Games and Economic Behavior 158 (2026) 737–762

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

Eliciting and distinguishing between weak and incomplete preferences: Theory, experiment and computation

Georgios Gerasimou

Adam Smith Business School, University of Glasgow, Scotland, United Kingdom

ARTICLE INFO

Keywords:
Weak preferences
Incomplete preferences
Revealed indifference & indecisiveness
Choice correspondences
Jaccard-Houtman-Maks method
Choice deferral
Lab experiment

ABSTRACT

We propose an experimental design and data-analytic method for eliciting and distinguishing the strict-preference, indifference, and indecisiveness components of individual preferences in general choice environments. The design combines a forced-choice treatment with a free-choice treatment. In both treatments, subjects may select multiple alternatives from a menu. In the free-choice treatment, subjects may also avoid or delay choice at a small expected cost. To analyze such data, we extend a standard non-parametric goodness-of-fit criterion to accommodate multi-valued choices. We apply it to evaluate the consistency of subjects' 50 decisions with utility maximization and two models of incomplete-preference maximization. Around 55% of subjects are well explained by one of these models, with 33% and 22% best explained by utility and incomplete-preference maximization, respectively. Revealed preferences typically feature non-trivial indifferences, and those that are incomplete often exhibit the predicted theoretical distinctions between indifference and indecisiveness, which are documented empirically for the first time.

1. Introduction

The possibility in which a decision maker is indifferent between two or more choice alternatives plays a prominent role in many domains of economic analysis. It is therefore important to understand what kinds of observable decision environments and data-analytic methods can in principle allow for extracting an individual's potentially weak preferences, and distinguishing between their strict-preference and indifference parts. This is particularly pertinent if one also accepts that the individual in question may be indecisive, with preferences that are stable and transitive but potentially incomplete. In this situation, it is not obvious what kinds of observable behavioral data could be used—and how—in order to separate indifference from incomparability/indecisiveness.

These problems are relevant both from a theoretical and a practical perspective. Since many economic models of individual, collective or strategic decisions assume that agents have preferences with non-degenerate indifference parts, eliciting the agents' weak preference relations would allow for testing those models' descriptive relevance more accurately than if indifferences were assumed away. Additionally, knowing, for example, how many decision makers in a community consider a tree-planting program to be equally good to the development of a playground, and how many of them have a strict preference instead, would generally lead to a better collective decision than if all preferences were mistakenly interpreted to be strict. In the example summarized in Table 1, this is illustrated with Basel and Cora's indifference between the two options, and Anna's strict preference for tree-planting. The latter option is the majority-rule outcome associated with the three agents' preferences, and is indeed obtained when multiple choices are allowed. When single choices must be made, however, it is possible that majority rule will lead to the playground outcome instead. Similarly, discerning when agents are indifferent or indecisive can help planners understand if they should immediately make

E-mail address: Georgios.Gerasimou@glasgow.ac.uk

https://doi.org/10.1016/j.geb.2026.05.011
Received 12 September 2025
Available online 26 May 2026

Table 1
Hypothetical example where ignoring potential indifferences by forcing single choices could lead to inefficient social decisions.

	Anna	Basel	Cora	Majority-rule outcome
True preferences	$TreePlanting \succ_A PlayGround$	$TreePlanting \sim_B PlayGround$	$TreePlanting \sim_C PlayGround$	$TreePlanting$
Possible revealed preferences when single choices allowed	$TreePlanting \succ^s_A PlayGround$	$PlayGround \succ^s_B TreePlanting$	$PlayGround \succ^s_C TreePlanting$	$PlayGround$
Revealed preferences when multiple choices allowed	$TreePlanting \succ^m_A PlayGround$	$PlayGround \sim^m_B TreePlanting$	$PlayGround \sim^m_C TreePlanting$	$TreePlanting$

an active choice that will impact everyone in a group, or delay such a choice and, for example, help group members resolve their indecisiveness – e.g. via targeted information campaigns.

This paper attempts to elicit human decision makers' potentially weak or incomplete preferences and, where relevant, separate their revealed strict preference, indifference and indecisiveness components. It contributes in this direction with a combination of methodological and empirical innovations that bring together, and extend, recent developments in choice theory, computational revealed preference analysis and experimental economics. In the next few paragraphs we provide an overview of these innovations.

First, the paper reports on a new lab experiment on riskless choice. This is based on a between-subjects design that elicits observable behavioral data that could be helpful toward recovering an individual's possibly weak and incomplete preferences. The design features a standard Forced-Choice treatment as well as a Free-Choice treatment. In both treatments, subjects can make multiple choices from every menu of alternatives they see during the main part of the experiment. Subjects in the Free-Choice treatment can, in addition, incur a small expected cost to avoid or delay choosing from a menu, at no informational benefit (henceforth to *defer* choice, possibly indefinitely). On the other hand, subjects in the otherwise standard Forced-Choice treatment must always choose some alternative(s).

