School of Economics and Finance Discussion Papers

University of St Andrews

Choice, Deferral and Consistency

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

School of Economics and Finance Discussion Paper No. 1416
12 Dec 2014 (revised 18 Dec 2020)
JEL Classification: C91, D01, D03, D11, D12,
Keywords: Choice deferral; active choices; choice consistency; revealed preferences; decision difficulty; experiments.

School of Economics and Finance
Online Discussion Paper Series
issn 2055-303X
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

Choice, Deferral and Consistency*†

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

December 18, 2020

Abstract

We report on two novel choice experiments with real goods where subjects in one treatment are forced to choose, as is the norm in economic experiments, while in the other they are not but can instead incur a small cost to defer choice. Using a variety of measures we find that the active choices of subjects in the latter treatment are generally more consistent. Our results suggest that non-forced-choice experiments can be helpful in separating subjects' rational, hesitant and genuinely inconsistent behavior, and can potentially offer important new insights in revealed preference analysis.

Keywords: Choice deferral; active choices; choice consistency; revealed preferences; decision difficulty; experiments.

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

*Authors are ordered alphabetically.

†This is a significantly revised version of a working paper with the same title that was originally circulated in December 2014. We thank Syngjoo Choi, Eric Danan, Itzhak Gilboa, Jan Heufer, Larry Samuelson and audiences at the Stanford Institute for Theoretical Economics (2019), Risk, Uncertainty and Decision (2017), Asian Meetings of the Econometric Society (2017), Annual Congress of the European Economic Association (2017), Workshop on Behavioural Game Theory (2016), World Congress of the Econometric Society (2015), Bounded Rationality in Choice (2014), Foundations of Utility and Risk (2014), Behaviour, Incentives and Contracts (2014), UEA, Georgetown, Johns Hopkins, William & Mary, Durham, Bath, WZB Berlin, Cyprus, Technion, Alicante, Aberdeen and St Andrews for helpful feedback. Gerasimou and Costa-Gomes gratefully acknowledge financial support from the British Academy (2012-13 SRG). Cueva acknowledges support from the Spanish Ministry of Science and Innovation (Grant PID2019-108193GB-I00) and the Generalitat Valenciana (Grant SEJI/2019/005). Any errors are our own.

‡School of Economics & Finance, University of St Andrews.

§Fundamentos del Análisis Económico (FAE), Universidad de Alicante.

¶School of Economics & Finance, University of St Andrews. Corresponding author: gg26@st-andrews.ac.uk.

‖School of Computer Science, University of St Andrews.

1 Introduction

Economic experiments on individual decision making typically require participants to choose a market alternative from those that the experimenter makes available to them. A possibility that arises when choice is forced in this way is that participants may be asked to "actively" choose an alternative in situations where they would have otherwise opted for the "choice deferral" outside option instead (Tversky and Shafir, 1992; Anderson, 2003; Bhatia and Mullett, 2016). For example, Shafir, Simonson, and Tversky (1993) illustrated such experimental findings in psychology with the following story: *"At the bookstore, [Thomas Schelling] was presented with two attractive encyclopedias and, finding it difficult to choose between the two, ended up buying neither –this, despite the fact that had only one encyclopaedia been available he would have happily bought it."*

When an individual is forced to choose in a situation like this she may do so in a way that is incompatible with utility maximization. Indeed, as pointed out by Luce and Raiffa (1957), *"intransitivities often occur when a subject forces choices between inherently incomparable alternatives"*. This argument naturally leads to the following question:

Do people make more consistent choices when they are not forced to choose?

If a decision maker is rational in the sense that her behavior is consistent with the maximization of a stable, complete and transitive preference relation, then she will always choose a most preferred feasible alternative from every menu. Considering, however, the large body of work which shows that people often do not behave in such a utility-maximizing fashion, the above-raised question becomes pertinent. Yet, despite its methodological, theoretical and empirical significance, no experimental/empirical study that we are aware of has studied it before.

In this paper we raise and attempt to answer this question for the first time. In summary:

1. We report on two lab experiments that implemented a new binary-treatment design that elicited forced as well as deferred/non-forced active choices from 26 menus derived from the same set of 5 real goods.
2. Using as our consistency criteria the proportions of WARP and Congruence violators (Richter, 1966) as well as the Houtman and Maks (1985); Apesteguia and Ballester (2015) rationality indices and an adaptation of the Selten (1991) measure of predictive success, we find that non-forced choice subjects were generally

more likely to be consistent and their active choices were closer to being rational compared to forced-choice subjects.

3. The primary source of deviation from rational choice for most deferring non-forced-choice subjects is their occasionally hesitant behavior that manifests itself in deferrals, and not the potential inconsistency of their active choices.
4. Analyzing additional data from an indecisiveness personality questionnaire, we also find that:
   (a) Within the non-forced-choice treatment, subjects who occasionally deferred had higher indecisive-personality scores than those who did not.
   (b) Indecisive forced-choice subjects were more likely to make inconsistent choices than decisive ones.

Our choice-based findings suggest that non-forced-choice economic experiments can complement the standard forced-choice ones in a fruitful way. In particular, revealed-preference analyses that are appropriate for such data can offer new and potentially important insights that may be helpful towards distinguishing between rational, hesitant and genuinely inconsistent behaviors.

2 Experimental Design

We conducted two between-subjects choice experiments with real goods at the University of St Andrews Experimental Economics Lab in September 2013 - February 2014 (henceforth Exp1) and again in September - October 2018 (Exp2). We recruited subjects using ORSEE (Greiner, 2015). In both experiments we used a between-subjects design with two treatments: Forced-Choice (FC) and Non-Forced-Choice (NFC).

In the main phase of the experiment all subjects were presented with a sequence of menus that were generated from a set of five headsets. Their brand names and models were chosen so that the products' prices were similar but their attributes differed in ways that made comparisons between them non-trivial. For instance, some headsets were basic but with well-known brand names, whereas others were more sophisticated or had some superior or distinctive features but were associated with less recognizable brand names (e.g. the headset with the less commonly known brand name was wireless/bluetooth whereas all others were not). 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.[1]

[1]Subjects could not access the internet during the experiment, and therefore could not read product reviews either

The order in which menus appeared was random and varied across sessions but was constant within sessions. Each item appeared top-left, middle, top-right etc. in an even manner across menus. Subjects in the FC treatment were asked to choose an item from all menus, without being able to defer choice. Subjects in the NFC treatment had the opportunity to choose one item or to select *"I'm not choosing now"* in each menu. No subject in either treatment could go back to review and change their decision after they were past a menu during this "main phase" of the experiment.

Once all subjects in both treatments completed this task, they proceeded to the "final phase" of the experiment, where one menu was randomly selected separately for each of them. Subjects were then reminded of the decision they had made at that menu in the "main phase". Subjects knew from the beginning that 1 out of every 4 of them (this feature of the design was entirely driven by the experimenters' budget constraints) would be randomly selected to win the item of their *final* choice from their randomly selected menu at the end of the experiment. We refer to such subjects as *"winners"*. Participants also knew from the beginning that winners might face some costs which would be deducted from their initial monetary endowment, I. These costs depended on the subjects' decisions at the randomly selected menu in the "main" and "final" phase of the experiment.

In particular, if a subject that was later drawn as winner had decided to choose an option other than the one she selected from this menu in the main phase, an amount $c_r < I$ was taken away from her initially allocated I. In contrast, there was no deduction if the subject opted for the same headset when she chose from that menu in the final phase. These features were shared by both the FC and NFC treatments. Subjects in the NFC treatment who deferred at their randomly selected menu in the main phase and were later drawn as winners incurred the cost of a $c_d < c_r$ deduction from the initial allocation of I. Finally, participants were told from the beginning that if they were *not* selected to win a headset they would receive their endowment, I, irrespective of their main- and final-phase decisions at the randomly selected menu.

The FC treatment follows procedures used in standard forced-choice experiments. On the other hand, the NFC treatment is structured in a way that parallels real-world situations where delaying an active choice is costly but changing such a choice is even costlier, e.g. the decision to delay or buy immediately from a store with a restrictive returns policy.

All 26 menus with 2–5 headsets were shown to subjects in both experiments, using the same z-Tree experimental interface (Fischbacher, 2007). Singleton menus were also included in Exp1, even though they involved no meaningful choice (note that deferral

through the PCs they were using or through their smartphones, as using the latter was not allowed.

is dominated in such menus). Those menus were not shown to subjects in Exp2.

