternative in  $B$  that  $x$  is optimal across all elements in  $A$  and  $B$ . This implies, in particular, that the following is satisfied:

$$x \in C(B), y \in B \text{ and } y \in C(D) \implies C(B \cup D) \neq \emptyset \quad (8)$$

This is a mild weakening of the *Strong Expansion* axiom which requires  $x \in C(B \cup D)$  whenever  $x \in C(B)$ ,  $y \in B$  and  $y \in C(D)$ .<sup>14</sup> Indeed, (8) only requires that the agent does not defer at the menu  $B \cup D$  when  $x$  and  $y$  are choosable and feasible at  $B$ , respectively, and  $y$  is choosable at  $D$ . Together with WARP, this restriction again ensures that  $x \in C(B \cup D)$  must actually hold.

Both (7) and (8) impose an intuitive and normatively appealing consistency structure on the agent's revealed preferences that goes beyond the no-choice-reversal restrictions imposed by WARP. Specifically, if choice is rational in the sense that it reveals weak preference, then  $x$  being choosable at  $B$  and feasible at  $D \subset B$  means that  $x$  is weakly preferred to everything in  $D$  and hence that it should remain choosable at this menu. Deferral in this case is therefore irrational. Similarly, if  $x$  is weakly preferred to  $y$ , then it is also a basic rationality principle that  $x$  is weakly preferred to everything that  $y$  is weakly preferred to. Hence, deferring at a menu  $B \cup D$  when  $x \in C(B)$ ,  $y \in B$  and  $y \in C(D)$  is also irrational in the same way.

We note that, while conditions (5) – (8) are implied by the structure that we have imposed on this Bayesian model, it is unknown to us if they are also sufficient for the existence of a state space  $S$ , an action set  $A$ , a function  $v : S \times A \rightarrow \mathbb{R}$ , a collection of menu-dependent beliefs  $\{q_B, \alpha_B\}_{B \in \mathcal{M}}$  and a function  $u : X \rightarrow \mathbb{R}$  that would make (1), (2) and (4) true. This is in contrast to the model that we lay out below, an important special case of which is fully characterized by these axioms.

#### 4.1.2 A Model of Rational Indecisiveness

Suppose now that, rather than being a fully rational Bayesian, the decision maker is *indecisive*. Formally, this means that he has preferences over  $X$  that are captured by a reflexive and transitive relation (preorder)  $\succsim$  that may be *incomplete*. In other words, there may exist alternatives  $x$  and  $y$  that are *incomparable* in the sense that  $x \not\prec y$  and  $y \not\prec x$ . That such situations often arise and lead to choice deferral is introspectively obvious and also supported by data from (usually hypothetical) consumer decisions over multi-attribute options that feature value conflicts/trade-offs.<sup>15</sup> As opposed to the Bayesian model, subjects who defer due to indecisiveness during the main phase of our experiment can be thought of as doing so in order to *complete* their

<sup>14</sup>This axiom in turn (which is referred to as *Property  $\gamma+$*  in Salant and Rubinstein, 2008) is stronger than the well-known *Property  $\gamma$*  which states that if  $x \in C(A)$  and  $x \in C(B)$ , then  $x \in C(A \cup B)$  (Sen, 1971).

<sup>15</sup>See, for example, Tversky and Shafrir (1992), Dhar (1997), Dhar and Simonson (2003), Anderson (2003) and Bhatia and Mullett (2016).

preferences in the second stage after receiving additional information about the alternatives. As already mentioned, Sen (1997) has referred to this kind of preference incompleteness that is rooted in the lack of information as *tentative*, and used the term *assertive incompleteness* to refer to situations where the alternatives in question are not rankable even after all the available information about them has been received.

Faced with such indecisiveness constraints, a rational agent will obviously choose a most preferred alternative from every menu where such an item exists. If this is not possible, however, then she must decide between deferring and choosing an *undominated* feasible option. Conditional on there being no most preferred alternative in the given menu, an undominated option is not worse than any other item in the menu without at the same time being better than everything else (i.e. for each such option there is at least one other feasible alternative with which it cannot be compared). We denote the sets of most preferred/dominant and non-inferior/undominated feasible options at menu  $D$  by

$$B_{\succsim}(D) := \{x \in D : x \succsim y \text{ for all } y \in D\} \quad \text{and} \quad M_{\succsim}(D) := \{x \in D : y \not\succeq x \text{ for all } y \in D\}$$

respectively. Since preferences are allowed to be incomplete, the set  $B_{\succsim}(D)$  may be empty. However, since they are also assumed to be transitive, the set  $M_{\succsim}(D)$  is always nonempty.

In the absence of a most preferred item in the menu, the decision between deferring and choosing an undominated option ultimately depends on the costs associated with these decisions. To organize our thinking about such costs we define a *decision cost function* as a mapping  $\phi_{\succsim} : \mathcal{M} \rightarrow \mathbb{R}_+$  and interpret  $\phi_{\succsim}(D)$  as the cost of choosing from menu  $D$  or the cost of deferring if  $D = \emptyset$ . Intuitively, choice from  $D$  is decision-costless if and only if a most preferred option exists in that menu. On the other hand, if such an option does not exist, choice is costly as it entails the possibility that the chosen item will turn out to be inferior to some other non-chosen alternative after the agent's preferences have been completed. In this case the cost could be monetary (e.g. having to return the item to the store) or psychological (e.g. regret). We will refer to a decision cost function  $\phi_{\succsim}$  as *rational* if, for every menu  $D$ ,

$$\phi_{\succsim}(D) = 0 \iff B_{\succsim}(D) \neq \emptyset \tag{9}$$

It is easy to show that a rational decision cost function exists.<sup>16</sup>

Suppose now that the agent is endowed with such a function  $\phi_{\succsim}$ . Suppose also that, as in our experimental design, deferral is costly for this individual, and this cost is the same for every menu in which deferral takes place, i.e.  $\phi_{\succsim}(\emptyset) := p_d > 0$ . Rational decision making in this

<sup>16</sup>Such an example is the function  $\phi_{\succsim}$  defined by

$$\phi_{\succsim}(D) = \begin{cases} p_d, & \text{if } D = \emptyset \\ |D| - \max_{z \in D} |L_{\succsim}(z) \cap D|, & \text{if } D \neq \emptyset \end{cases}$$

where  $p_d \geq 0$  is a constant,  $L_{\succsim}(z) := \{y \in X : z \succsim y \text{ for all } y \in X\}$  and  $|D|$  is the cardinality of  $D$ .

environment entails conforming with the choice rule whereby

$$C(D) = \begin{cases} B_{\succsim}(D), & \text{if } \phi_{\succsim}(D) = 0 \\ \emptyset, & \text{if } \phi_{\succsim}(D) > p_d \\ M_{\succsim}(D), & \text{if } 0 < \phi_{\succsim}(D) < p_d \end{cases} \quad (10)$$

In words, this procedural rule predicts that whenever the agent is presented with a menu, she first checks whether she can find a most preferred alternative or not. If she can, then (9) and (10) imply that she chooses one of these best options and is therefore a standard utility-maximizer. If she cannot, then she decides between deferring and choosing randomly from the set of undominated feasible options on the basis of whether deferring is less or more costly compared to choosing an undominated option from that menu.<sup>17</sup>

Importantly, if  $\phi_{\succsim}(D) > p_d$  for all menus  $D$  in which no dominant option exists, then (10) reduces to

$$C(D) = \begin{cases} B_{\succsim}(D), & \text{if } B_{\succsim}(D) \neq \emptyset \\ \emptyset, & \text{otherwise} \end{cases} \quad (11)$$

In this reduced model of *rational indecisiveness* the agent actively chooses from a menu if the latter contains a most preferred option, and defers otherwise. Although this agent is not Bayesian in the sense discussed above because her ex ante preferences are generally incomplete, when she operates in the incentive environment that is put forward by our experimental design it is natural for her decision cost  $\phi_{\succsim}(D)$  of choosing from a menu  $D$  where a most preferred option does not exist to be thought of as coinciding with the *expected cost* of a choice reversal at that menu. Given that our experiment features  $p_r = 4p_d$  if a risk-neutrality assumption is again employed, by letting  $s_r^D$  denote the agent's subjective belief that she will change her choice in menu  $D$  after receiving more information about the elements of that menu we get  $\phi_{\succsim}(D) = p_r \cdot s_r^D > p_d = \phi_{\succsim}(\emptyset)$  if and only if  $s_r^D > \frac{1}{4}$ . This condition is mild and will always be satisfied, for example, if at least one of the five alternatives that our subjects were presented with was considered to be dominated at the grand menu that contained all these five options.<sup>18</sup>

Importantly, if  $s_r^D > \frac{1}{4}$  is satisfied by all subjects and in all menus, then the hypothesis that decision makers are "rational indecisives" in the sense of (11) becomes fully testable. Indeed,

<sup>17</sup>This procedural model can be further enriched by letting the choice-reversal probabilities depend on the menu as well as on the alternatives in the set  $M_{\succsim}(D)$ . For example, if  $D = \{w, s, y, z\}$  and  $M_{\succsim}(D) = \{s, y\}$ , then  $s_r^{D,s}$  and  $s_r^{D,y}$  denote the probabilities with which the agent will reverse her choice if she chooses  $s$  and  $y$ , respectively, from menu  $D$ . An intuitive restriction to impose in this more refined version of the model is to let the agent's decision cost at  $D$  coincide with the *minimum* expected cost of a choice reversal that is associated with the elements of  $M_{\succsim}(D)$ , because a rational decision maker would obviously choose the undominated option that minimizes the probability of a choice reversal. Formally,  $\phi_{\succsim}(D) = \min_{x \in M_{\succsim}(D)} s_r^{D,x} \cdot p_r$ .