One of the menus that were shown during the main part of the experiment was drawn randomly at the end. If a subject had chosen all alternatives at that menu previously, one of these was selected at random. The subject was offered this alternative as an in-kind reward, and also received a payment in cash. This coincided with subjects' initial monetary endowment, and represented the maximum possible cash reward. By contrast, if a subject had not chosen all alternatives at that menu previously, they were asked to choose one then. If this final choice was among those they had previously made at that menu, then they were offered their chosen alternative and received the maximum cash payment. On the other hand, if their final choice was not among the choices they had made at that menu previously, then their cash payment was markedly lower. Finally, if the subject deferred at that menu originally, then their cash payment was lower than the maximum possible one, but higher than what it would have been in the above choice-reversal scenario. A utility-maximizing agent who makes decisions in this environment, and who has no preference for or against randomization, is strictly incentivized never to choose a preference-dominated alternative, as this could result in a sub-optimal combined reward. In addition, such an agent is weakly incentivized to choose all feasible alternatives that they consider to be best at a menu, as doing so can never lead to a sub-optimal combined reward. Finally, such an agent will never defer, because doing so could lead to a sub-optimal cash reward.

Given that deferring does not come with any informational benefits in this experiment, however, a non-utility maximizing subject may be willing to incur the deferral cost in the Free-Choice treatment in situations where solving the decision problem—i.e. contemplating until a best option is determined—may be even costlier. In addition, and independently, the possibly multiple choices per menu that are made by a non-utility maximizing subject will not generally reveal a stable, complete and transitive ordering. For this reason, even though the design's incentives are more closely aligned with testing the benchmark model of utility maximization under weak or strict preferences, we also consider two models of stable and transitive but *incomplete* preference maximization: (i) *undominated* choice, whereby an individual is described as choosing the feasible alternative(s) that are not worse than anything else; (ii) *dominant* choice, whereby the agent chooses the most preferred feasible alternative(s) if and only if those exist, and defers otherwise.

In this connection, the paper's second methodological contribution consists of analyzing the experimental data by applying a novel computational method. The method is based on brute-force discrete optimization and allows for a model-rich approach towards recovering an individual's potentially weak and/or incomplete preferences. It extends in the direction of multi-valued choice—and in a model-based way—the classic (Houtman and Maks, 1985) method that is routinely used in empirical revealed preference tests of utility maximization. The original method computes the maximal subset of a subject's dataset that is consistent with rational choice. This allows, essentially, for an intuitive quantification of the subject's behavioral proximity with that model. The novel extension that we propose and utilize—which we refer to as the *Jaccard*-Houtman-Maks method due to its leveraging the classic Jaccard dissimilarity metric—accounts for the generally multi-valued nature of choice data in these computations. We motivate and introduce this extension in Section 2.3, and in Section 5 apply it simultaneously to the model of rational choice and the two models of incomplete-preference maximization outlined above.

Our experiment was conducted in a laboratory environment. There were 138 and 135 evaluable participants in the Forced- and Free-Choice treatment, respectively. Each was presented with 50 distinct menus (decision problems). These were drawn from an underlying set of 6 riskless goods. A good was a bundle of two different £10 gift cards that were issued by popular retail brand names in the grocery, coffee, bookshop and restaurant sectors. Every gift-card pair had a total value of £20. The 50 decision problems consisted of all menus with 2, 3 and 4 gift-card pairs, and were presented in subject-specific random orders.

Despite the relatively large number of decisions, more than 55% of all subjects are well-approximated by one of the above three models of preference maximization, in the sense that they are no more than 10% away from being explained by them perfectly. Furthermore, and perhaps surprisingly, the model-optimally recovered preferences of 79% of all subjects in this group feature at least one indifference comparison between distinct alternatives and, on average, almost 2.5. This represents 15% of all possible binary comparisons between the 6 distinct choice alternatives in our experiment. The relatively high such frequency of indifferences that is detected by this theory-guided method in our discrete choice environment is important and, in our view, merits further exploration. As far as the decomposition of preference-maximization types is concerned, 33% and 22% of all subjects are best matched by utility maximization and the two models of incomplete-preference maximization, respectively. In addition, a *single* optimal preference relation is recovered from 81% of these subjects. This showcases the method's potential to deliver sharp preference identification when applied on relatively rich—even if incomplete—data sets using models that are in principle uniquely identifiable.

Importantly, the preferences elicited from nearly two thirds of all incomplete-preference maximizers document empirically—and for the first time—the distinct theoretical separations between *revealed indifference and revealed indecisiveness* that are afforded by the two relevant models. Both these separations assume the analyst has access to the kind of multi-valued choice data that are elicited in our experiment. Their documentation is one of the paper's main empirical contributions, as it demonstrates the relevance of theories of incomplete weak preferences; confirms that their predictions can indeed be tested by observable choices; and suggests a way that such a test could be done.