Exp1 subjects in both treatments were told that they would be allowed to try out the headphones in the randomly selected menu before making their second and final choice from that menu. It is possible, therefore, that the decision to opt for costly deferral in Exp1 could be interpreted as the rational decision to buy more information before making an active choice. We stress, however, that we did not design the treatment with that in mind, and no additional information was given to subjects in writing or in some other objective way. Subjects in Exp2 on the other hand could not inspect the headphones in their randomly selected menu between their first and final choice, and neither did they receive any additional information. Deferring in this case can in principle be driven by an urge to avoid the task of making a preference-guided active choice when the menu in question generates decision difficulty. We set $c_r = £4$, $c_d = £1$, $I = £7$ in Exp1 and $c_r = £6$, $c_d = £0.5$, $I = £8$ in Exp2 to compensate for this fact and to account for inflation.

In Exp1 we also tried to elicit subjects' potential indifference between alternatives.[2] Following their choices at each binary menu, we asked whether their choice reflected a preference for their chosen alternative over the other, whether they found both to be equally good and so chose randomly, or whether they chose for another reason. Subjects were told that if they stated that they found both to be equally good, we would pick one of the alternatives at random for them at that menu, and they would not have the opportunity to revise their choice should this menu be selected in the final phase of the experiment. Subjects, therefore, did not have a strict incentive to state "indifference" in their responses.[3] To simplify and facilitate the subjects' understanding of the experimental instructions we did not elicit indifferences in Exp2.

The main differences between our NFC treatment design and a related one that was proposed in Danan and Ziegelmeyer (2006) that also allows for choice delay can be summarized as follows: (i) we framed decision problems as ones where subjects had to choose *from* menus, not *between* them (Kreps, 1979; Dekel, Lipman, and Rustichini, 2001); (ii) we presented subjects with decision problems of many sizes, not only binary ones; (iii) we allowed costly switching to a different active choice in the final

[2]In an earlier version of this paper we used these responses to transform the single- or empty-valued choice functions into multi- or empty-valued choice correspondences, aiming to give both FC and NFC subjects the benefit of the doubt and maximize their behavioral consistency. We acknowledge, however, that any choice-augmentation method of this kind is bound to make assumptions, which in some cases may involve giving an equal weight to the non-strictly incentivized indifference data and the payoff-relevant choice data. For this reason, we are focusing here on the primitive choice data only. More information about the above analysis is available upon request.

[3]It is possible, however, that some subjects who occasionally declared themselves to be indifferent did so because of a "preference for randomization" that was analogous to the one reported in Agranov and Ortoleva (2017). Subjects in these authors' experiments –which were conducted independently of our own and with a different focus– were faced with the same decision problem several times and, in addition to making different choices in the same menu, they were often willing to incur a small cost to have their choice made randomly. Unlike their experiments, ours was not specifically designed to test for such a preference. This is reflected in our non-repeated decision problems and non-strictly incentivized indifference statements.

phase of the experiment for subjects who had previously made such a choice at their randomly selected menu [switching was not possible for the subjects' "commitment" singleton or "flexibility" doubleton menu in Danan and Ziegelmeyer (2006)]; (iv) we introduced explicit deferral costs that were taken off some initial endowment instead of making the rewards marginally more appealing with non-deferral. Finally, because our experiments are predominantly or exclusively choice-based and feature choices over general choice alternatives (which in this case are real goods), our design is also distinct from the imprecise-interval elicitation ones that have been used in several experiments with binary menus of lotteries (see Cubitt, Navarro-Martinez, and Starmer (2015) and references therein).

3 Rationality and Active-Choice Consistency

Given a finite set X of general choice alternatives and a collection of nonempty subsets of X (to be called *menus*), the decision maker's observable behavior is described by a choice correspondence C that satisfies $C(A) \subseteq A$ for every menu A in that collection. By letting $C(A) = \emptyset$ here we model the situation where the decision maker, by opting for the no-choice/choice-deferral outside option, has chosen none of the feasible *market* alternatives. The agent's weak, strict and indirect weak revealed preference relations $\succsim^R$, $\succ^R$ and $\succsim^{\widehat{R}}$ are defined, respectively, by $x \succsim^R y$ if there is a menu A such that $x \in C(A)$ and $y \in A$; $x \succ^R y$ if there is a menu A such that $x \in C(A)$ and $y \in A \setminus C(A)$; and $x \succsim^{\widehat{R}} y$ if there are alternatives $x_1, \ldots, x_n$ such that $x = x_1$, $y = x_n$ and $x_i \succsim^R x_{i+1}$ for all $i = 1, \ldots, n-1$. Given these concepts and notation, we can now introduce compactly the following fundamental principles of choice consistency (see also Chambers and Echenique (2016) and references therein):

Weak Axiom of Revealed Preference (WARP)
$x \succ^R y \implies y \not\succsim^R x.$

Congruence
$x \succ^R y \implies y \not\succsim^{\widehat{R}} x.$

In words, Congruence amounts to active-choice acyclicity, whereas WARP is its special case that rules out choice reversals/choice cycles of length two.

A choice correspondence C is *rational* if there exists a complete and transitive preference relation $\succsim$ on X such that, for every menu A,

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

Rationality implies conformity with WARP and Congruence. In our environment of non-forced choice, however, the converse is not necessarily true: even when active choices satisfy these axioms, the overall behavior will be incompatible with that model if the decision maker defers at some menu(s). Two equally important questions naturally arise in this case for such an agent's behavior:

1. Are *all* her decisions compatible with (1)?
2. If not, are her *active-choice* decisions compatible with (1)?

Clearly, when the answer to the first question is positive, the agent behaves as if she were a utility maximizer. But even when that's not true and the answer to the second question is positive instead, rationality is not contradicted in a strong sense because an incomplete but transitive preference relation is recoverable from her active choices and, by standard extension results, this can in theory be extended at a later stage to a complete relation. Therefore, rather than suggesting "irrationality", this case signifies a "not-yet-rational" behavior. Our primary aim is to investigate the existence and frequency of precisely this unexplored kind of behavior.

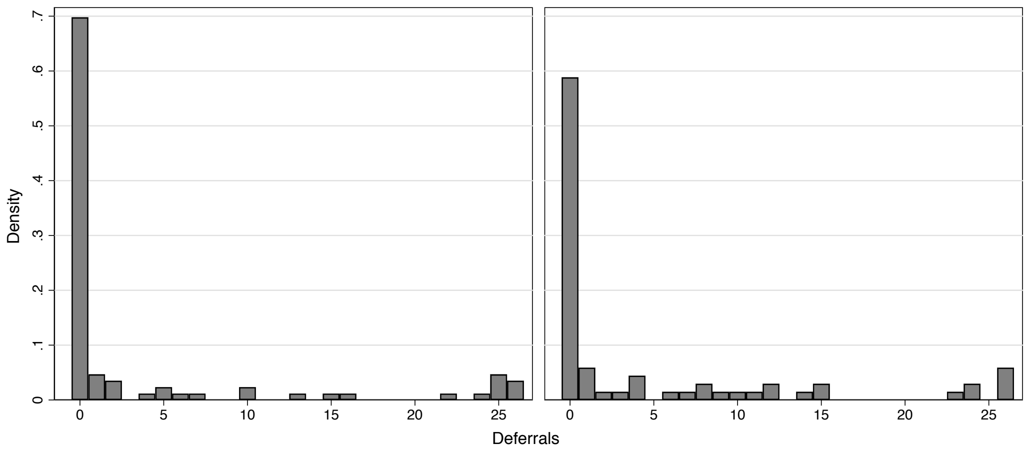
Now, when an agent's active-choice decisions are incompatible with rational choice, one is naturally interested in measuring the severity of this incompatibility. Two such measures are suggested by the classic Houtman-Maks (1985) (henceforth "HM") index and the more recently proposed Apesteguia-Ballester (2015) (henceforth "Swaps") index. The former corresponds to the smallest number of choices that need to be changed in this dataset in order for the remaining ones to be perfectly consistent with some instance of utility maximization, i.e. some complete and transitive relation $\succsim$ that explains the data as in (1). The Swaps index instead is computed by minimizing over all such instances the sum of the total number of alternatives in each menu that are better than the agent's chosen alternative.

Let us illustrate the differences between the two measures with the following example (we write $C(\{a,b\}) \equiv C(a,b)$ etc., to ease notation): $C(a,b) = a$, $C(a,c) = c$, $C(b,c) = c$, $C(a,b,c) = b$ on $X := \{a,b,c\}$. The corresponding HM and Swaps scores here are 1 and 2, respectively: changing $C(a,b,c)$ from b to c makes that dataset compatible with the preference order $c \succ a \succ b$ that is suggested by the binary choices; however, although $c \succ a \succ b$ is indeed the best-matching ordering for that dataset, Swaps also takes into account that the "erratic" choice of b in $C(a,b,c)$ is two ranks away from the optimal choice of c in that menu, and is therefore associated with a score of 2. Thus, Swaps incorporates menu-specific cardinal information on the agent's deviations from rational choice that might be interpretable as quantifying the welfare loss associated with the agents' "wrong" choices.