<sup>18</sup>Indeed, since all other menus in our experiment contained four or fewer options, the presence of one dominated option ensures that the set  $M_{\succsim}(D)$  contains at most three options at all menus  $D$  with up to four elements except, possibly, the four-element menu where the dominated option is non-feasible (in which case  $s_r^D = \frac{1}{4}$  may hold).

the model that is laid out in (11) is characterized by WARP/(5), Desirability/(6), Contraction Consistency/(7) and Strong Expansion/(8).<sup>19</sup> In particular, unlike the Bayesian model where these four conditions are only known to be necessary, in the context of the indecisiveness model they are both necessary and sufficient.<sup>20</sup>

## 4.2 Unattractiveness-Constrained Utility Maximization

Even when an agent can completely order all feasible alternatives according to his preferences and is absolutely certain that no additional information is going to change them, she may still defer choice at menus where she considers no alternative to be “sufficiently” attractive or desirable.<sup>21</sup> Such a decision maker can be portrayed as having a desirability or unattractiveness threshold that takes the form of some alternative  $x^*$ , and also as choosing like a standard utility maximizer whenever the most preferred feasible alternative is strictly preferred to  $x^*$  and deferring otherwise. Formally, the agent here behaves as if his preferences over  $X$  were represented by a utility function  $u : X \rightarrow \mathbb{R}$  and, for every menu  $D$ ,

$$C(D) = \begin{cases} \arg \max_{x \in D} u(x), & \text{if } \max_{x \in D} u(x) > u(x^*) \\ \emptyset, & \text{otherwise} \end{cases} \quad (12)$$

This model of *unattractiveness-constrained* utility maximization is characterized<sup>22</sup> by WARP, Contraction Consistency and the following two axioms:

$$C(\{x\}) = \emptyset \text{ for some } x \in X \quad (13)$$

$$C(B) = \emptyset \text{ and } D \subset B \implies C(D) = \emptyset \quad (14)$$

These will be referred to as *Undesirability* and *Contractive Undesirability*, respectively. The former simply ensures the existence of some unattractive alternative. The latter imposes consistency on the agent’s perception of unattractiveness in the sense that if all feasible options in some menu are of that kind, then this remains the case in every submenu as well.

We note that our experimental design provides incentives for subjects to *not* conform with (13), even if their actual preferences *do* feature an undesirability threshold that renders some alternatives unacceptable. In particular, when singleton menus are presented during the main phase of the experiment, it is rational for all decision makers to actively choose instead of deferring. This is so because deferring at a singleton menu has no effect on the probability of

<sup>19</sup>See Gerasimou (2015) for more details.

<sup>20</sup>A noteworthy feature of this model is that the “revealed indecisiveness” relation that is induced by it (in the sense that  $x \not\prec y$  and  $y \not\prec x$  if and only if there is no menu  $B$  where either option is chosen and the other is feasible) actually coincides with the revealed indecisiveness relation that is associated with the Bernheim-Rangel (2009) definition of *strict* welfare preferences, according to which an analyst who deals with a complete choice dataset (i.e. with no deferrals) may conclude that  $x$  is strictly preferred to  $y$  if  $y$  is never chosen in the presence of  $x$ . Gerasimou (2015) discusses this further.

<sup>21</sup>See Zakay (1984), Mochon (2013), Dhar (1997) for relevant experimental evidence.

<sup>22</sup>See Gerasimou (2015) for the details.

that menu being randomly selected at the end as the “winning” menu. Thus, a subject that is randomly selected to win the item in that singleton menu will win this item regardless of whether she had initially chosen it or not. The important difference, however, is that in the latter case her £7 endowment would be reduced by £1. Therefore, every subject who prefers more money to less should choose at all singletons. We also note, however, that the other predictions of this model are compatible with the behavior that is induced by our design.

Table 6 summarizes the similarities and differences of the three models of rational choice deferral that were presented above by identifying which of the six behavioral axioms they are compatible with.

Table 6: Similarities and differences in the predictions of the three models of rational choice deferral

	<b>Bayesian</b>	<b>Indecisiveness</b>	<b>Unattractiveness</b>
(5) - <b>WARP</b>	✓	✓	✓
(6) - <b>Desirability</b>	✓	✓	×
(7) - <b>Contraction Consistency</b>	✓	✓	✓
(8) - <b>Strong Expansion</b>	✓	✓	✓
(13) - <b>Undesirability</b>	×	×	✓
(14) - <b>Contractive Undesirability</b>	×	×	✓

### 4.3 Generalizing and Extending the Houtman-Maks Rationality Index

In addition to the frequency of choice reversals which we used in Section 2 to demonstrate a negative effect of forced-choice on consistency, various other indices that capture the degree of a choice dataset’s rationality have been proposed in recent years.<sup>23</sup> Some of them are applicable in choice over competitive budget sets only, while others can in principle be used in more general choice environments. As we noted in the introduction, one well-known and widely applied general measure of this kind is the so-called Houtman-Maks (1985) (henceforth HM) 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 utility maximization. When the dataset under investigation is complete in the sense that it contains no deferral observations, an equivalent way in which the HM index can be interpreted is that it counts the smallest number of changes that need to be made on such a dataset for it to be as if it was generated by utility maximization, with a count or “score” of zero obviously corresponding to perfect conformity with this model. In particular, positive but low HM scores may suggest that the relevant subject does not conform perfectly with utility maximization solely due to one or two errors rather than that his behavior systematically deviates from this model.

While all the above-mentioned measures of consistency have been proposed in the context of complete datasets, an interesting feature of the HM index which we note and build on here is that the reduced model of rational indecisiveness that is laid out in (11), together with the

<sup>23</sup>For example, Echenique et al (2011), Apesteguia and Ballester (2015) and Halevy et al (2012).

above equivalent interpretation of the index, suggests a natural generalization that makes this method applicable also to incomplete datasets that contain deferral observations. Specifically, our proposed *generalized HM* index is defined by the smallest number of changes that need to be made on a possibly incomplete dataset for the latter to be compatible with maximization of a *possibly incomplete* preorder. Here, an HM score of zero can arise not only in complete datasets that are perfectly compatible with maximization of a complete preorder (i.e. with utility maximization) but also in incomplete datasets that are perfectly compatible with the model of rational indecisiveness that is captured in (11), which in turn reduces to utility maximization in the special case where the preorder is complete.

Given this approach through which the HM index can be generalized, and given the fact that there are 6942 distinct *possibly* incomplete preorders on a set of five elements (with 4231 of them being *strict* in the sense that they do not allow for indifference), to make the generalized HM index operational and applicable to the incomplete datasets that have been derived from our experiment both when indifference statements are accounted for and when they are not, we developed a computer algorithm that calculates this index when attention is restricted to the 6401 *incomplete* preorders (respectively, the 4111 strict incomplete preorders) after excluding the 541 complete preorders (respectively, the 120 strict complete preorders).<sup>24</sup> The resulting number for each dataset captures its proximity with the model of rational indecisiveness (or with the Bayesian model for some suitable parameterization of the latter). At the same time, the proximity of each dataset with the standard model of *unconstrained* utility maximization over the set of five options is captured by the original HM method that is only applied over the 541 *complete* preorders (respectively, the 120 strict complete preorders).

Similarly, to find the proximity of each incomplete dataset with the model of unattractiveness-constrained utility maximization we extended the original HM index in a different direction. Specifically, as suggested by the relevant model that is laid out in (12), we developed a computer algorithm that calculates the smallest number of changes that need to be made in such a dataset for it to be as if it was generated by maximization of a “truncated” complete preorder, whereby any possible such ordering that features anything between one and five options that fall below a desirability/choosability threshold are included in the computational process. The latter effectively checks for conformity with *unconstrained* utility maximization over sets of zero (trivial case), one, two, three and four elements.

## 4.4 Empirical Findings

The procedure that was described in the previous subsection equips us with four HM-based “scores”: the standard HM score which corresponds to unconstrained utility maximization;

---

<sup>24</sup>Details on the total number of orderings that are of interest to us in this paper can be found 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>.

the generalized HM score which corresponds to proximity of a dataset with maximization of a possibly incomplete preorder; the rational indecisiveness HM score which is derived from the generalized HM method when the latter is restricted to strictly incomplete preorders; and the extended HM score that corresponds to unattractiveness-constrained utility maximization. Our classification of each dataset into one of the available models of utility maximization or rational choice deferral is made according to the corresponding HM method that has resulted in the lowest score.

#### 4.4.1 Non-Choice-Reversal Inconsistencies Under Non-Forced-Choice

An interesting and novel finding that is derived from our generalized HM analysis and which we report in Table 7 is that, despite making significantly fewer choice reversals (as we discussed in Section 2), subjects in the non-forced-choice treatment have significantly *higher* generalized HM scores than the standard HM scores of subjects in the forced-choice treatment, both when indifference is accounted for and when it is not. The reason for this is that, unlike the original HM score, the generalized one is strictly positive not only when the dataset in question is complete and features WARP violations, but also when it is incomplete and incompatible with any of the *Desirability*, *Contraction Consistency* and *Strong Expansion* axioms that were introduced in Section 4.1 and which characterize the rational indecisiveness model that the generalized HM index sets as its benchmark. Thus, complete datasets are associated with an upper-bound HM score that is significantly lower than the one corresponding to incomplete datasets, precisely because in the latter case there are three additional behavioral requirements that must be satisfied and which are independent of WARP. This disparity between the two upper bounds and the overall distributions of the generalized HM scores is indeed verified in the simulations of one million random-behaving subjects per treatment which we report in the Online Appendix 4.