Another empirical contribution, derived from analyzing data at the aggregate level, is the documentation of a significant negative correlation between subjects' choice consistency and their average response times. Although seemingly counter-intuitive, and despite being derived from collections that comprise both binary and larger menus, we argue that this finding is broadly in line with a key prediction of sequential learning models such as the influential *drift-diffusion* one that pertains to binary forced choices (Section 4). The prediction might be interpreted as suggesting that shorter response times are more likely when the individual has a clear preference between the feasible options. Such clarity in turn might translate into more consistent active choices, which is indeed what we find.

Furthermore, although subjects in our experiment were allowed to—and typically did—make multiple choices, those in the Free-Choice treatment behaved significantly more consistently. This provides a positive robustness check of the main finding in Costa-Gomes, Cueva, Gerasimou and Tejiščák (2022) —henceforth frequently referred to as CCGT22. Unlike the present work, that study—which we recall in detail below—was predominantly designed to test for such potential differences in consistency, and did so by analyzing single-valued choice data.[1] It might have been hypothesized ex ante that the additional freedom of subjects—particularly in the Forced-Choice treatment—to select more than one alternative could alleviate such differences in consistency. Our results suggest otherwise.

1.1. Related literature

The core of this paper's experimental design extends in the multi-valued choice direction the design of CCGT22. As already mentioned, that study also reported on forced- and free-choice treatments, but elicited single-valued choices. This made it impossible to raise the questions that the present paper is mainly concerned with. Although a pre-publication version of that paper (Costa-Gomes et al., 2016) had proposed a method of multi-valued choice construction, that method relied on survey data obtained after choices from binary menus. As such, it was considerably more restrictive than the one advanced here. Furthermore, in distinct work following that study, but predating the present one, Bouacida (2021) proposed a design to elicit unrestricted multi-valued choices directly, paying subjects a fixed extra amount for every additional alternative that they chose from a menu, and assigning them one randomly as a reward. By contrast, the design in the present paper penalizes inconsistent choices towards incentivizing subjects to reveal their true preferences.

Regarding the pre-existing literature on the experimental elicitation of incomplete preferences, we highlight that this has mainly revolved around choices from binary menus of lotteries or uncertain acts, with designs that explicitly utilize this fact. That literature is reviewed in CCGT22. More summarily here: to identify incompleteness in such environments, those studies have used choice deferral alongside preference-for-flexibility models (Danan and Ziegelmeyer, 2006); partially incentivized methods of imprecise-preference revelation from menu lists (Cubitt, Navarro-Martinez and Starmer (2015) and references therein); and incentivized methods where incompleteness might be thought of as being revealed via certain patterns of randomized choice (Cettolin and Riedl, 2019; Agranov and Ortoleva, 2025). In general choice environments on the other hand, evidence for *strict* incomplete preferences was documented

[1] Aside from the significance of such findings for revealed preference analysis, there is at least one real-world parallel with large-stakes decisions: high-court judicial decision making. Indeed, justices in such courts often have the luxury of choosing which cases to hear—and make a decision on—and which to ignore. As Hitt (2019) notes for the case of the US Supreme Court (pp. 6–7), *"the Court could protect the logical consistency and quality of its opinions by ignoring complex and multifaceted cases. […] Essentially, [the Supreme Court Case Selections Act of 1988] gave the justices almost total freedom to opt* not *to decide most disputes. As such, whether consciously or not, the modern Supreme Court actively evolved away from decisiveness".* The author further argued that such potentially very impactful choices may have traceable and behaviorally intuitive underlying patterns, noting that *"prioritizing consistency means that the Court will leave numerous important questions and conflicts unresolved"* because *"a desire to produce 'good' (consistent) law may induce the justices to prioritize consistency over decisiveness."*

in the free-choice treatment of CCGT22 via costly choice deferrals and a restricted application of the Houtman-Maks method on the dominant-choice model in Gerasimou (2018a, Section 2) in which the decision maker's incomplete preferences were assumed to be strict. The present paper improves upon CCGT22 by enabling tests of the hypothesis of incomplete-preference maximization in the most general choice environments possible.

2. Theoretical background

2.1. Three deterministic models of preference-maximizing choice

In the main part of our individual-level analysis, presented in Section 5, we consider the following three general choice models of deterministic preference maximization:

I. Rational Choice/Utility Maximization.
II. Undominated Choice with Incomplete Preferences.
III. Dominant Choice with Incomplete Preferences.