4 Results

There were 75 FC, 86 NFC evaluable subjects in Exp1 and 53 FC, 68 NFC such subjects in Exp2.[4] Figure 1 shows the distribution of deferrals of NFC subjects in Exp1 and Exp2. We find no statistical difference between them ($p = 0.2$, two-sided Mann-Whitney U-test). In particular, Exp1 and Exp2 subjects deferred on average at 3.8 and 4.8 menus, i.e. in about one out of every six menus. Moreover, 35% of these subjects deferred at least once, and did so at an average of 12.5 (Exp1) and 11.5 (Exp2) menus, with most subjects deferring at between one and sixteen menus, and 8 subjects deferring at either all or all but one or two menus. The proportions of deferring NFC subjects and their rates of deferral were also not statistically different between Exp1 and Exp2 ($p = 0.2$, two-sided Fisher's exact test and $p = 0.9$, two-sided Mann-Whitney U test). Overall, we can thus conclude that deferring behavior was similar in both experiments, with around one third of the subjects deferring, and with a rate of deferrals at non-singleton menus of almost one half among these subjects.

Figure 1: Deferral relative frequencies in Exp1 (left) and Exp2.



We now compare FC and NFC subjects' rationality/conformity with (1) and their active-choice consistency/conformity with Congruence and WARP within each experiment separately, as well as when the data from the two experiments are pooled.

[4]Some participants were excluded from the analysis due to: (i) failing to demonstrate that they understood the instructions (12 and 17 quiz failures, respectively); (ii) deferring at singleton menus in Exp1 (58); (iii) behaving randomly (1 FC and 3 NFC subjects in Exp1; 1 FC subject in Exp2). A participant was thought to behave randomly if the HM score on his/her active choices was weakly greater than the 2.5% cut-off value of the subject-specific HM distribution that resulted from simulated uniform-random active choices at only those menus where (s)he made active choices. Our results without the latter two kinds of exclusions are presented in the Online Appendix.

Although the two experiments differed in some important ways, their primary features (namely their choice menus, experimental interface and general reward structure) were identical. Together with the results from the previous tests which show that the distribution of deferrals in the two experiments were not significantly different and further tests showing that the same is true for the various measures of choice consistency used throughout the paper, this fact suggests that the pooled data –and the resulting larger sample size– allow for an additional informative comparison.[5]

Table 1 presents the results from comparing the proportions of subjects who violated WARP and Congruence in the two treatments. Every Congruence violator turned out to also violate WARP, in both experiments and treatments. Moreover, the proportion of subjects who violated these axioms is 14 percentage points higher in the FC treatment in each of Exp1 and Exp2. The difference between the proportions is statistically significant at the 5% level when the data from the two treatments is pooled, even if is not within each treatment.[6]

Table 1: Proportions of subjects with zero active-choice cycles (p-values from two-sided Fisher exact tests).

	WARP/Congruence		
	Exp1	Exp2	Pooled
FC	55% (41/75)	60% (32/53)	57% (73/128)
NFC	69% (59/86)	74% (50/68)	71% (109/154)
p-value	0.075	0.170	0.018
N	161	121	282

Table 2 shows the results from comparing the distributions of HM and Swaps scores across treatments. These results reinforce the ones reported in Table 1. In particular, NFC subjects' active choices are closer to the rational choice model relative to FC subjects according to both the HM and Swaps indices, and the difference between the respective distributions in each case is significant at the 5% level in the pooled data. Notably, the average scores are generally lower than or very close to 1 in both experiments and treatments, suggesting that subjects were on average less than one decision away from perfect conformity with utility maximization, conditional on the menus where they made active choices.

[5]In particular, we find no statistically significant differences between Exp1 and Exp2 in the proportions of WARP/Congruence violating subjects in either treatment ($p > 0.5$, Fisher's exact tests), nor in their Houtman-Maks or Swaps scores on active choices ($p > 0.8$, Mann Whitney U tests).

[6]Prior to conducting Exp2, the weak evidence for the existence of binary choice cycles (i.e. Congruence violations restricted to menus with two alternatives) in the 10 binary menus over 5 goods in Exp1 prompted us to investigate further the comparative incidence of such cycles in a more focused experiment that only featured choices from the 15 possible binary menus over a set of 6 lotteries that had three strictly positive monetary outcomes and were pairwise unrelated by second-order stochastic dominance. That experiment was conducted at the University of Alicante Experimental Economics Lab in January – March 2018. Binary cycles were relatively scarce in both treatments of that experiment too, with no significant differences between FC and NFC subjects. Because the structure of that experiment was very different from those of Exp1 and Exp2, we will report on it in more detail elsewhere.

Table 2: Mean Houtman-Maks and Swaps scores on active choices (number of subjects in parenthesis; p-values from two-sided Mann-Whitney U tests).

	Houtman-Maks			**Swaps**		
	Exp1	Exp2	Pooled	Exp1	Exp2	Pooled
FC	0.80 (75)	0.98 (53)	0.88 (128)	0.88	1.09	0.97
NFC	0.58 (86)	0.75 (64)	0.65 (150)	0.62	0.86	0.72
p-value	0.088	0.205	0.032	0.108	0.199	0.035
N	161	117	278	161	117	278

The findings in Tables 1 and 2 provide indications that subjects who are not forced to choose make more consistent active choices than subjects who are forced to do so. Importantly, however, although rational decision makers would have zero axiom violations and HM/Swaps scores regardless of whether they operated in an FC or an NFC environment, for agents whose behavior is incompatible with this model one naturally expects that the incidence and magnitude of deviations from it will depend on the number of their active choices and on the specific menus where these choices were made. To address this comparability issue, we also perform a "like-for-like" cross-treatment comparison of Selten's (1991) measure of *predictive success*, the relevance of which in revealed preference analysis is increasingly appreciated since Beatty and Crawford (2011).

To this end let us first recall that, when applied to the rational choice model, Selten's measure is defined by subtracting from the proportion of actual subjects whose behavior is perfectly explained by it (called the "pass rate") the proportion of uniform-random simulated subjects that are also explained by that model perfectly (called the "area"). The latter proportion is informative of the *power* of the revealed-preference test, a concept that originated in Bronars (1987). The difference between these two proportions lies in the $[-1, 1]$ interval, and higher values point to a higher predictive power for the model relative to the uniform-random model, conditional on the given decision environment.

Let us denote by m^i_{FC}, p^i_{FC} and a^i_{FC} the predictive success rate, pass rate, and area that correspond to the rational choice model in the FC treatment of experiment $i \in \{1, 2\}$. By definition,

$$m^i_{FC} = p^i_{FC} - a^i_{FC}.$$

To apply this method to the NFC treatment as well, we proceed as follows. For a given subject n in that treatment, we first find the proportion of random-behaving simulated subjects whose active choices are rational once they are confined to the same menus as those where subject n made her own active choices. We then define the pass rate p^i_{NFC} in that treatment and experiment to be the proportion of actual NFC subjects whose corresponding active choices were rational, and the area a^i_{NFC} to be the average proportion of the random-behaving subjects –conditioned as above– that were rational. The predictive success rate of rational choice in the NFC treatment is then defined by

$$m^i_{NFC} = p^i_{NFC} - a^i_{NFC}.$$

A comparison between m^i_{FC} and m^i_{NFC} therefore accounts for the fact that subjects in the latter treatment make fewer active choices in general, and corrects the possibly higher pass rates that may emerge as a result of this fact by conditioning the area-determining simulations accordingly.

Because our subjects made decisions from the full set of 26 possible menus that can be derived from a set of 5 alternatives, the power of the revealed-preference test for this model is very high in the FC environment, with $a^i_{FC} \approx 0$. That is, such randomly made forced-choice decisions can be consistent with rational choice very rarely (at a roughly 1 out of 2.83 billion rate). It follows, therefore, that $m^i_{FC} \approx p^i_{FC}$ in our case or, equivalently, that the rational choice model's predictive success rates in the FC treatment can be taken to coincide with the proportion of such subjects that this model explains *perfectly*. Because approximately 54.6% and 60.4% of FC subjects in Exp1 and Exp2, respectively, behaved as if they were perfect utility maximizers, this in turn translates into the Selten rates $m^1_{FC} \approx 0.546$ and $m^2_{FC} \approx 0.604$ for utility maximization in the two experiments, with $m_{FC} \approx 0.570$ in the pooled data.