Table 7: Average generalized HM scores in each treatment. Standard errors in parenthesis and  $p$ -values from Mann-Whitney  $U$  tests

	Generalized HM (w/ indifference)	Generalized HM (w/o indifference)
<b>NFC</b>	2.350	2.364
	(0.201)	(0.197)
$n =$	143	143
<b>FC</b>	0.706	0.8
	(0.131)	(0.129)
$n =$	75	75
$p$ -value	<0.001	<0.001

We now discuss some possible behavioral explanations that may lead subjects to violations of the above three WARP-independent conditions. As was mentioned earlier, the *Desirability* condition is necessarily violated when decision makers consider some alternatives to be unacceptable and hence non-choosable. Yet, as the model of unattractiveness-driven deferral which

is captured in (12) also shows, this behavior (which is also present in our data, as we discussed) may well be compatible with rationality. More of a puzzle are violations of *Contraction Consistency* and *Strong Expansion* as stated in (7) and (8), respectively. It is worth noting, however, that experimental evidence in existing studies shows that both these conditions can be expected to be violated in certain non-forced-choice settings.

Dhar and Simonson (2003) have shown, for example, that the well-known “decoy” or “attraction effect” is strengthened when choice is not forced, because even though most subjects choose the asymmetrically dominating option  $x$  in the relevant three-element menu  $\{x, x', y\}$  where  $x'$  is a decoy for  $x$ , many subjects actually defer at the binary menu  $\{x, y\}$  where the decoy  $x'$  has been removed. This is captured by  $C(\{x, x', y\}) = \{x\}$  and  $C(\{x, y\}) = \emptyset$ , which is a choice pattern that violates *Contraction Consistency* but not WARP. Although our experimental design features no objective dominance structures between alternatives and hence is not suitable for re-testing the Dhar-Simonson hypothesis, we note that the above choice pattern is nevertheless occasionally present in our datasets too.

A general behavioral reason why *Strong Expansion* may be violated on the other hand is the choice-overload effect that was discussed above. Decision makers who are overloaded in menus with three or more elements, for example, may exhibit the choice pattern  $C(\{x, y\}) = \{x\} = C(\{x, z\})$  and  $C(\{x, y, z\}) = \emptyset$ , which contradicts this axiom. Although the analysis that was provided in Section 2 suggests that there is no overload-like behavior at the aggregate level in our sample, we note that several subjects (including ones with zero WARP violations) exhibited this pattern sporadically.

#### 4.4.2 Subject Categorization

Table 8 presents our primary subject classification that uses revealed-preference data only and builds on the original as well as on the extended HM indices that were introduced above. This categorization includes three main types: unconstrained utility maximizers (UM), unattractiveness-constrained utility maximizers (UCUM) and either rationally indecisive or Bayesian decision makers.<sup>25</sup> As shown in the table, 43% and 30% of the subjects are best explained by the models of unconstrained and unattractiveness-constrained utility maximization, respectively. Therefore, the textbook model and its most straightforward deferral-permitting extension can account for nearly three quarters of the observed behavior in our NFC treatment. However, while an average of roughly 0.55 changes is required in the (complete) datasets of subjects in the first category for them to be perfectly compatible with utility maximization, an average of 2.3 changes are necessary for the incomplete datasets of subjects in the second category in order to be compatible

<sup>25</sup>In some cases, there is more than one preorder that minimizes a subject’s HM score. We apply the following tie-breaking rules: for the UM model, we select the preorder with the fewest indifferences; for UCUM model, we select the preorder with the fewest unattractive alternatives, then the preorder with the fewest indifferences; for the rationally indecisive/Bayesian model, we select the least incomplete preorder, then the preorder with the fewest indifferences. After applying this tie-breaking procedure, 126 subjects out of the 143 subjects in the NFC treatment are assigned a unique preorder while the remaining 17 subjects are assigned two preorders. Online Appendix 5 includes three examples from our subjects’ data that illustrate the results of this process.

with the corresponding model ( $p < 0.001$  with and without indifference). As Table 6 shows and the reader can easily verify, the model of unattractiveness-driven deferral also leads to behavior that conforms with the *Contraction Consistency* and *Strong Expansion* axioms. Therefore, the discussion of the previous subsection remains relevant here as a framework through which these higher HM scores can be explained.

Table 8: Percentage of subjects categorized as following each model and average model-based HM scores for each category. Subjects always fall in the same category regardless of whether indifference statements are included or not

	#	%	Model-HM (w/ indifference)	Model-HM (w/o indifference)
<b>Unconstrained Utility Maximizers (UM)</b>	61	42.7	0.541 (0.111)	0.590 (0.113)
<b>Unattractiveness-Constrained UM (UCUM)</b>	43	30.0	2.326 (0.308)	2.326 (0.303)
<b>Rationally Indecisive/Bayesian</b>	30	21.0	2.633 (0.443)	2.6 (0.433)
<b>Rationally Indecisive/Bayesian or UCUM</b>	6	4.2	3.667 (0.803)	3.667 (0.803)
<b>UM or UCUM</b>	2	1.4	3 (2)	3 (2)
<b>UM or Rationally Indecisive/Bayesian</b>	1	0.7	7 -	7 -
<i>n</i> =	143	100		

The third main category comprises 21% of the subjects for whom the best matches are the rational indecisiveness and Bayesian models which are confounded in the context of our experimental design. The average number of changes that are necessary for perfect conformity with the former model under some incomplete preference preorder, or with the latter model under some collection of menu-dependent beliefs, is 2.6 and is not significantly different than the average of 2.3 that corresponds to the model of unattractiveness-driven deferral ( $p = 0.807$  with indifference;  $p = 0.815$  without indifference). Yet, the difference between HM scores in the unconstrained utility maximization model and the one corresponding to this category is significant ( $p < 0.001$  with and without indifference). Finally, the remaining 6% of the subjects remained uncategorized by this method as there was more than one model that explained their behavior equally well.

With regards to subjects whose behavior was explained *perfectly* by the models in the above three categories, we note that there were 40 (28%) unconstrained and 10 (7%) unattractiveness-constrained utility maximizers of this kind, respectively, and 7 (5%) rationally indecisive or Bayesian subjects. Therefore, 40% of the NFC subjects' behavior is explained perfectly by the three observationally distinct models that we consider.

To further validate our subject classification into the various models, we make use of the two sets of questionnaire data collected in the experiment.<sup>26</sup> The first set is the subjects' responses to

<sup>26</sup>The use of non-choice data to validate the assignment of subjects to different models has been used in the literature previously.

the question that appeared following a deferral at a binary menu. Recall that the three options that subjects were given at this point were the following: i) “I could not decide which one I prefer”; ii) “I found both to be equally good”; iii) “other reason”. It is reasonable to hypothesize that subjects categorized as rationally indecisive/Bayesian would choose the first option more frequently than subjects categorized as unattractiveness-constrained utility maximizers. Similarly, subjects classified as the latter might be expected to state “other reason” following a deferral more frequently than subjects categorized as rationally indecisive/Bayesian. The data presented in Table 9 shows that our extended HM categorization method indeed supports this hypothesis. In particular, focusing on the sub-sample of subjects who deferred in at least one binary menu, those subjects who are categorized as rationally indecisive or Bayesian selected the first response 75% of the time, on average. The modal response for subjects in the unattractiveness-driven deferral category, on the other hand, was “other reason”. A Fisher exact test rejects the null of no difference between the frequency of deferral statements across the unattractiveness-constrained utility maximizers and rationally indecisive/Bayesian categories ( $p < 0.001$ ).<sup>27</sup>

Table 9: Frequency of statements following deferrals in binary menus for each subject category. (1) *indecisive*, (2) *indifferent* and (3) *other* measure the proportion of times a subject stated that they deferred because (1) “they could not decide which one they prefer”, (2) “they found both to be equally good”, or (3) “other reason”, respectively. Standard errors in parentheses.

	<i>indecisive</i>	<i>indifferent</i>	<i>other</i>	subjects	obs.
<b>Unattractiveness-Constrained UM</b>	0.361 (0.073)	0.179 (0.057)	0.460 (0.085)	33	107
<b>Rationally Indecisive/Bayesian</b>	0.757 (0.073)	0.148 (0.060)	0.095 (0.047)	27	130

While this data supports our HM-based categorization, it also suggests the possibility that the majority of subjects in the third category may in fact be indecisive. To test this further we counted the number of subjects in this category for whom the frequency of “indecisiveness” responses in the above question was strictly higher than 0.5. Nineteen out of the thirty subjects in this category satisfied this condition, which corresponds to an approximate two-thirds majority in favor of indecisiveness. The latter type of subjects in turn represents 13% of the total in the NFC treatment.

The second source of questionnaire data is the subjects’ psychological decisiveness scores. In this regard, the data reported in Table 10 provides further evidence in support of our HM-based categorization by showing that the subjects in the rationally indecisive/Bayesian category were indeed the least decisive on average ( $p = 0.064$ , Mann-Whitney  $U$ -test). Interestingly, if we focus on the 19 subjects in this category who were declared indecisive according to the above

See for example Costa-Gomes, Broseta, and Crawford (2001) who use subjects’ information searches of elements of a game to validate the assignments to different models using subjects’ actions, as well as to improve the assignment when two models predict the same actions in all of a set of games.

<sup>27</sup>Subjects categorized as following the unconstrained utility maximization model or falling into more than one category are not considered in this analysis due to the very low number of binary menu deferral observations.

Table 10: Mean psychological decisiveness scores for each category. Standard errors in parentheses. Subjects categorized as Rationally Indecisive/Bayesian are further subdivided into either Indecisive or Bayesian according to their binary menu statements following deferral as detailed in the text.

UM	Rationally Indecisive/Bayesian			UCUM
	All	Indecisive	Bayesian	
0.677	0.596	0.561	0.656	0.681
(0.024)	(0.041)	(0.053)	(0.064)	(0.029)
$n = 61$	$n = 30$	$n = 19$	$n = 11$	$n = 43$

procedure (i.e. those who stated “I could not decide which one I prefer” more than half of the time as their reason for deferring), then the average psychological decisiveness score decreases and the difference between this value and the corresponding value for the two categories of utility maximization becomes significant at the 5% level ( $p = 0.026$ ).