We focus on these models for the following reasons:

1. They feature stable preferences and predict both single-valued and/or multi-valued choices under different "instances".[2] Under all three models, moreover, a predicted single choice at a menu means that the alternative in question is strictly preferred to all other feasible ones. Conversely, multiplicity of the set of predicted optimal choices means that the decision maker is indifferent between the goods in question (I, III) or, more generally, unable to compare them by strict preference (I–III).
2. They impose strong behavioral restrictions. All models satisfy the fundamental *Property α* or *Contraction Consistency* principle (Sen, 1971, 1997). This requires an option that is chosen at some menu to also be chosen at every submenu where it remains feasible. In addition, I and III predict active choices that are consistent with the Strong Axiom of Revealed Preference.
3. They are defined in terms of—and hence in principle allow for recovering—a single preference relation. This makes the welfare-relevant part of their analysis unambiguous.
4. They are uniquely identifiable. That is, if a decision maker's observable behavior is perfectly compatible with a model, and data from sufficiently many choice problems are available, then there is a unique preference relation that explains the individual's behavior under this model.
5. They are sufficiently tractable computationally to allow for the behaviorally intuitive optimization-based goodness-of-fit test that we describe below.
6. Under the assumptions laid out in our discussion in Section 3.5 concerning the structure and interpretation of the experimental design, the models predict the kinds of active-choice or deferring behaviors that one might expect to observe in our data.

To state the models formally we first define a decision maker's choice dataset $\mathcal{D} = \left(A_i, C(A_i)\right)_{i=1}^{k}$ on a finite grand choice set X to be a collection of pairs that comprise a non-empty menu $A_i \subseteq X$ and a—possibly empty—set of alternatives that were chosen at this menu when the decision maker was presented with it. That is, $\emptyset \subseteq C(A_i) \subseteq A_i$ holds for all $i \leq k$. Dataset $\mathcal{D}$ is explainable by *Rational Choice/Utility Maximization* if there exists a complete and transitive preference relation $\succsim$ on X such that, for all $i \leq k$,

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

If, instead, (1) is true for all $i \leq k$ with respect to a reflexive and transitive but incomplete preference relation, then $\mathcal{D}$ is explainable by the model of (maximally) *Dominant Choice with Incomplete Preferences*. In that case we have

$$\begin{aligned} C(A) \neq \emptyset &\iff \text{there is } x \in A \text{ such that } x \succsim y \text{ for all } y \in A \\ C(A) = \emptyset &\iff \text{for all } x \in A \text{ there is } y \in A \text{ such that } x \not\succsim y \end{aligned}$$

Finally, $\mathcal{D}$ is explainable by the model of *Undominated Choice with Incomplete Preferences* (also known as *maximal-element* choice) if there is a reflexive, transitive and incomplete preference relation $\succsim$ whose asymmetric part is $\succ$, such that, for all $i \leq k$,

$$C(A_i) = \{x \in A_i : y \not\succ x \text{ for all } y \in A_i\} \tag{2}$$

Richter (1966) characterized the rational choice model in a general environment which encompasses that of the present paper. The model of undominated choice with incomplete preferences was, to the best of our knowledge, introduced by Schmeidler (1969) in a general equilibrium setting. It has been studied choice-theoretically under a variety of decision environments and preference structures by, most notably, Sen (1971), Schwartz (1976), Bossert, Sprumont and Suzumura (2005), Eliaz and Ok (2006),

[2] Throughout, by a model's "instance" we mean either a specific weak order (cf. Rational Choice/Utility Maximization) or a specific incomplete preorder (cf. Undominated or Dominant Choice with Incomplete Preferences) that may explain some data when it is maximized in the way that is dictated by that model.

Table 2
Enumeration of incomplete preorders that allow for non-trivial indifferences.

$|X|$	All indifference-permitting incomplete preorders	Regular indifference-permitting incomplete preorders	%
3	3	0	0.00%
4	85	54	63.53%
5	2290	1705	74.45%
6	75,541	60,455	80.03%
7	3,363,129	2,799,615	83.24%

Bossert and Suzumura (2010) and Stoye (2015).[3] Dominant choice with incomplete preferences was introduced and studied in Gerasimou (2018a, Section 2).[4]

The two models of incomplete-preference maximization are logically distinct. Moreover, if we replace the term "incomplete" with "possibly incomplete" in their respective statements, these models generalize rational choice in different ways. The first does so by relaxing active-choice consistency while retaining the decisiveness (non-emptiness) assumption that requires $C(A_i) \neq \emptyset$ for all $i \leq k$. The second model achieves this by relaxing the decisiveness assumption while retaining active-choice consistency.

2.2. Distinguishing between indifference and incomparability/indecisiveness

For a decision maker with incomplete preferences who is also indifferent between some alternatives, a natural but non-trivial question is how one might use observable behavioral data in conjunction with some model in order to separate those pairs of alternatives between which the agent is indifferent from those where the agent is indecisive. Eliaz and Ok (2006) were the first to raise and provide an answer to this question. Taking the undominated-choice model as their primitive, the authors focused on the special case where the model's rationalizing incomplete preference relation $\succsim$ is "regular". This refers to the property that, whenever $x \not\succsim y$ and $y \not\succsim x$ are both true, then there is $z \in X$ such that either $x \not\succsim z$, $z \not\succsim x$ and $y \succ z$ or $z \succ y$ holds, or $y \not\succsim z$, $z \not\succsim y$ and $x \succ z$ or $z \succ x$ holds.