In the NFC treatment on the other hand, some of the collections of menus where subjects made active choices were associated with non-negligible areas in the corresponding subject-specific simulations. Taking both these areas and the pass rates of utility maximization into account, the Selten measure in the NFC treatment was $m^1_{NFC} \approx 0.686 - 0.103 = 0.583$ in Exp1 and $m^2_{NFC} \approx 0.718 - 0.035 = 0.683$ in Exp2, with $m_{NFC} \approx 0.626$ in the pooled data. Therefore, rational choice predicts much better than random behavior in both treatments, and more so in the NFC than in the FC treatment.

The results presented so far suggest that a significant fraction of subjects are willing to opt for costly choice deferral when given the opportunity to do so and that active choices are more likely to be consistent with rationality in such an environment. A possible interpretation of these findings is that some subjects in our experiment may have been *indecisive* at some menus and so opted to defer when this option was available to them. The increase in inconsistencies observed in the FC treatment could then be driven by the fact that such potentially indecisive subjects who are forced to choose may be more likely to violate rationality.

Table 3: Active-choice consistency for decisive and indecive subjects
(pooled data; number of subjects in parentheses)

	WARP/Congruence			**Houtman-Maks**			**Swaps**		
	Decisive	Indecisive	p-value	Decisive	Indecisive	p-value	Decisive	Indecisive	p-value
FC	69% (27/39)	46% (23/50)	0.033	0.54 (39)	1.18 (50)	0.022	0.59	1.34	0.018
NFC	71% (39/55)	76% (34/45)	0.656	0.54 (54)	0.57 (44)	0.458	0.61	0.64	0.495
p-value	1.00	0.004		0.987	0.003		0.942	0.003	
N	94	95		93	94		93	94	

Notes: WARP/Congruence lists the percentage of subjects with zero active-choice cycles (p-values from two-sided Fisher's exact tests); Houtman-Maks/Swaps list the respective mean scores on active choices (p-values from two-sided Mann-Whitney U tests).

To test this prediction we examine the relationship between subjects' behavior in the experiment and their responses to the *indecisive personality* questionnaire of Germeijs and Boeck (2002) that we administered at the end of each session. The questionnaire elicited 0–7 Likert-scale responses in 11 original statements and also in 11 statements that made the exact opposite claim. Examples include *"I find it easy to make decisions"*, *"It is hard for me to come to a decision"*, *"I delay deciding"*, *"I don't postpone making decisions to a later date"*. In what follows, we refer to subjects with relatively higher scores as *indecisive* and those with relatively lower scores as *decisive*.[7] We ask:

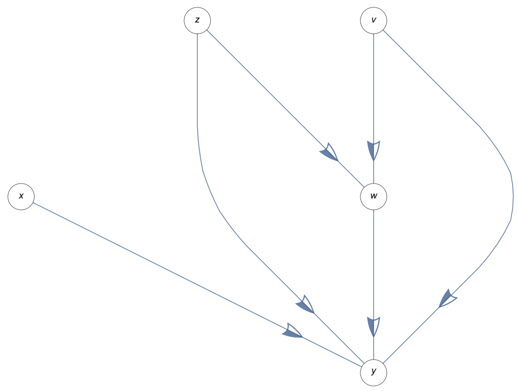
1. Are deferring subjects more likely to be indecisive?
2. When forced to choose, are indecisive subjects more inconsistent than decisive ones?
3. Is the observed forced-choice treatment effect driven by indecisive subjects?

We find evidence suggesting a positive answer to each of these questions. Specifically, the average indecisive-personality score of the 57 deferring and 100 non-deferring NFC subjects in the two experiments were 0.420 and 0.348, respectively ($p = 0.052$; two-sided Mann-Whitney U test). In addition, Table 3 shows that: (i) decisive FC subjects –defined as those in the bottom third of the indecisive-personality distribution– were significantly more consistent than indecisive ones –defined as those in the top third; (ii) indecisive subjects were significantly more consistent in NFC than in FC while no forced-choice treatment effect is observed on decisive subjects.

Taken together, the preceding results (as well as those presented in Table 4 in the Online Appendix) suggest that the primary source of deviation from rational choice for most deferring NFC subjects is their occasionally hesitant behavior that manifests itself in deferrals, and not the potential inconsistency of their active choices. As we pointed out in Section 3, although such behavior is incompatible with rationality, it is perhaps

[7]The indecisiveness score was calculated as the average response to the 22 items, normalized between 0 and 1. We discarded subjects' answers to an item when their answer to the oppositely worded item differed by 3 or more points in the Likert scale. Subjects' average response consistency defined in this way was approximately 90%.

Figure 2: The partial preferences recovered from an NFC subject (arrows point to dominated options).



better viewed as "not-yet-rational" rather than as "irrational". Observing such "not-yet-rational" behavior raises the question of whether any additional information may also be recovered about deferring subjects' decision process and preferences.

A natural starting point in this regard is to investigate whether the deferring subjects' behavior is potentially compatible with the preference-maximization principle laid out in (1) but when preferences are strictly incomplete. This model of dominant choice with incomplete preferences (Gerasimou, 2018, Section 2) portrays the decision maker as making an active choice at a menu if and only if there is a most preferred alternative in that menu, and as deferring otherwise. As such, it predicts behavioral patterns like the one quoted in the opening paragraph.

When we analyze the deferring NFC subjects' *combined* deferral and active-choice data we find that this model provides a better fit than rational choice for more than 80% of those subjects in the two experiments, with goodness of fit now captured by the extension of the HM method that detects how close the combined deferral and active-choice data of a given subject are with (1) when preferences are potentially incomplete (see Table 5 in the Online Appendix). This fit was perfect for some subjects and allowed for the recovery of a unique partial preference ordering (see Figure 2 for an example and Figure 3 in the Online Appendix for the behavior that generated it). Although this fact does not prove that those subjects deferred because their preferences over the five headsets were indeed incomplete, it does suggest that their behavior can be thought of *as if* it had been generated by such a decision process and preference relation. Finally, because our experimental data feature single- or empty-valued choices, we cannot use them to conduct a proper test of the complementary class of undominated-choice models of incomplete preference maximization such as those analyzed in Schwartz (1976) and Eliaz and Ok (2006). We leave the task of testing these no-deferral models for future work.

5 Concluding Remarks

In this paper we raised the question of whether not forcing experimental subjects to make active choices from the menus they are presented with, and allowing them instead to delay making such choices, could enable researchers to recover more consistent behaviors and stable preferences from such subjects, possibly over only a subset of the original set of alternatives. To answer this question we implemented a new experimental design with a forced and a non-forced choice treatment. Analyzing the data from two experiments that implemented this design on 26 menus over 5 real goods, and using several non-parametric methods that have been suggested in the literature, we find evidence that higher proportions of subjects make consistent active choices when not forced to choose, and that the active-choice behavior of subjects in that treatment is generally closer to being considered rational. Moreover, the combined deferral and active-choice decisions of non-forced-choice subjects and the data from an indecisive-personality questionnaire provide independent indications that some of the deferrals and active-choice inconsistencies may have originated in menus that generated higher decision difficulty. Our analysis and results suggest that non-forced-choice experiments can offer new and potentially important insights on subjects' behavior and preferences.

References

AGRANOV, M. AND P. ORTOLEVA (2017): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy*, 125, 40–68.

ANDERSON, C. J. (2003): "The Psychology of Doing Nothing: Forms of Decision Avoidance Result from Reason and Emotion," *Psychological Bulletin*, 129, 139–167.

APESTEGUIA, J. AND M. A. BALLESTER (2015): "A Measure of Rationality and Welfare," *Journal of Political Economy*, 123, 1278–1310.

BEATTY, T. K. M. AND I. A. CRAWFORD (2011): "How Demanding is the Revealed Preference Approach to Demand," *American Economic Review*, 101, 2782–2795.

BHATIA, S. AND T. L. MULLETT (2016): "The Dynamics of Deferred Decision," *Cognitive Psychology*, 86, 112–151.

BRONARS, S. G. (1987): "The Power of Non-parametric Tests of Preference Maximization," *Econometrica*, 55, 693–698.

CHAMBERS, C. P. AND F. ECHENIQUE (2016): *Revealed Preference Theory*, Econometric Society Monograph 56, Cambridge: Cambridge University Press.

Cubitt, R., D. Navarro-Martinez, and C. Starmer (2015): "On Preference Imprecision," *Journal of Risk and Uncertainty*, 50, 1–34.

Danan, E. and A. Ziegelmeyer (2006): "Are Preferences Complete? An Experimental Measurement of Indecisiveness Under Risk," *Working Paper*.

Dekel, E., B. L. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

Eliaz, K. and E. A. Ok (2006): "Indifference or Indecisiveness? Choice-theoretic Foundations of Incomplete Preferences," *Games and Economic Behavior*, 56, 61–86.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 10, 171–178.