## 5 Related Literature

Our paper appears to be the first in the literature to test for a negative effect of forced choice on the subjects’ consistency. However, studies that have used non-experimental data on health insurance, pension savings and voting decisions and which have shown, among other things, a tendency of decision makers to choose their non-market outside option/default alternative (i.e. what we have been referring to as “choice deferral” in this paper) include Samuelson and Zeckhauser (1988), Madrian and Shea (2001), Iyengar et al (2004), Iyengar and Kamenica (2010), Carroll et al (2009) and Augenblick and Nicholson (2016). In some of these studies this behavior is traceable in 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 we argued above, is not relevant for our data.

As far as other experimental-economics approaches to incomplete-preference elicitation are concerned, the designs proposed in Danan and Ziegelmeyer (2006) and Cettolin and Riedl (2015) aimed to test for the presence of such preferences in the domain of money lotteries and/or ambiguous acts, and without allowing subjects to defer as we did. Danan and Ziegelmeyer (2006) found that a high proportion of the subjects in their two-stage experiment initially chose a menu that included a risky lottery and a certain cash amount over the singleton menus that comprised either the lottery or the certain monetary amount, respectively. They interpreted these first-stage preferences as evidence for incompleteness. The structure of their design and the reported findings, however, are also compatible with a positive test for “preference for flexibility” in the sense of Kreps (1979).

The approach followed by Cettolin and Riedl (2015) builds on a design that features uncertain options with one positive and one zero monetary outcome, and gives participants the possibility to delegate to a uniform randomization device the choice within each pair of uncertain options that were sequentially presented to them by stating that they were “indifferent” between these

---

options. Making use of relevant technical properties of expected utility functionals and some of their extensions, the authors concluded that subjects who conform with these models must choose the “indifference” option exactly once in the particular sequence of pairs that they saw, while they also argued that the higher frequency of indifference statements that they actually found in their data is compatible with the Bewley (2002) model of indecisiveness in *beliefs*. Our design clearly differs in many respects from these authors’, and, most importantly, contributes to the elicitation of incompleteness/indecisiveness in *tastes* over riskless and non-monetary alternatives.

## 6 Concluding Remarks

The first conclusion that can be derived from our analysis is that forcing decision makers to always make active choices can have a detrimental effect on their consistency when the latter is captured by the number or relative frequency of choice reversals/WARP violations. In light of this finding, the analysis of 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 in a suitably incentivized way, then some of the inconsistency that would have been observed if they were forced to choose and which may be driven by unattractiveness or indecisiveness constraints may in fact disappear, and a stable preference ordering may be elicited. At the same time, our analysis also suggests that giving subjects the opportunity to defer making an active choice can and does make them prone to decision inconsistencies that are independent of WARP. Unlike choice reversals that in theory render a decision maker vulnerable to exploitations of a money-pump type, however, such exploitation is not possible for this other type of inconsistencies which are manifested in cases where the agent defers even though his general behavior would suggest that he should have made an active choice.

The effect of forced choice on choice reversals that we report here has implications for empirical/experimental work on individual decision making. At one level, it suggests that violations of WARP and GARP which have been documented in existing studies<sup>28</sup> where subjects were forced to choose from competitive budget sets may decrease or even disappear once subjects are allowed to defer choice. At another level, it is potentially relevant to many recently developed models of bounded-rational decision making where the agent is portrayed as having, for example, cyclic preferences or as suffering from limited attention. Most of these models build on the assumption that active choices are always made and that choice observations from all possible decision problems/menus are available to the analyst.<sup>29</sup> Our study suggests that a robustness check for future experimental studies that aim to test these theories is the inclusion of a non-forced-choice treatment like the one we are proposing in this paper, which would allow for ruling out alterna-

---

<sup>28</sup>See, for example, Sippel (1997), Harbaugh, Krause, and Berry (2001) or Andreoni and Miller (2002).

<sup>29</sup>de Clippel and Rozen (2014) re-examines some of these models when this assumption is relaxed.

---

tive explanations for the subjects' underlying decision process and/or preferences over the given set of choice options.

Our novel model-proximity/goodness-of-fit analysis that builds on combinatorial-optimization methods provides strong support to the textbook model of utility maximization and its straightforward unattractiveness-constrained deferral-permitting variant, which jointly account for the behavior of almost three quarters of our non-forced-choice datasets. This analysis also suggests that about one fifth of the subjects in this treatment exhibit behavior that is best matched by the models of rational indecisiveness and Bayesian-rational information acquisition that are confounded in the context of our experimental design. Our additional data that utilizes questionnaire responses following subjects' deferrals in binary menus together with responses in an indecisiveness personality questionnaire suggest that the majority of the subjects in this category deferred because they were indecisive in the sense that they were unable to compare the relevant alternatives. Further support for the rational indecisiveness model finally comes from noticing that if the Bayesian model provided a good explanation for the behavior of what we referred to in Section 3 as "psychologically indecisive" subjects, then WARP violations for them should be similar in the forced-choice and non-forced-choice treatments. However, we observe a fivefold increase in the average number of violations for such subjects in the former treatment compared to the latter.

We conclude with a remark on experimental methods in deferral-permitting environments. The novel protocol that we implemented in order to collect choice data with deferral observations offered subjects the opportunity to obtain additional information about the alternatives in question. As we noted in the introduction, this is in line with natural existing ideas that have been discussed by economists and psychologists and which suggest that indecisive people may welcome such information as a means for resolving their so-called *tentative* indecisiveness. A simple variation of this protocol that could be employed in future studies and which may provide a means towards eliciting subjects' *tentative or assertive* indecisiveness by means of revealed-preference data alone would deprive them of this opportunity and thus break the confounding link between rational indecisiveness and Bayesian rationality.

## References

- ANDERSON, C. J. (2003): "The Psychology of Doing Nothing: Forms of Decision Avoidance Result from Reason and Emotion," *Psychological Bulletin*, 129, 139–167.
- ANDREONI, J., AND J. MILLER (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70, 737–753.
- APESTEGUIA, J., AND M. BALLESTER (2015): "A Measure of Rationality and Welfare," *Journal of Political Economy*, pp. 1278–1310.

- 
- AUGENBLICK, N., AND S. NICHOLSON (2016): “Ballot Position, Choice Fatigue and Voter Behaviour,” *Review of Economic Studies*, 83, 460–480.
- AUMANN, R. J. (1962): “Utility Theory without the Completeness Axiom,” *Econometrica*, 30, 445–462.
- BERNHEIM, B. D., AND A. RANGEL (2009): “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *Quarterly Journal of Economics*, 124, 51–104.
- BEWLEY, T. F. (2002): “Knightian Decision Theory. Part I,” *Decisions in Economics and Finance*, 25, 79–110.
- BHATIA, S., AND T. L. MULLETT (2016): “The Dynamics of Deferred Decision,” *Cognitive Psychology*, 86, 112–151.
- CARROLL, G. D., J. J. CHOI, D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2009): “Optimal Defaults and Active Decisions,” *Quarterly Journal of Economics*, 124, 1639–1674.
- CETTOLIN, E., AND A. RIEDL (2015): “Revealed Incomplete Preferences under Uncertainty,” *mimeo*.
- CHERNEV, A., U. BÖCKENHOLT, AND J. GOODMAN (2015): “Choice Overload: A Conceptual Review and Meta-Analysis,” *Journal of Consumer Psychology*, 25, 333–358.
- CHOI, S., R. FISMAN, D. GALE, AND S. KARIV (2007): “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *American Economic Review*, 97, 1921–1938.
- CHOI, S., S. KARIV, W. MÜLLER, AND D. SILVERMAN (2014): “Who is (More) Rational?,” *American Economic Review*, 104, 1518–1550.
- COSTA-GOMES, M., B. BROSETA, AND V. P. CRAWFORD (2001): “Cognition and Behavior in Normal-Form Games: An Experimental Study,” *Econometrica*, 69, 1193–1235.
- DANAN, E., AND A. ZIEGELMEYER (2006): “Are Preferences Complete? An Experimental Measurement of Indecisiveness Under Risk,” *mimeo*.
- DE CLIPPEL, G., AND K. ROZEN (2014): “Bounded Rationality and Limited Datasets,” *mimeo*.
- DEAN, M., AND D. MARTIN (2016): “Measuring Rationality with the Minimum Cost of Revealed Preference Violations,” *Review of Economics and Statistics*, 98, 524–534.
- DHAR, R. (1997): “Consumer Preference for a No-Choice Option,” *Journal of Consumer Research*, 24, 215–231.
- DHAR, R., AND I. SIMONSON (2003): “The Effect of Forced Choice on Choice,” *Journal of Marketing Research*, 40, 146–160.

- 
- ECHENIQUE, F., T. IMAI, AND K. SAITO (2016): “Testable Implications of Models of Intertemporal Choice: Exponential Discounting and Its Generalizations,” *mimeo*.
- ECHENIQUE, F., S. LEE, AND M. SHUM (2011): “The Money Pump as a Measure of Revealed Preference Violations,” *Journal of Political Economy*, 119, 1201–1223.
- FAMULARI, M. (1995): “A Household-Based, Nonparametric Test of Demand Theory,” *Review of Economics and Statistics*, 77, 372–383.
- FERRARI, J. R., AND J. F. DOVIDIO (2001): “Behavioral Information Search by Indecisives,” *Personality and Individual Differences*, 30, 1113–1123.
- FISCHBACHER, U. (2007): “z-Tree: Zurich Toolbox for Ready-Made Economic Experiments,” *Experimental Economics*, 10, 171–178.
- FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): “Individual Preferences for Giving,” *American Economic Review*, 97, 1858–1876.
- GERASIMOU, G. (2015): “Indecisiveness, Undesirability and Overload Revealed Through Rational Choice Deferral,” *mimeo*.
- GERMEIJS, V., AND P. D. BOECK (2002): “A Measurement Scale for Indecisiveness and its Relationship to Career Indecision and Other Types of Indecision,” *European Journal of Personality Assessment*, 18, 113–122.
- GILBOA, I., F. MACCHERONI, M. MARINACCI, AND D. SCHMEIDLER (2010): “Objective and Subjective Rationality in a Multiple Prior Model,” *Econometrica*, 78, 755–770.
- GROSS, J. (1995): “Testing Data for Consistency with Revealed Preference,” *Review of Economics and Statistics*, 77, 701–710.
- HALEVY, Y., D. PERSITZ, AND L. ZRILL (2012): “Parametric Recoverability of Preferences,” *mimeo*.
- HARBAUGH, W. T., K. KRAUSE, AND T. R. BERRY (2001): “GARP for Kids: On the Development of Rational Choice Behavior,” *American Economic Review*, 91, 1539–1545.
- HEUFER, J., AND P. HJERTSTRAND (2015): “Consistent Subsets: Computationally Feasible Methods to Compute the Houtman-Maks-Index,” *Economics Letters*, 128(87-89).
- HOUTMAN, M., AND J. A. MAKES (1985): “Determining All Maximal Data Subsets Consistent with Revealed Preference,” *Kwantitatieve Methoden*, 19, 89–104.
- IYENGAR, S. S., G. HUBERMAN, AND W. JIANG (2004): “How Much Choice is Too Much? Contributions to 401(k) Retirement Plans,” in *Pension Design and Structure: New Lessons from Behavioral Finance*, ed. by O. S. Mitchell, and S. P. Utkus, pp. 83–96. New York: Oxford University Press.