Given their important role in this theoretical distinction, before proceeding we wish to understand how big—in absolute and relative terms—is the collection of regular incomplete preorders that, in addition, allow for non-trivial indifferences. To this end, Table 2[5,6] presents novel enumeration results that clarify how the number of incomplete preorders that are both regular and indifference-permitting evolves when X contains anything between 3 and 7 items, inclusive. Our main finding here is that their number, relative to those preorders with the same properties that are not necessarily regular, is increasing at a decreasing rate in the cardinality of X, and amounts to 80% when $|X| = 6$ –the case of special interest here.

With this definition and clarifications in place, we can proceed with a summary of the distinction proposed in Eliaz and Ok (2006):

> An agent whose incomplete preferences are captured by a regular preorder and who maximizes these preferences according to the undominated-choice model is revealed to be:
>
> *indifferent* between x and y ***only if*** $\big(x, y \in A \;\&\; y \in C(A)\big) \Rightarrow x \in C(A)$;
>
> *indecisive* between x and y ***only if*** $x \in C(A)$, $y \in A \setminus C(A)$, $y \in C(B)$ and $x \in B$ for distinct menus A and B.

In words, the agent is indifferent only if the two options are either chosen or rejected together when both are feasible, and indecisive only if one is chosen over the other in some menu and the latter is chosen in the presence of the former in another menu. Importantly, the revealed indifference relation here is transitive, whereas the revealed indecisiveness one is not (see also Mandler, 2009). Also importantly, this intuitive choice-reversal based distinction between the two notions is not robust: any behavior that is compatible

[3] This model is implicitly also used in stochastic-dominance applications of portfolio efficiency, along the lines studied in Levy (2016) and Arvanitis et al. (2023), for example. Bouacida (2021) tests the no-indifference version of this model with the data collected under that study's experimental design.

[4] A related pre-dating study is Dean (2008), which proposed a class of *decision-avoidance* models that focused on explaining the increasing prevalence of *status quo bias*—a phenomenon that is distinct from *choice deferral*—as menu size increases. When the decision problem contains no natural status quo and no dominant alternative, these models predict a compensatory decision process via which (generally inconsistent) active choices are always made. Dominant choice with incomplete preferences on the other hand focuses on decision problems without a natural status quo option and features a non-compensatory decision process whereby the agent makes (consistent) active choices when and only when a most preferred alternative exists.

[5] This output was produced from constraint satisfaction problems written in the Essence′ language and solved by the open-source "Minion" solver (Gent et al., 2006) using the "Savile Row" constraint modelling assistant (Nightingale et al., 2014).

[6] For example, the set of all possible incomplete preorders with non-trivial indifferences on $X = \{x, y, z\}$ are: (1) $x \sim y$; $x \not\succsim z \not\succsim x$; $z \not\succsim y \not\succsim z$. (2) $x \sim z$; $x \not\succsim y \not\succsim x$; $z \not\succsim y \not\succsim z$. (3) $y \sim z$; $x \not\succsim y \not\succsim x$; $z \not\succsim x \not\succsim z$

with such an indifference-permitting instance of that model is observationally equivalent to behavior generated by the instance that is identical to it except that indifferences are replaced by incomparabilities (see also Theorem 1 in , Bossert et al., 2005 and Theorem 3.3 in Bossert and Suzumura, 2010).

More recently, in Gerasimou (2018a) this author noted a distinct and robust behavioral separation between indifference and indecisiveness that is afforded by the dominant-choice model:

> A decision maker whose incomplete preferences are captured by a preorder and who maximizes these preferences according to the dominant-choice model is revealed to be:
>
> *indifferent* between x and y ***if and only if*** $\big(x, y \in A \;\&\; y \in C(A)\big) \Rightarrow x \in C(A)$;
>
> *indecisive* between x and y ***if and only if*** $x, y \in A \Rightarrow x, y \notin C(A)$.

The agent here is indifferent iff both options are either chosen or rejected together when both are feasible, and indecisive iff neither is ever chosen in the presence of the other. The reason why this distinction is robust is that reducing a relation's indifferences to incomparabilities here leads to different active-choice and deferring behavior.

2.3. How close to a model is a choice correspondence? The Jaccard-Houtman-Maks index

Our non-parametric model-fitting analysis extends the classic Houtman and Maks (1985) (HM) method that was outlined in the Introduction by accounting for multi-valued—as well as empty-valued—choice data *and* the possibility that the agent generating such data behaves under a model that is more general than utility maximization. Specifically, by finding which model is closest to explaining each subject's behavior, it effectively allows for recovering—possibly approximately, and with the standard *as if* qualifications in place—both the individual's deterministic decision rule *and* their preferences conditional on that rule.