Gerasimou, G. (2018): "Indecisiveness, Undesirability and Overload Revealed Through Rational Choice Deferral," *Economic Journal*, 128, 2450–2479.

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.

Greiner, B. (2015): "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

Houtman, M. and J. A. Maks (1985): "Determining All Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden*, 19, 89–104.

Kreps, D. M. (1979): "A Representation Theorem for 'Preference for Flexibility'," *Econometrica*, 47, 565–577.

Luce, R. D. and H. Raiffa (1957): *Games and Decisions*, New York: Dover.

Richter, M. K. (1966): "Revealed Preference Theory," *Econometrica*, 34, 635–645.

Schwartz, T. (1976): "Choice Functions, "Rationality" Conditions, and Variations of the Weak Axiom of Revealed Preference," *Journal of Economic Theory*, 13, 414–427.

Selten, R. (1991): "Properties of a Measure of Predictive Success," *Mathematical Social Sciences*, 21, 153–167.

Shafir, E., I. Simonson, and A. Tversky (1993): "Reason-Based Choice," *Cognition*, 11, 11–36.

Tversky, A. and E. Shafir (1992): "Choice under Conflict: The Dynamics of Deferred Decision," *Psychological Science*, 3, 358–361.

Online Appendix

Experiment 1 – Non-Forced Choice Treatment Instructions

General procedure

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

At the start of the experiment you will be allocated £7. You will then be presented with a sequence of 31 *menus* of HSs (a menu is simply a collection of HSs). Each menu may have 1 to 5 HSs. When a menu appears on your screen you will have the opportunity to look at the image of each HS in that menu and also to read a short description of its main features. You will then be able to choose one of the available HSs, or to select the option *"I'm not choosing now"*.

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

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

You will then get to examine the actual HSs contained in menu R and to try them while listening to a song. Lastly, you will be asked to make a ***final choice from R*** (not choosing a HS is not possible at this stage). One in every four participants will be randomly selected to win the HS **of their final choice** from their randomly selected menu R.

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

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

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

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

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

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

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

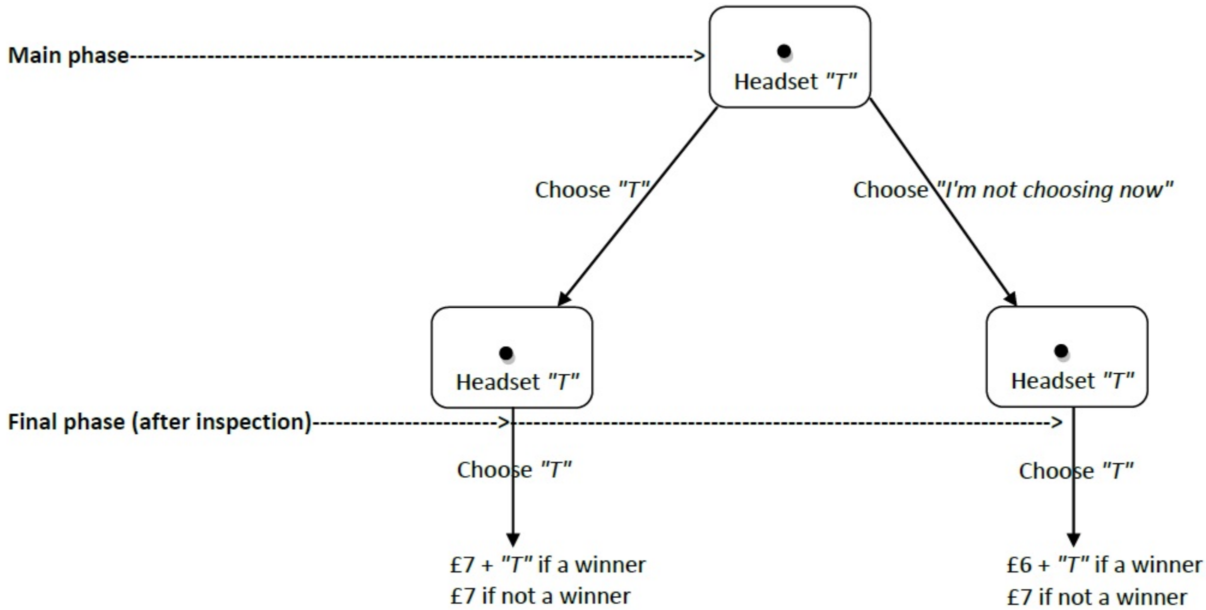
During the phase when you are presented with the 31 menus, whenever ***a menu of exactly two HSs*** comes up ***and you have chosen one of them***, a short follow-up question will ask you to state if you *preferred* the chosen HS over the non-chosen one, or if the non-chosen one was *equally good* to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

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

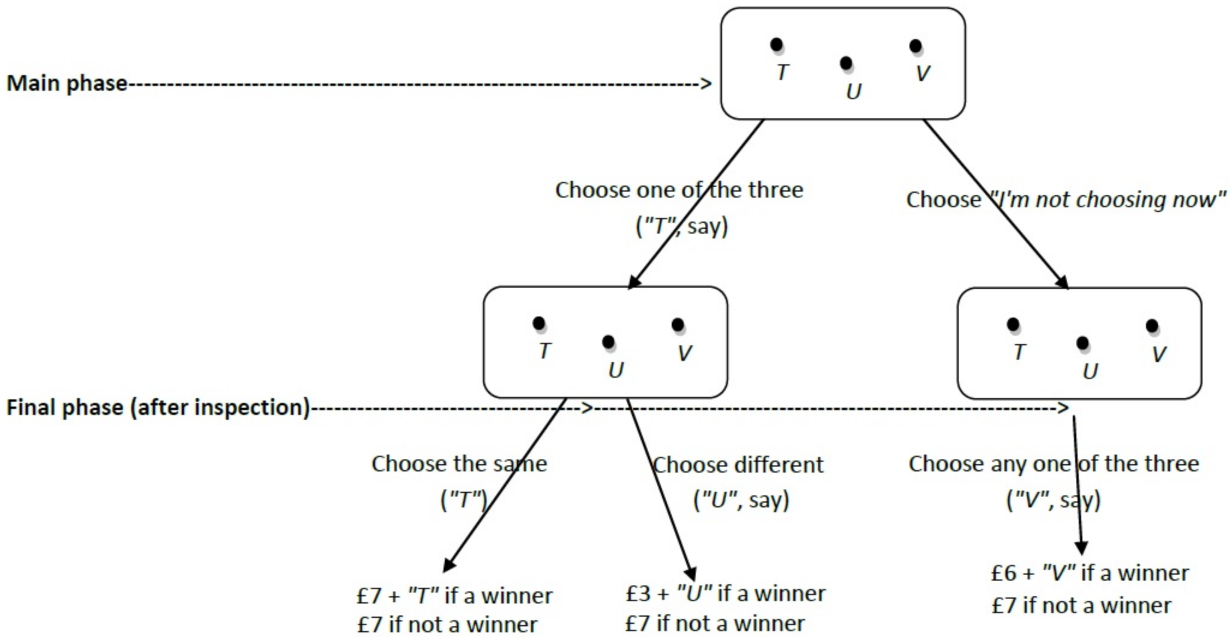
1) *If you had chosen a headset from R initially,* you will *not be able to change your decision at this stage*. *One of the two HS will be randomly selected* and you will win this HS if you are picked as a winner.

2) *If you had chosen "I'm not choosing now" at R initially*, then *one of the two HSs will be randomly selected* and you will win this HS if you are picked as a winner.