- 
- IYENGAR, S. S., AND E. KAMENICA (2010): “Choice Proliferation, Simplicity Seeking, and Asset Allocation,” *Journal of Public Economics*, 94, 530–539.
- IYENGAR, S. S., AND M. R. LEPPER (2000): “When Choice is Demotivating: Can One Desire Too Much of a Good Thing?,” *Journal of Personality and Social Psychology*, 79, 995–1006.
- KREPS, D. M. (1979): “A Representation Theorem for ‘Preference for Flexibility’,” *Econometrica*, 47, 565–577.
- LEVI, I. (1986): *Hard Choices: Decision Making under Unresolved Conflict*. Cambridge: Cambridge University Press.
- MADRIAN, B. C., AND D. F. SHEA (2001): “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics*, 116, 1149–1187.
- MOCHON, D. (2013): “Single-Option Aversion,” *Journal of Consumer Research*, 40, 555–566.
- RASSIN, E., P. MURIS, E. BOOSTER, AND I. KOLSLOOT (2008): “Indecisiveness and Informational Tunnel Vision,” *Personality and Individual Differences*, 45, 96–102.
- SALANT, Y., AND A. RUBINSTEIN (2008): “ $(A, f)$ : Choice with Frames,” *Review of Economic Studies*, 75, 1287–1296.
- SAMUELSON, W., AND R. ZECKHAUSER (1988): “Status Quo Bias in Decision Making,” *Journal of Risk and Uncertainty*, 1, 7–59.
- SAVAGE, L. J. (1954): *The Foundations of Statistics*. New York: Wiley.
- SEN, A. (1971): “Choice Functions and Revealed Preference,” *Review of Economic Studies*, 38, 307–317.
- (1997): “Maximization and the Act of Choice,” *Econometrica*, 65, 745–779.
- SHAFIR, E., I. SIMONSON, AND A. TVERSKY (1993): “Reason-Based Choice,” *Cognition*, 11, 11–36.
- SIPPEL, R. (1997): “An Experiment on the Pure Theory of Consumer’s Behaviour,” *Economic Journal*, 107, 1431–1444.
- SMEULDERS, B., F. C. R. SPIEKSMAN, L. CHERCHYE, AND B. DE ROCK (2014): “Goodness-of-Fit Measures for Revealed Preference Tests: Complexity Results and Algorithms,” *ACM Transactions on Economics and Computation*, 2, 1–16.
- TVERSKY, A., AND E. SHAFIR (1992): “Choice under Conflict: The Dynamics of Deferred Decision,” *Psychological Science*, 3, 358–361.
- VON NEUMANN, J., AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

---

ZAKAY, D. (1984): ““To Choose or not to Choose”: On Choice Strategy in Face of a Single Alternative,” *American Journal of Psychology*, 97, 373–389.

## Online Appendix 1: Instructions in the Non-Forced Choice Treatment

### 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.

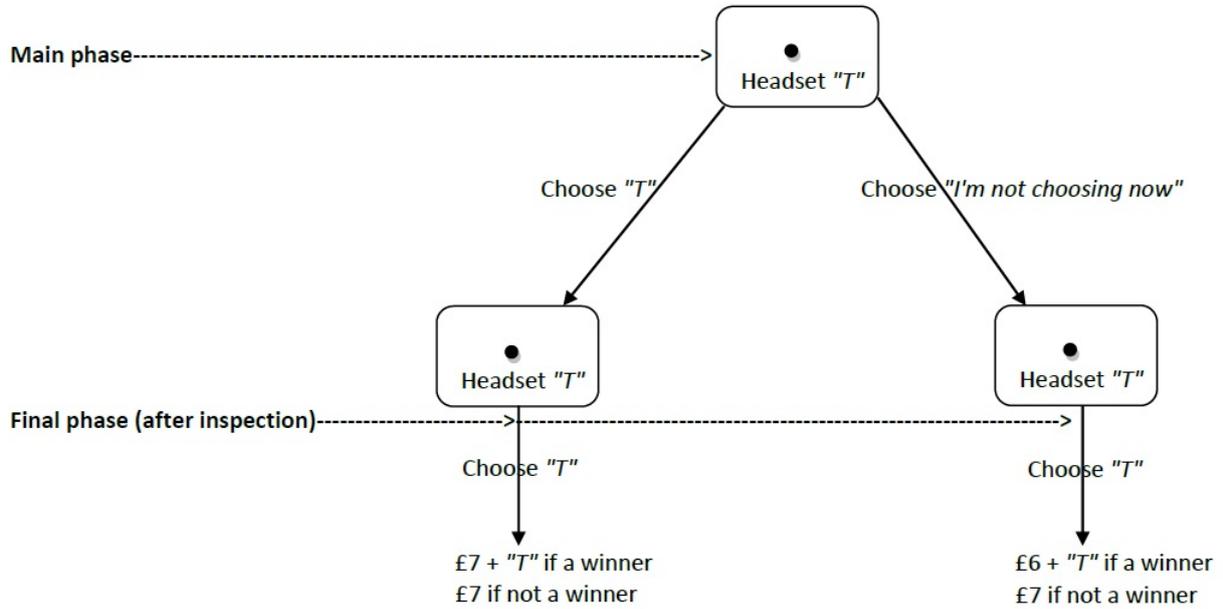
*If your randomly selected menu R contains two HSs and you had previously stated that both were equally good, then:*

1) *If you had chosen a headset from R initially, you will not be able to change your decision at this stage.*

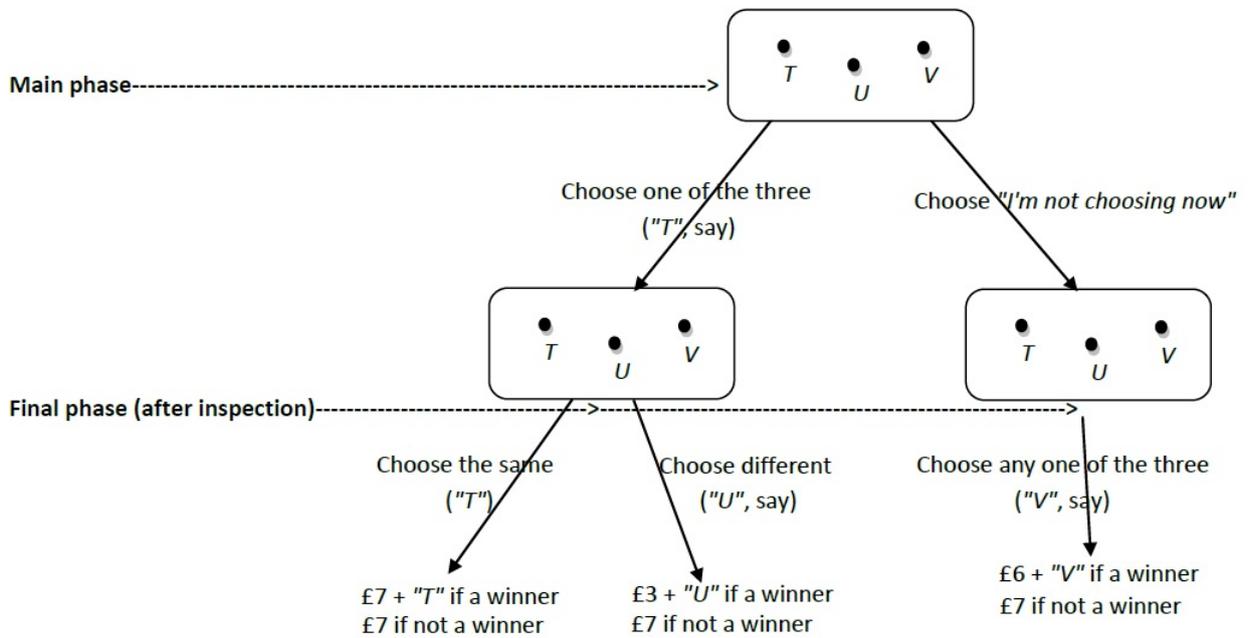
One of the two HS will be randomly selected and you will win this HS if you are picked as a winner.

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.