To account for the potential multiplicity of both the subjects' choices and the model-predicted optimal choices at some menu, our hereby proposed extension of HM is based on computing the *(dis-)similarity* between these sets. Intuitively, we wish to capture the proximity between observed and model-predicted choices at each menu because this would allow for a more nuanced understanding of when a deviation of a subject's set of chosen alternatives from the model-predicted ones is a relatively "severe" or "mild" mistake. This is important and stands in contrast to the binary "pass/fail" approach of the straightforward extension of the HM method in this environment.

While several dissimilarity measures have been introduced to study specific problems, here we adopt the classic *Jaccard metric* (Jaccard, 1901) between finite sets of discrete objects. This function, which is a proper metric (Marczewski and Steinhaus, 1958; Levandowsky and Winter, 1971), identifies the dissimilarity between two sets as the proportion of their non-common elements relative to the total number of distinct elements between them. Since its introduction, it has been widely applied in the life sciences and several other fields. An axiomatic characterization of this metric was recently proposed by this author in Gerasimou (2024).

Formally, recall that a dataset $\mathcal{D}$ consists of a finite collection of observations $\big(A, C(A)\big)$, where $\emptyset \neq A \subseteq X$ and $\emptyset \subseteq C(A) \subseteq A$. Denote by $\mathbf{D}$ the collection of all such datasets. In the important special case where $|C(A)| = 1$ for every pair $\big(A, C(A)\big)$ in $\mathcal{D} \in \mathbf{D}$ one can measure the proximity of $\mathcal{D}$ to the model of utility maximization with strict preferences via the function $\mathtt{HM} : \mathbf{D} \rightarrow \mathbb{N}_0$, due to Houtman and Maks (1985). This specifies

$$\mathtt{HM}(\mathcal{D}) \;:=\; \min_{\succ \in \mathcal{P}} \Big| \big(A, C(A)\big) \in \mathcal{D} \,:\, C(A) \neq C_{\succ}(A) \Big|, \tag{3}$$

where $\mathcal{P}$ is the set of strict linear orders on X and $C_{\succ}(A)$ the unique $\succ$-maximizer at menu A. That is, $\mathtt{HM}(\mathcal{D})$ is the smallest number of choices that need to be changed at (or dropped from) $\mathcal{D}$ in order for these data to be rationalizable by utility maximization under some strict preference relation.

Now suppose $|C(A)| \geq 1$ for every $\big(A, C(A)\big)$ in $\mathcal{D}$. Replacing in (3) the set $\mathcal{P}$ with the collection of *complete* preorders on X, which we denote by $\mathcal{R}$, and $C_{\succ}(A)$ with $C_{\succsim}(A)$ for any $\succsim$ in $\mathcal{R}$, provides a direct translation of the HM index with strict preferences to its weak-preference counterpart. There is, however, a conceptual subtlety that is lost in this translation: while there is a unique way in which a single-valued choice $C(A)$ could differ from the $\succ$-optimal choice $C_{\succ}(A)$ (i.e. either the two coincide or they do not), there are many ways in which a *multi-valued* choice $C(A)$ can differ from the $\succsim$-optimal choice $C_{\succsim}(A)$. We illustrate this with the following example that presents three hypothetical subjects' choices at the same menu:

$\succsim$:	$a \sim b \succ c \succ d$	
A	$=$	$\{a, b, c\}$
$C_{\succsim}(A)$	$=$	$\{a, b\}$
$C_1(A)$	$=$	$\{a\}$
$C_2(A)$	$=$	$\{c, d\}$
$C_3(A)$	$=$	$\{a, b\}$

Here, the $\succsim$-optimal choices at A are in the set $\{a, b\}$ and coincide with the choices made by subject 3. Those made by subjects 1 and 2 on the other hand both deviate from the optimal choice. However, there is a natural sense in which the former subject's choice is "closer" to being optimal than that of subject 2: it includes one of the two $\succsim$-best alternatives at the menu, whereas the choice of subject 2 contains neither. In the language of mistakes, one might label the first subject's as a "partially" mistaken decision, while that of the second subject as "fully" so. The straightforward extension of the HM index presented above does not account for this nuance and "penalizes" both $C_1(A)$ and $C_2(A)$ in the same way.