Example 1:
Randomly selected menu with only one headset

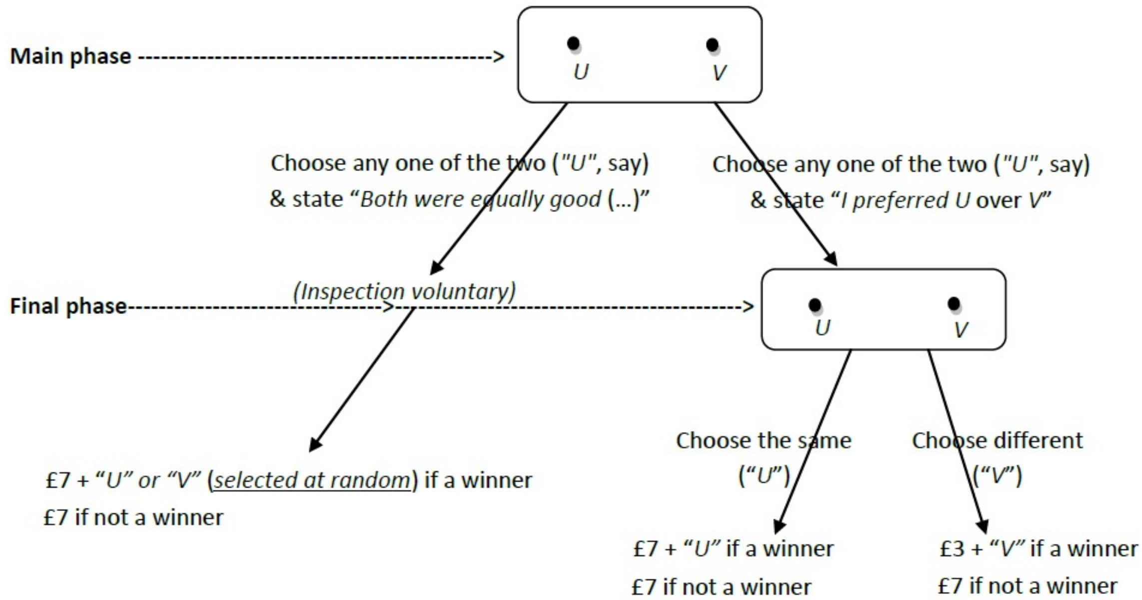


Example 2:
Randomly selected menu with three headsets



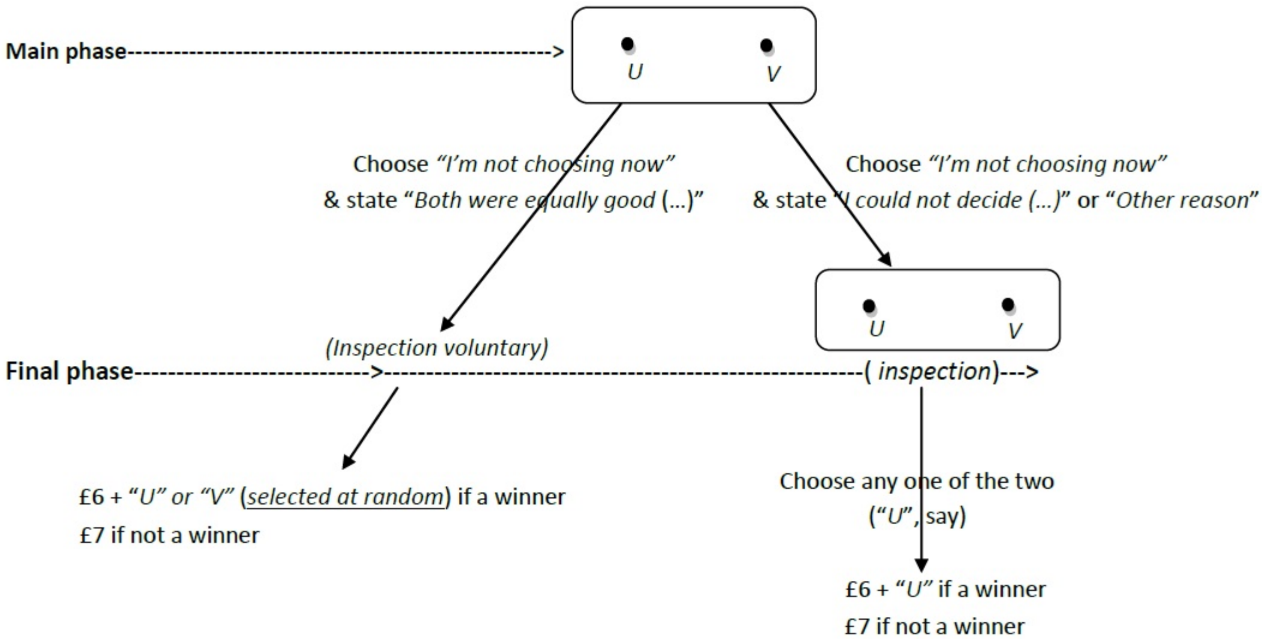
Example 3:

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



Experiment 1 – Forced Choice Treatment Instructions

General procedure

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

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

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

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

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

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

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

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

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

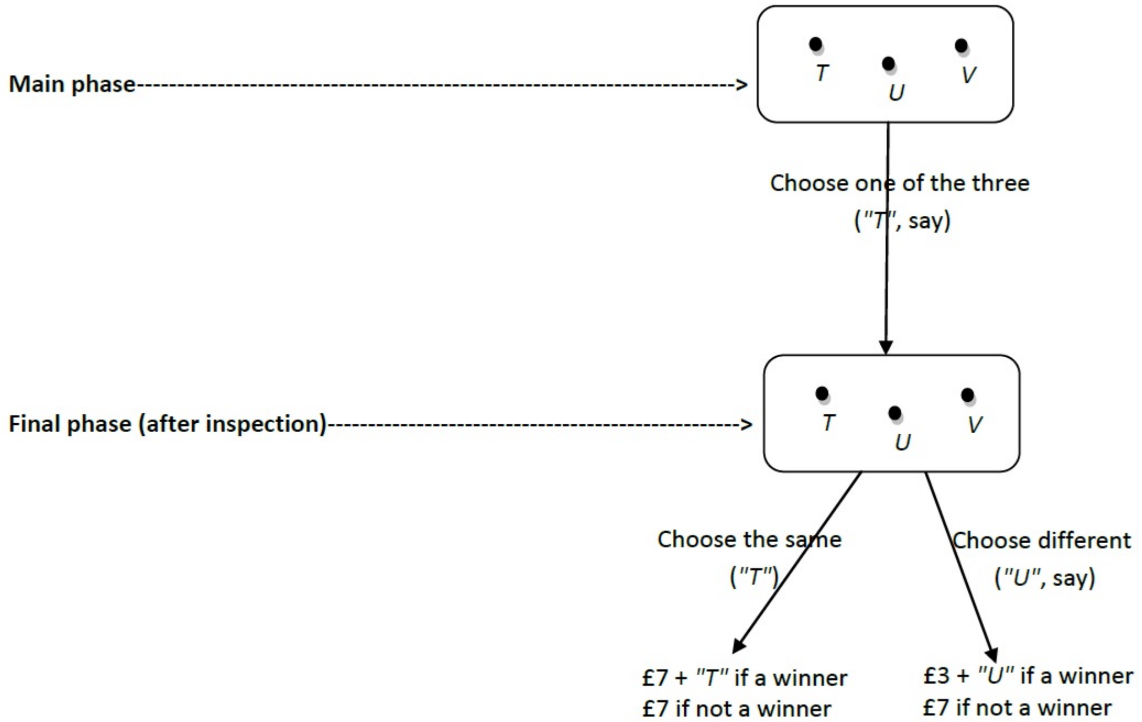
B) Suppose that when you first saw menu R ***you had chosen some HS from it. If you chose a different one the second time, you will receive £3 of the £7*** initially allocated to you.

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

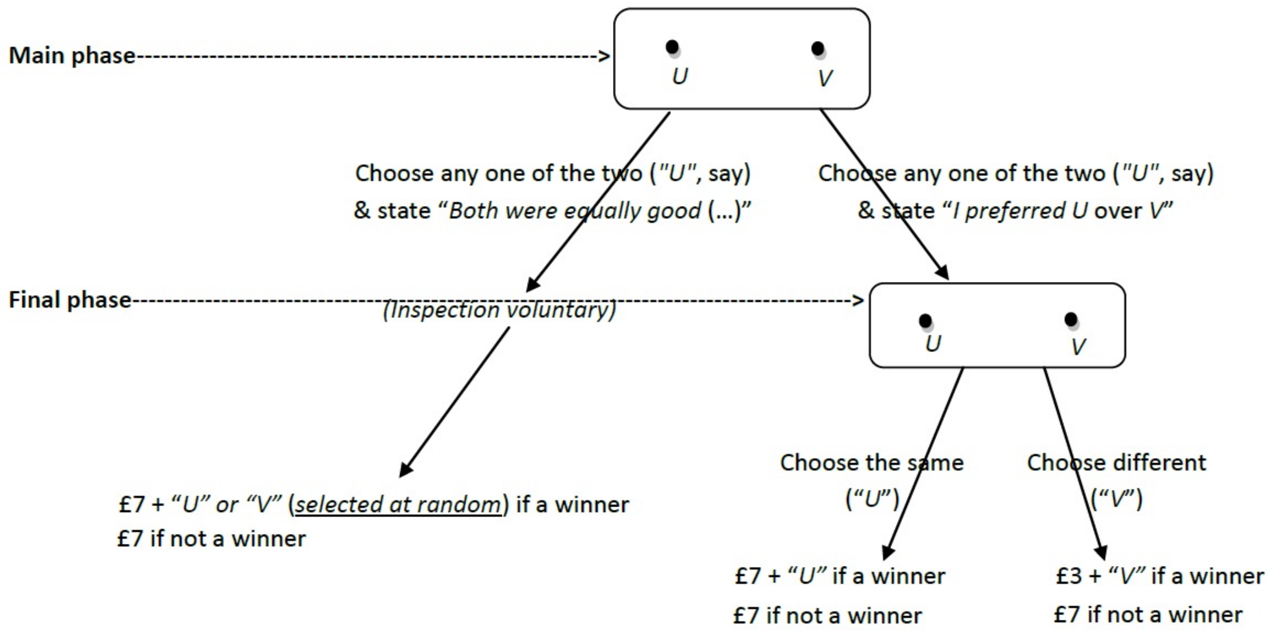
During the phase when you are presented with the 31 menus, whenever ***a menu of exactly two HSs*** comes up and you have chosen one of them, a short follow-up question will ask you to state if you *preferred* the chosen HS over the non-chosen one, or if the non-chosen one was *equally good* to the one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

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

Example 1:
Randomly selected menu with three headsets



Example 2:
Randomly selected menu with two headsets



Experiment 2: Non-Forced Choice Treatment Instructions

General procedure

This experiment aims to study people's choice behaviour.