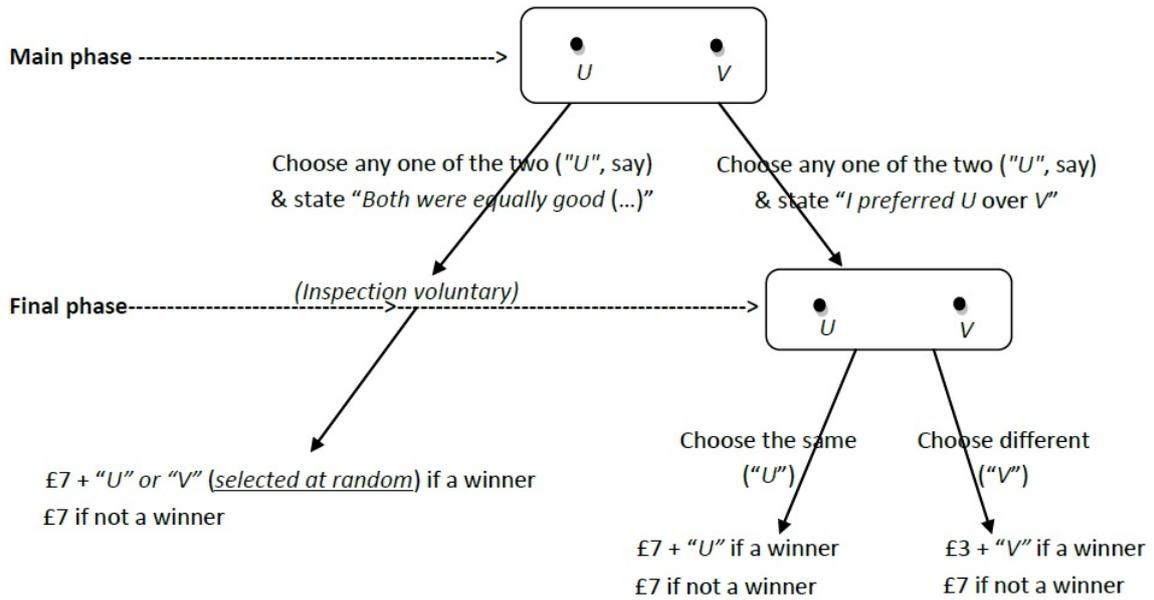
**Example 1:**  
Randomly selected menu with only one headset



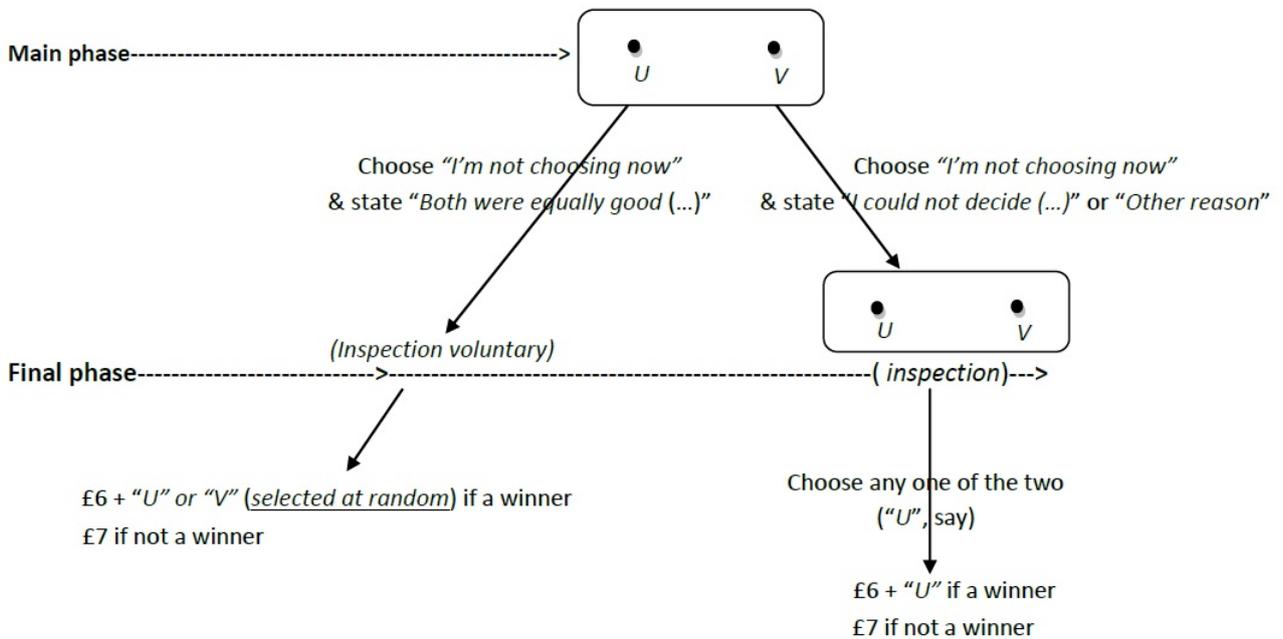
**Example 2:**  
Randomly selected menu with three headsets



**Example 3:**  
Two-headset randomly selected menu where a choice had been made initially



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



---

## Online Appendix 2: Instructions in the Forced Choice Treatment

### 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:

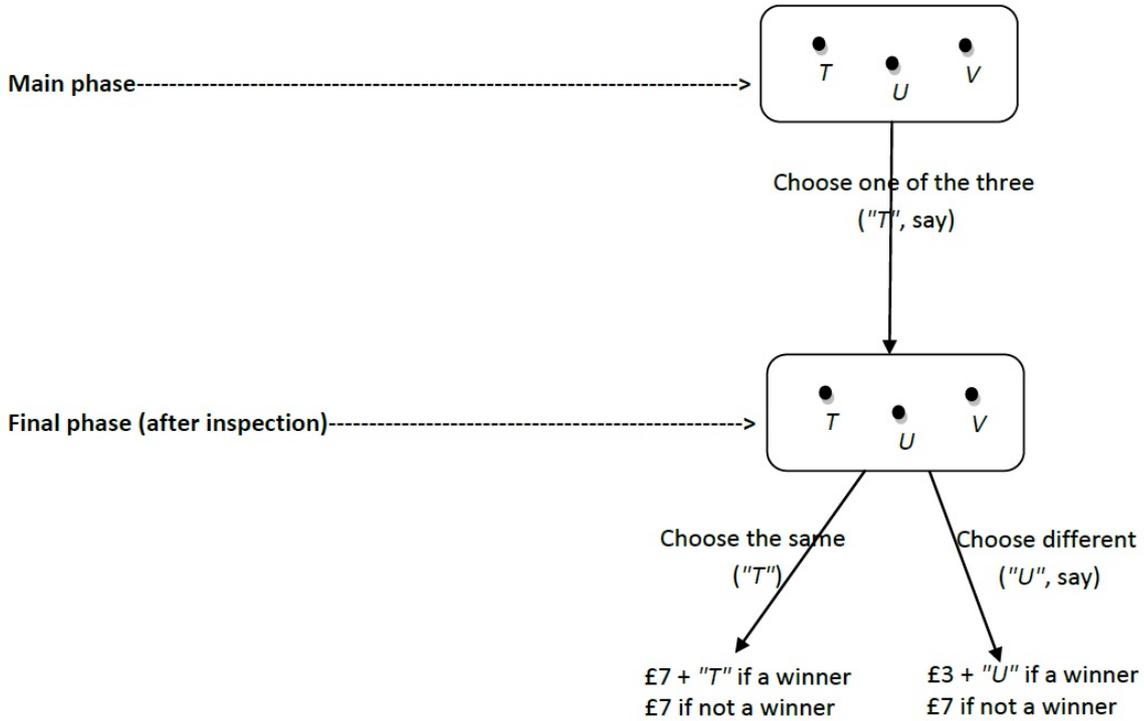
- A) Suppose that when you first saw menu *R* *you had chosen some HS* from it. *If you chose the same menu the second time*, you will receive the £7 initially allocated to you.
- B) Suppose that when you first saw menu *R* *you had chosen some HS* from it. *If you chose a different one the second time,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

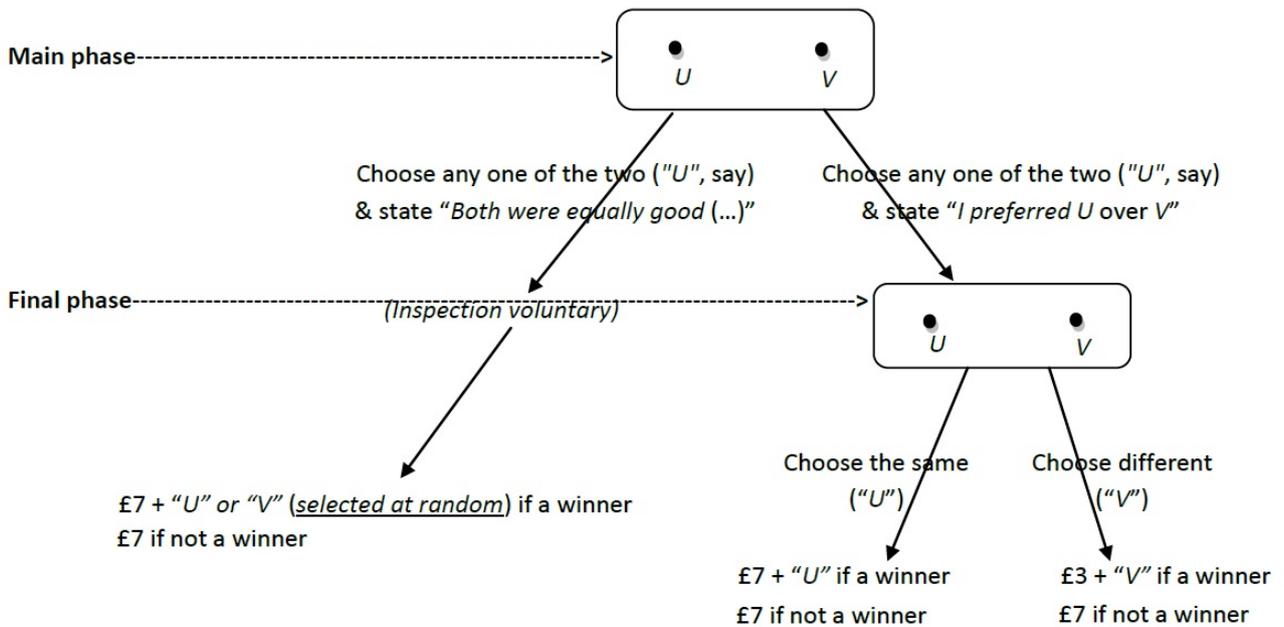
During the phase when you are presented with the 31 menus, whenever *a menu of exactly two HSs* comes up and you have chosen one of them, a short follow-up question will ask you to state if you *preferred* the chosen HS over the non-chosen one, or if the non-chosen one was *equally good* to the one you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