For this reason, instead of generalizing (3) to the multi-valued choice case in what appears to be a counter-intuitively punitive way, we propose a generalization that explicitly accounts for the *content-similarity* of sets $C(A)$ and $C_{\succsim}(A)$ across all pairs $\big(A, C(A)\big) \in \mathcal{D}$ and every weak preference relation $\succsim \in \mathcal{R}$. More specifically, and as already anticipated, to assess how dissimilar is the observed choice $C(A)$ to the choice $C_{\succsim}(A)$ that is optimal under some $\succsim \in \mathcal{R}$ we use the Jaccard dissimilarity metric, whereby

$$\begin{aligned} J\big(C(A), C_{\succsim}(A)\big) &:= \frac{|C(A) \cup C_{\succsim}(A)| - |C(A) \cap C_{\succsim}(A)|}{|C(A) \cup C_{\succsim}(A)|} \\ &= 1 - \frac{|C(A) \cap C_{\succsim}(A)|}{|C(A) \cup C_{\succsim}(A)|} \qquad (4) \\ &\in [0,1]. \qquad (5) \end{aligned}$$

With this definition in place, we can now formally introduce our proposed *Jaccard-Houtman-Maks* generalization of (3) for the case of possibly multi-valued choice data as the function $\mathtt{JHM} : \mathbf{D} \to \mathbb{Q}_+$ where

$$\mathtt{JHM}(\mathcal{D}) := \min_{\succsim \in \mathcal{R}} \sum_{(A, C(A)) \in \mathcal{D}} J\big(C(A), C_{\succsim}(A)\big) \qquad (6)$$

In words, JHM differs from the cruder multi-valued choice extension of HM in that it distinguishes between (and penalizes accordingly) "partially" and "fully" mistaken decisions at a menu according to the Jaccard similarity between each set of actual choices and the corresponding set of model-optimal choices at every menu. As such, one readily observes that $\mathtt{JHM}(\mathcal{D}) \leq \mathtt{HM}(\mathcal{D})$ for all $\mathcal{D} \in \mathbf{D}$, and $\mathtt{JHM} \equiv \mathtt{HM}$ in the subdomain of $\mathbf{D}$ where $|C(A)| = 1$ for all $\big(A, C(A)\big) \in \mathcal{D}$. Applied to the example given above, JHM assigns *distance scores*[7] (or simply JHM scores) $\frac{1}{2}$, 1 and 0 to the first, second and third subject, respectively, whereas the "binary" or "pass/fail" extension of HM assigns scores 1, 1 and 0 instead. The analysis of Section 4.3 builds on this index, while the results of Section 5 also rely on the natural adaptations of this index to the two models of choice with incomplete preferences that were discussed in Section 2.1.[8]

3. Experimental design

3.1. General remarks

The experiment was conducted between September 2019 and 2021 at the University of St Andrews Experimental Economics Lab with a total 282 subjects (more details in Table 3.3). The experimental goods were 6 distinct *pairs* of gift cards. Each individual gift card was worth £10; hence, the value of each experimental good in every menu was £20. All gift cards came from popular UK or international brands: two supermarkets, two coffee shops, one bookshop, and a gift card that enabled dining at one of nine restaurants. All 6 cards in these 6 pairs could be redeemed in at least one venue in the local town centre (subjects were explicitly informed about this), with the restaurant gift card being redeemable in 3 such venues. Each of the 6 gift cards appeared in exactly 2 of the 6 pairs, once showing up as the first item in the pair and once as the second (Fig. 1).

There are several reasons why the experimental choice alternatives were selected to be gift-card pairs, and those ones in particular. First, gift cards can be thought of as restricted forms of cash that can be used for consumption, informally traded or, indeed, gifted. As such, they are intrinsically valuable. Second, these particular gift cards were issued by some of the most popular leisure and grocery destinations for the local student population, and were all within a short walking distance from each other. Hence, all were expected to be desirable to everyone in this experiment's subject pool. Finally, presenting the choice alternatives as gift card *pairs* proxies a realistic situation (many online retailers invite consumers to choose between multi-store gift card bundles rather than individual-store gift cards) and could potentially lead to some relatively hard decisions.

Out of the 63 non-empty menus that are derivable from this set of 6 experimental goods, in the main part of either treatment subjects were presented sequentially with the 15, 20 and 15 menus with 2, 3 and 4 goods, respectively (see Fig. 2 for an example). Each of the 50 distinct menus was presented once. The order of menu presentation was randomized and differed between subjects. The experiment's choice domain is therefore relatively rich, and also heterogeneous and symmetric in the sense that all alternatives are feasible in exactly the same number of menus (see also Online Appendix D for more on this point).

The experiment's computer interface was programmed in Qualtrics and executed in full-screen mode on a web browser that prevented subjects from exiting the interface without the experimenter's intervention. Subject recruitment was done with ORSEE (Greiner, 2015). All menus appeared as unnumbered vertical choice lists. The *"I'm not choosing now"* option in the Free-Choice treatment was always the last item.[9] There were 141 participants in each of the two treatments. The 9 subjects who always deferred or always chose everything were excluded from the analysis because their choice behavior is completely uninformative. Every subject

[7] We use the term *"distance score"* rather than *"distance"* here partly because the underlying function (HM; JHM) is not a proper metric and partly to highlight the ranking aspect of the JHM-based goodness-of-fit approach when we apply it in Section 5.