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

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

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

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

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

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

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

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

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

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

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

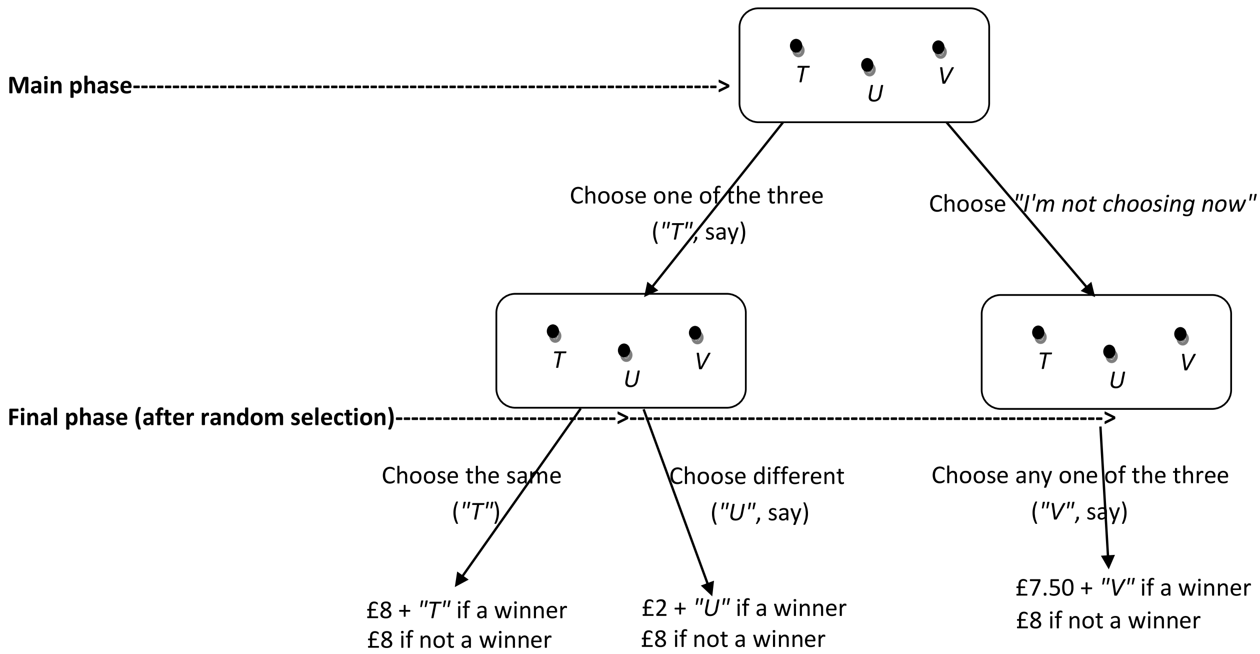
B) Suppose that when you first saw menu R you had chosen some HS from it. If you chose a different HS the second time, ***you will receive £2 of the £8 initially allocated to you***.

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

Example:

Randomly selected menu with three headsets

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



Experiment 2: Forced Choice Treatment Instructions

General procedure

This experiment aims to study people's choice behaviour.

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

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

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

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

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

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

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

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

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

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

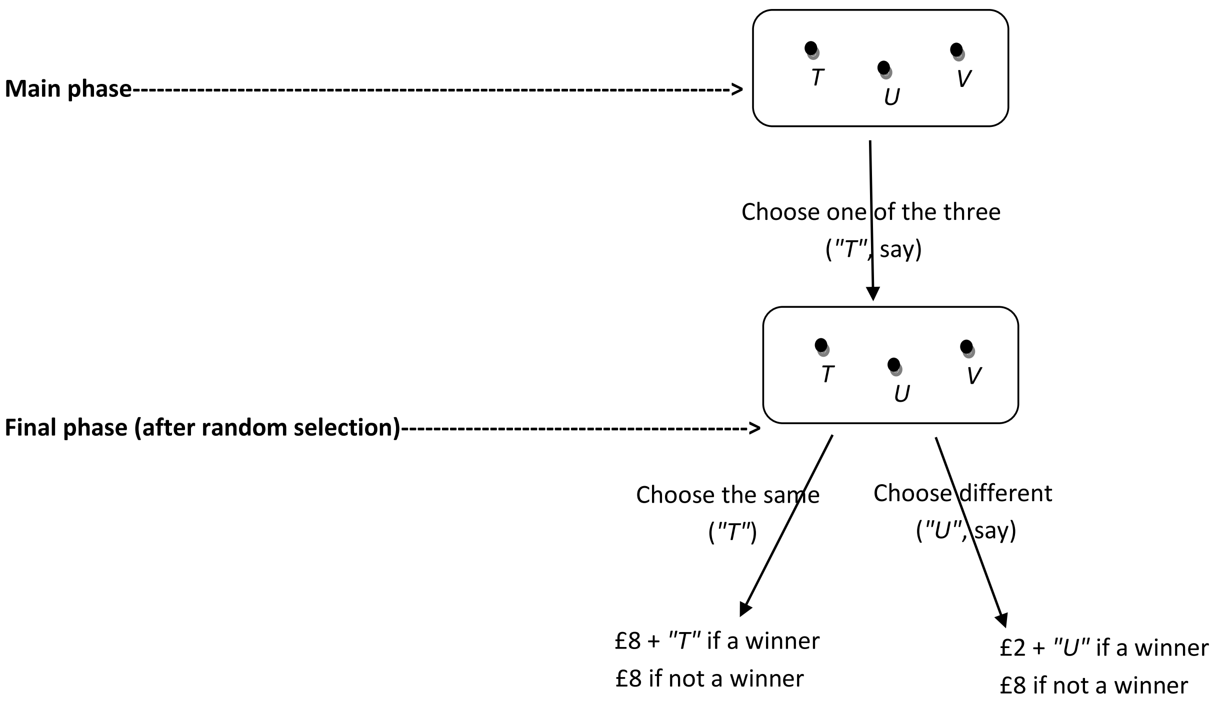
A) Suppose that when you first saw menu R you had chosen some HS from it. If you chose the same HS the second time, ***you will receive the £8 initially allocated to you***.

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

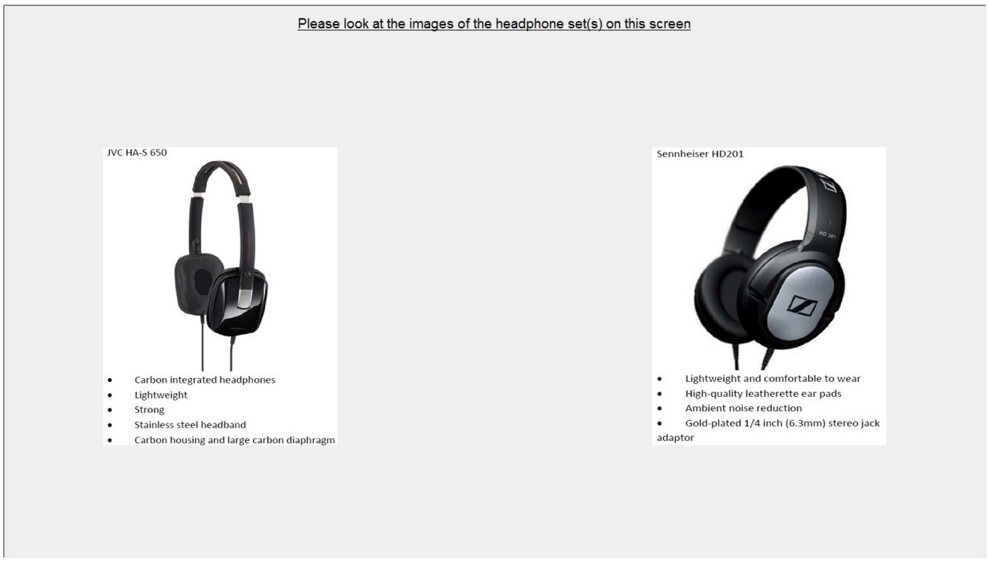
Example:

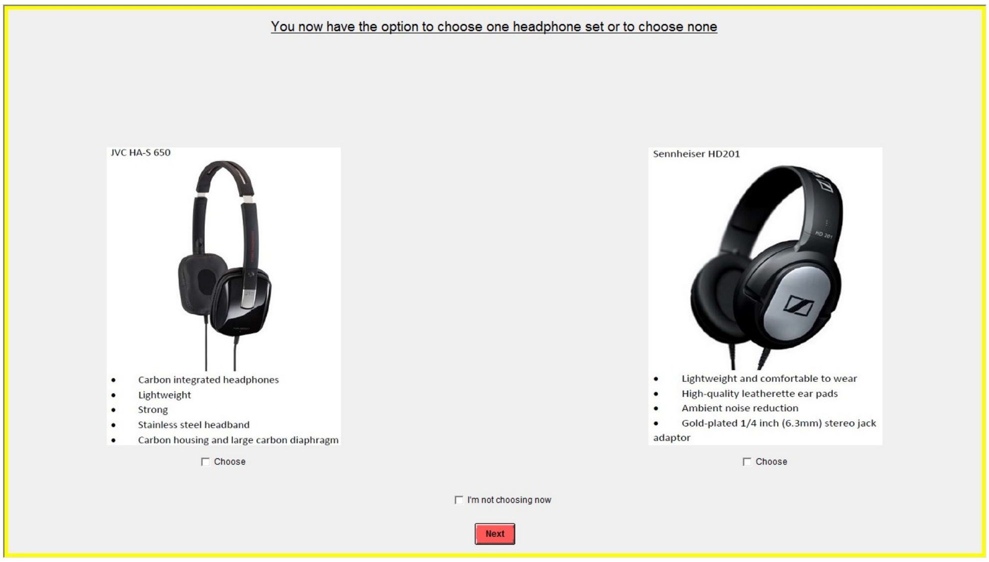
Randomly selected menu with three headsets

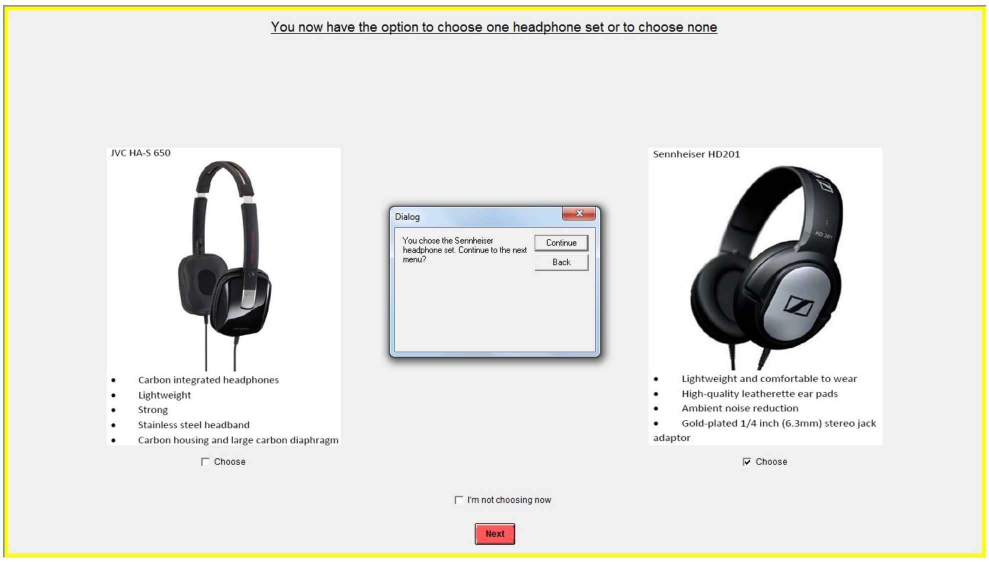
(the stated amounts assume that all quiz questions were answered correctly by the 3rd attempt)



Experiments 1 & 2 – Sample Screenshots







Non-Parametric Model-Based Analysis of All 26 Decisions

Table 4: Average HM distance scores for the models of rational choice and dominant choice with incomplete preferences over the combined active-choice and deferral decisions (ranges, medians and proportions of zero scores in parenthesis).

	Rational Choice	Dominant Choice with Incomplete Preferences	Best of Two Models
Experiment 1: FC (#75)	0.80 (0-5; 0; 55%)	1.80 (1-6; 1; 0%)	0.80 (0-5; 0; 55%)
Experiment 1: NFC (#86)	4.36 (0-26; 1; 44%)	1.86 (0-7; 1; 10%)	1.12 (0-7, 0; 54%)
Experiment 2: FC (#53)	0.98 (0-5; 0; 60%)	1.98 (1-6; 1; 0%)	0.98 (0-5; 0; 60%)
Experiment 2: NFC (#68)	5.46 (0-26; 1.5; 43%)	2.18 (0-10; 1; 10%)	1.53 (0-10; 0; 53%)
Pooled: FC (#128)	0.88 (0-5; 0; 57%)	1.88 (1-6; 1; 0%)	0.88 (0-5; 0; 57%)
Pooled: NFC (#154)	4.84 (0-26; 1; 44%)	2.00 (0-10; 1; 10%)	1.53 (0-10; 0; 54%)

Table 5: Proportions of subjects whose overall behavior is best explained by the two models (proportion of perfect fits out of all relevant subjects in parenthesis).

	Rational Choice	Dominant Choice with Incomplete Preferences
Experiment 1: FC (#75)	100% (55%)	0%
Experiment 1: NFC (#86)	76% (44%)	24% (11%)
Experiment 1: NFC-deferring (#26)	19% (0%)	81% (35%)
Experiment 2: FC (#53)	100% (60%)	0%
Experiment 2: NFC (#68)	66% (43%)	34% (10%)
Experiment 2: NFC-deferring (#28)	18% (0%)	82% (25%)
Pooled: FC (#128)	100% (57%)	0%
Pooled: NFC (#154)	71% (44%)	29% (10%)
Pooled: NFC-deferring (#54)	19% (0%)	81% (30%)

Notes: ties (one subject in each of Exp1 and Exp2) are broken in favor of rational choice.

Figure 3: An NFC subject's behavior that is explainable by dominant choice with incomplete preferences, with the partial preference ordering that is uniquely recoverable from that behavior shown in Figure 2.

$$
\begin{array}{lllll}
C(v)=v, & C(w)=w, & C(x)=x, & C(y)=y, & C(z)=z, \\
C(v,w)=v, & C(v,x)=\emptyset, & C(v,y)=v, & C(v,z)=\emptyset, & C(w,x)=\emptyset, \\
C(w,y)=w, & C(w,z)=z, & C(x,y)=x, & C(x,z)=\emptyset, & C(y,z)=z, \\
C(v,w,x)=\emptyset, & C(v,w,y)=v, & C(v,w,z)=\emptyset, & C(v,x,y)=\emptyset, & C(v,x,z)=\emptyset, \\
C(v,y,z)=\emptyset, & C(w,x,y)=x, & C(w,x,z)=\emptyset, & C(w,y,z)=z, & C(x,y,z)=\emptyset, \\
C(v,w,x,y)=\emptyset, & C(v,w,x,z)=\emptyset, & C(v,w,y,z)=\emptyset, & C(w,x,y,z)=\emptyset, & \\
C(v,w,x,y,z)=\emptyset & & & &
\end{array}
$$

Consistency Results without Noisiness/Singleton-Deferral Exclusions

Table 6: Proportions of subjects with zero active-choice cycles (p-values from two-sided Fisher exact tests).

	WARP/Congruence		
	Exp1	Exp2	Pooled
FC	54% (41/76)	59% (32/54)	56% (73/130)
NFC	70% (104/149)	74% (50/68)	71% (154/217)
p-value	0.018	0.121	0.005
N	225	122	347

Table 7: Mean Houtman-Maks and Swaps mean scores on active choices (number of subjects in parenthesis; p-values from two-sided Mann-Whitney U tests).

	Houtman-Maks			**Swaps**		
	Exp1	Exp2	Pooled	Exp1	Exp2	Pooled
FC	0.89 (76)	1.13 (54)	0.99 (130)	0.99	1.24	1.09
NFC	0.54 (149)	0.75 (64)	0.60 (213)	0.57	0.86	0.66
p-value	0.019	0.148	0.006	0.023	0.144	0.006
N	225	118	343	225	118	343