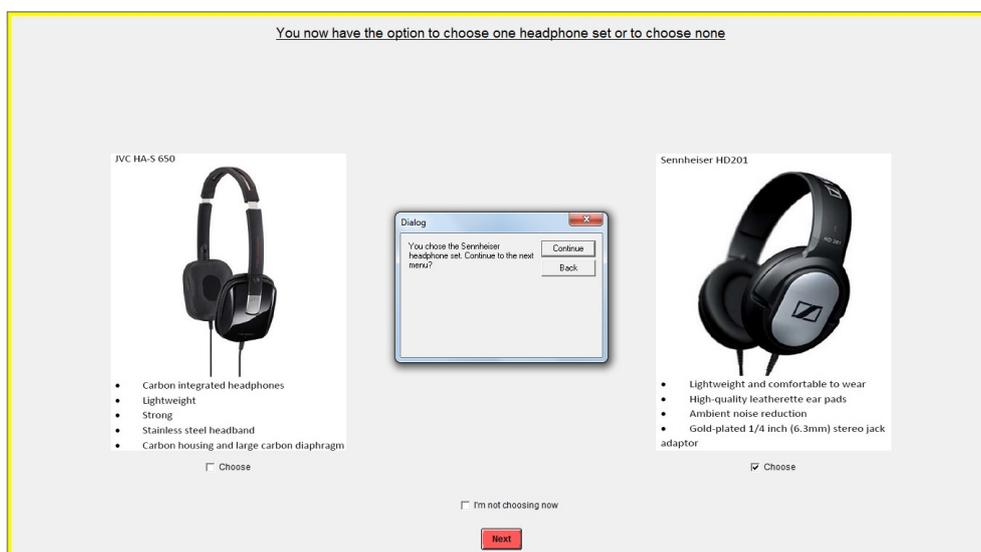
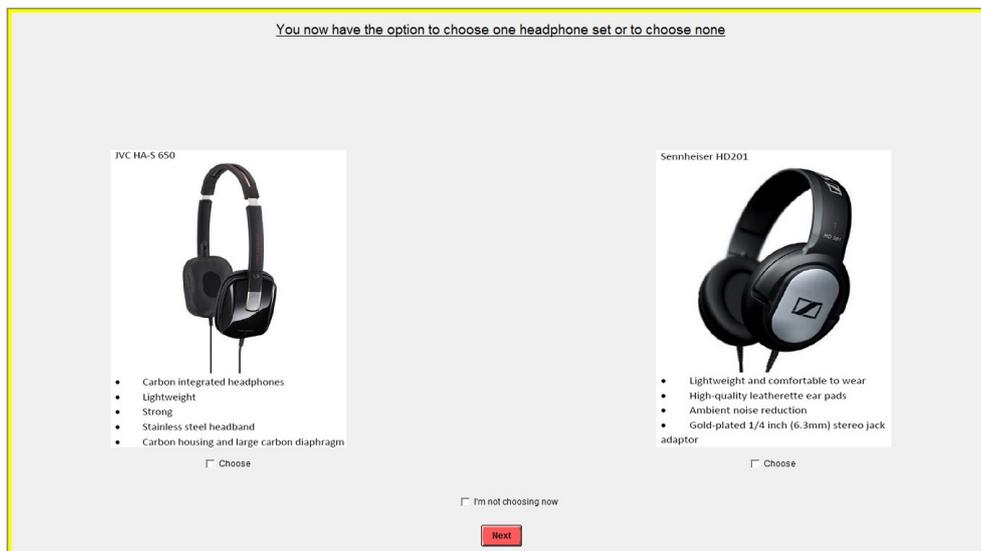
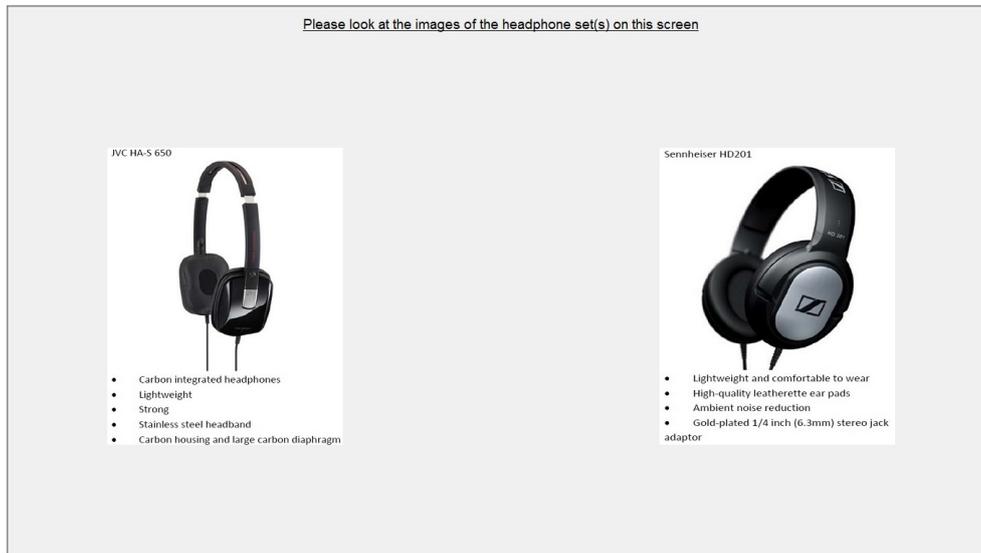
**Example 1:  
Randomly selected menu with three headsets**



**Example 2:  
Randomly selected menu with two headsets**



## Online Appendix 3: Sample Screenshots



Why did you choose this headphone set?  I chose this headphone set because I preferred it to the other headphone set in this menu  
 I found both equally good, so I chose randomly  
 Other reason

OK

You did not choose any headphone set from this menu. Which of these options best reflects your reasons for not choosing?  I could not decide which one I prefer  
 I found both equally good  
 Other reason

OK

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

Menu Selected

OK

## Online Appendix 4: Augmentation of Choice Data with Indifference Statements

The indifference-augmentation process comprises two parts: I) consistency check of indifference statements (which may remove some or all of them); II) the augmentation itself.

### I. Consistency check

The input to 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 correspondence  $C$  giving the choosable set  $C(M)$  of the subject for each menu  $M \in \mathcal{M}$ .

First, we determine the set  $\mathcal{D}$  of menus where the subject has deferred while stating indifference. As explained in Section 3.2 in the main text, we will disregard these statements.

$$\mathcal{D} := \{M \mid M \in \mathcal{I}, C(M) = \emptyset\} \quad (15)$$

**Preference multigraph** Then we construct a preference multigraph  $G = (V, E)$  with labelled edges, where each vertex  $v \in V$  corresponds to an option available in some menu. Then we traverse all binary menus  $M = \{u, v\}$  and for each one, we create edges as follows.

- If  $M \in \mathcal{I}$ , and  $C(M) = \emptyset$ , we do not create any edges. (This is a case where  $M \in \mathcal{D}$ .)
- If  $M \in \mathcal{I}$ , and  $C(M) \neq \emptyset$ , we create edges  $u \xrightarrow[\sim]{\{M\}} v$  and  $v \xrightarrow[\sim]{\{M\}} u$ .
- If  $M \notin \mathcal{I}$ , and  $C(M) = \{u\}$ , we create edges  $u \xrightarrow[\succ]{\{M\}} v$  and  $v \xrightarrow[\prec]{\{M\}} u$ .
- If  $M \notin \mathcal{I}$ , and  $C(M) = \{v\}$ , we create edges  $u \xrightarrow[\prec]{\{M\}} v$  and  $v \xrightarrow[\succ]{\{M\}} u$ .

Each edge is annotated with a list of menus involved (above the arrow), and with a comparison relation induced by the choices (below the arrow).

**Transitive closure** Next, we calculate the transitive closure  $\rightarrow^*$  of the edge relation  $\rightarrow$ , merging the menu sets above the arrows.

$$u \xrightarrow[\underset{R}{\leftarrow}]{\overset{M}{\rightarrow}} v \iff u \xrightarrow[\underset{R}{\leftarrow}]{\overset{M}{\rightarrow}} v \quad (16)$$

$$u \xrightarrow[\underset{R}{\leftarrow}]{\overset{M \cup N}{\rightarrow}} v \iff \left( u \xrightarrow[\underset{S}{\leftarrow}]{\overset{M}{\rightarrow}} w \right) \wedge \left( w \xrightarrow[\underset{T}{\leftarrow}]{\overset{N}{\rightarrow}} v \right) \wedge (R, S, T) \in \mathcal{T} \quad (17)$$

The set  $\mathcal{T}$  of triples of comparison relations  $(R, S, T)$  describes how comparisons  $S$  and  $T$  compose to imply comparison  $R$ .

$$\mathcal{T} := \{(\prec, \prec, \prec), (\prec, \prec, \sim), (\prec, \sim, \prec), (\sim, \sim, \sim), (\succ, \succ, \succ), (\succ, \succ, \sim), (\succ, \sim, \succ)\} \quad (18)$$

**Removal of conflicting indifference statements** The final step of the consistency check is based on the fact that (transitive) comparisons we have just built may be in conflict with stated comparisons – or, rather, their absence because the subject chose differently or not at all.

$$\mathcal{C} := \bigcup \left\{ M \mid u \xrightarrow[\prec]{M} v, \left( u \xrightarrow[\prec]{\{\{u,v\}\}} v \right) \notin E \right\} \cup \bigcup \left\{ M \mid u \xrightarrow[\sim]{M} v, \left( u \xrightarrow[\sim]{\{\{u,v\}\}} v \right) \notin E \right\} \quad (19)$$

The set of conflicting menus  $\mathcal{C}$  collects all binary menus that are involved in intransitivities – either by a conflict of a transitively required strict preference (because the subject deferred or compared the options differently), or by a conflict of a transitively required indifference.

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 saying which indifference statements we should preferentially keep.

Therefore, the updated set of indifference-inclusive binary menus is:

$$\mathcal{I}' := \mathcal{I} \setminus (\mathcal{D} \cup \mathcal{C}) \quad (20)$$

**Repeat** By removing some indifference statements, we may however create new conflicts because some indifferences become strict preference comparisons. Therefore, we repeat the above process of consistency check and indifference statement removal until  $\mathcal{I}' = \mathcal{I}$ .

## II. Augmentation

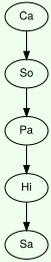
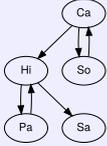
Given a set of indifference-inclusive binary menus  $\mathcal{I}$  that survives the above process, for each menu  $M$  and the corresponding choice set  $C(M)$  of a subject, we calculate the augmented choice set  $C'(M)$  as the union of all indifference-inclusive binary menus having common elements with  $C(M)$ , intersected with  $M$ :

$$C'(M) := M \cap \bigcup \{K \mid K \in \mathcal{I}, K \cup C(M) \neq \emptyset\} \quad (21)$$

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

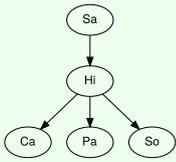
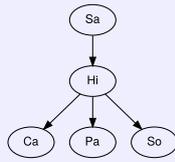
# Online Appendix 5: HM-Based Model-Proximity Examples

Note: Arrows point from less to more preferred options

				Raw	Augmented
		Best model		Unconstrained utility maximization	Unconstrained utility maximization
		Switches necessary		2	1
		Pairs of indifferent options		0	2
		Visualisation			
Menu	Subject raw choices	Subject augmented choices	Best order 32847951	Best order 32979039	
CaHiPaSaSo	Sa	Sa	Sa	Sa	
CaHiPaSa	Sa	Sa	Sa	Sa	
CaHiPaSo	Hi	HiPa	Hi	HiPa	
CaHiSaSo	Sa	Sa	Sa	Sa	
CaPaSaSo	Sa	Sa	Sa	Sa	
HiPaSaSo	Sa	Sa	Sa	Sa	
CaHiPa	Hi	HiPa	Hi	HiPa	
CaHiSa	Sa	Sa	Sa	Sa	
CaHiSo	Hi	Hi	Hi	Hi	
CaPaSa	Sa	Sa	Sa	Sa	
CaPaSo	Pa	Pa	Pa	Pa	
CaSaSo	Sa	Sa	Sa	Sa	
HiPaSa	Pa	HiPa	<b>Sa</b>	<b>Sa</b>	
HiPaSo	Pa	HiPa	<b>Hi</b>	HiPa	
HiSaSo	Sa	Sa	Sa	Sa	
PaSaSo	Sa	Sa	Sa	Sa	
CaHi	Hi	Hi	Hi	Hi	
CaPa	Pa	Pa	Pa	Pa	
CaSa	Sa	Sa	Sa	Sa	
CaSo	So (indiff.)	CaSo	So	CaSo	
HiPa	Hi (indiff.)	HiPa	Hi	HiPa	
HiSa	Sa	Sa	Sa	Sa	
HiSo	Hi	Hi	Hi	Hi	
PaSa	Sa	Sa	Sa	Sa	
PaSo	Pa	Pa	Pa	Pa	
SaSo	Sa	Sa	Sa	Sa	
Ca	Ca	Ca	Ca	Ca	
Hi	Hi	Hi	Hi	Hi	
Pa	Pa	Pa	Pa	Pa	
Sa	Sa	Sa	Sa	Sa	
So	So	So	So	So	

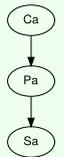
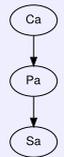
nfc-140129\_1407-phase3-subj11

(Arrows point from less desirable options to more desirable options.)

			Raw	Augmented
Best model			Rational indecisiveness	Rational indecisiveness
Switches necessary			0	0
Pairs of incomparable options			3	3
Pairs of indifferent options			0	0
Visualisation				
Menu	Subject raw choices	Subject augmented choices	Best order 17732577	Best order 17732577
CaHiPaSaSo	--	--	--	--
CaHiPaSa	--	--	--	--
CaHiPaSo	--	--	--	--
CaHiSaSo	--	--	--	--
CaPaSaSo	--	--	--	--
HiPaSaSo	--	--	--	--
CaHiPa	--	--	--	--
CaHiSa	Ca	Ca	Ca	Ca
CaHiSo	--	--	--	--
CaPaSa	--	--	--	--
CaPaSo	--	--	--	--
CaSaSo	--	--	--	--
HiPaSa	Pa	Pa	Pa	Pa
HiPaSo	--	--	--	--
HiSaSo	So	So	So	So
PaSaSo	--	--	--	--
CaHi	Ca	Ca	Ca	Ca
CaPa	--	--	--	--
CaSa	Ca	Ca	Ca	Ca
CaSo	--	--	--	--
HiPa	Pa	Pa	Pa	Pa
HiSa	Hi	Hi	Hi	Hi
HiSo	So	So	So	So
PaSa	Pa	Pa	Pa	Pa
PaSo	--	--	--	--
SaSo	So	So	So	So
Ca	Ca	Ca	Ca	Ca
Hi	Hi	Hi	Hi	Hi
Pa	Pa	Pa	Pa	Pa
Sa	Sa	Sa	Sa	Sa
So	So	So	So	So

nfc-131122\_1543-phase3-subj05

(Arrows point from less desirable options to more desirable options.)

			Raw	Augmented
		Best model	Unattractiveness-constrained utility maximization	Unattractiveness-constrained utility maximization
		Switches necessary	0	0
		Unattractive options	2	2
		Pairs of indifferent options	0	0
		Visualisation		
Menu	Subject raw choices	Subject augmented choices	Best order 24090711-HiSo	Best order 24090711-HiSo
CaHiPaSaSo	Sa	Sa	Sa	Sa
CaHiPaSa	Sa	Sa	Sa	Sa
CaHiPaSo	Pa	Pa	Pa	Pa
CaHiSaSo	Sa	Sa	Sa	Sa
CaPaSaSo	Sa	Sa	Sa	Sa
HiPaSaSo	Sa	Sa	Sa	Sa
CaHiPa	Pa	Pa	Pa	Pa
CaHiSa	Sa	Sa	Sa	Sa
CaHiSo	Ca	Ca	Ca	Ca
CaPaSa	Sa	Sa	Sa	Sa
CaPaSo	Pa	Pa	Pa	Pa
CaSaSo	Sa	Sa	Sa	Sa
HiPaSa	Sa	Sa	Sa	Sa
HiPaSo	Pa	Pa	Pa	Pa
HiSaSo	Sa	Sa	Sa	Sa
PaSaSo	Sa	Sa	Sa	Sa
CaHi	Ca	Ca	Ca	Ca
CaPa	Pa	Pa	Pa	Pa
CaSa	Sa	Sa	Sa	Sa
CaSo	Ca	Ca	Ca	Ca
HiPa	Pa	Pa	Pa	Pa
HiSa	Sa	Sa	Sa	Sa
HiSo	--	--	--	--
PaSa	Sa	Sa	Sa	Sa
PaSo	Pa	Pa	Pa	Pa
SaSo	Sa	Sa	Sa	Sa
Ca	Ca	Ca	Ca	Ca
Hi	--	--	--	--
Pa	Pa	Pa	Pa	Pa
Sa	Sa	Sa	Sa	Sa
So	--	--	--	--

---

## Online Appendix 6: Simulated Random Subjects

The simulations were done in the following way (the program that generates random subjects was written in the Rust programming language and the statistical analysis was carried out in the R programming language for statistical computing - further details are available from the authors on request):

To generate a random subject, iterate over all 31 menus. For each menu, with  $N$  being the size of that menu, do the following:

- If the menu is non-binary ( $N \neq 2$ ):
  - Generate a random number from
    - \* the range  $1, \dots, N$  (inclusive) for forced choice
    - \* the range  $1 \dots N + 1$  (inclusive) for non-forced choice
  - Outcomes in  $1 \dots N$  correspond to options of the menu
  - Outcome  $N + 1$  corresponds to deferral
- If the menu is binary ( $N = 2$ ):
  - Pick a random element from the set
    - \*  $\{1,2,INDIFF\}$  for forced choice
    - \*  $\{1,2,INDIFF,DEFER\}$  for non-forced choice
  - If the outcome is 1 or 2, pick the corresponding option from the menu
  - If the outcome is DEFER, defer
  - If the outcome is INDIFF, generate a number from
    - \* the set  $\{1,2\}$  for forced choice
    - \* the set  $\{1,2,DEFER\}$  for non-forced choice
    - \* If the outcome is 1 or 2, pick the corresponding option from the menu, while declaring indifference
    - \* If the outcome is DEFER, defer, while declaring indifference.

Throughout this process all random choices are done uniformly.

The distributions of HM violations for all simulated FC and NFC subjects when indifference statements are also included (and validated in the way explained in the text) are shown in Figure 4 for the original (“HM.rationality”), generalized (“HM.generalised), rational indecisiveness (“HM.indecisiveness”) and unattractiveness-constrained (“HM.unattractiveness”) HM index, together with the distributions of WARP violations (“warp.pairs”).

Figure 4: The distributions of generalized/extended HM scores and WARP violations for two million simulated random-behaving subjects (one million per treatment)