[8] For completeness, Online Appendix B reports the main results of Section 5 when the "binary" HM extension is used instead.

[9] Because subjects often had to scroll down to find and select that option, this positioning meant that it was often physically harder for them to defer.

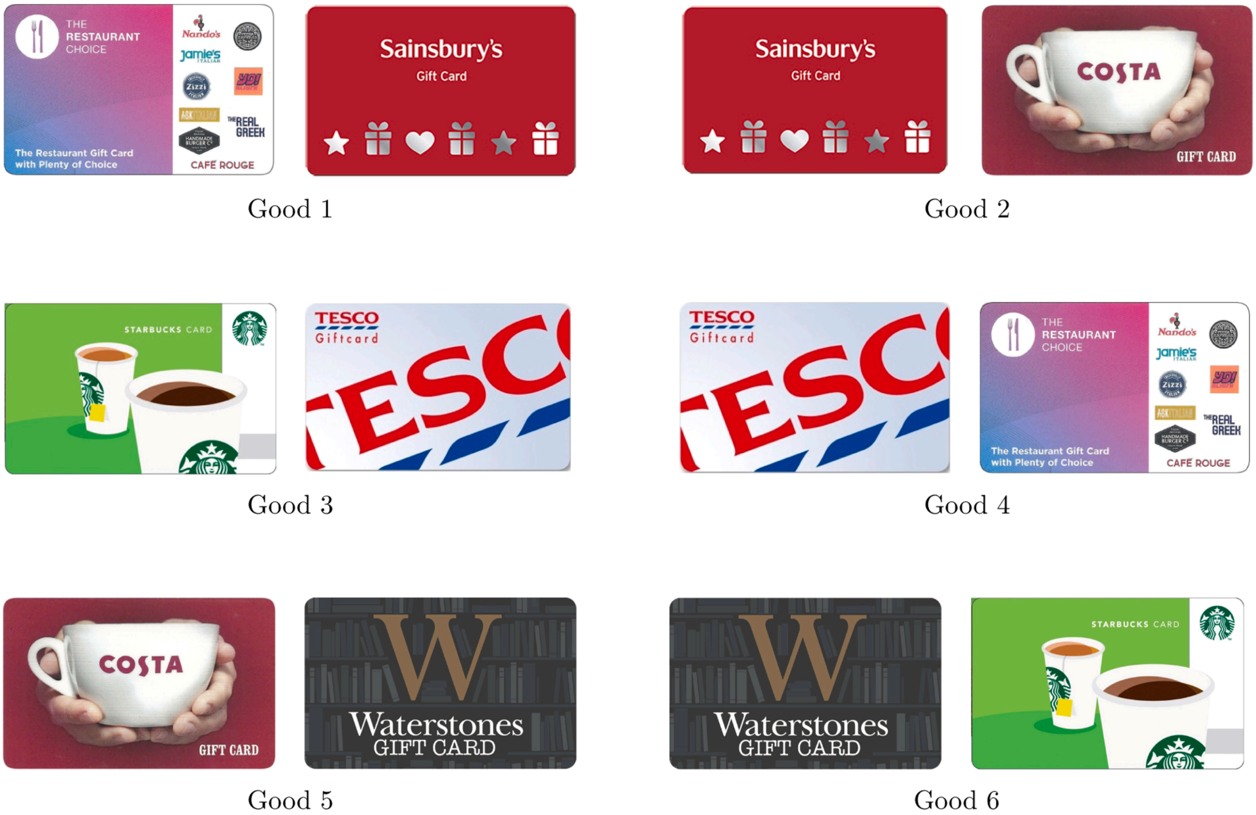

Fig. 1. Experimental goods: 6 pairs of gift cards (value in each pair: £20).

received both their gift-card and cash rewards, as explained in the next subsection. A nine-question understanding quiz preceded the main part of the experiment in each of the two treatments. Subjects could not proceed to the main part of the experiment until they answered all questions correctly.

3.2. *Forced-Choice treatment*

At the beginning of the experiment subjects in both treatments were allocated a monetary endowment of $I = £2.40$. When a menu was shown to subjects in the Forced-Choice treatment, they were asked to choose one or more items from that menu. Subjects knew that one menu would be picked at random for them at the end of the experiment. They also knew that they would be rewarded with an element of their randomly selected menu, and that the decision they made at that menu during the main part of the experiment would be reminded to them before they were asked to make their final, payoff-relevant decision there. Once subjects were past a menu during the main part of the experiment they never saw it again, unless that menu later turned out to be their randomly selected one. No additional information about the choice alternatives was provided at any point.

If a subject in this treatment chose one or more—but not all—goods from their randomly selected menu during the main part of the experiment, and also chose something from those previously selected options if that menu was their payoff-relevant one, then they received that item (in-kind reward) and I (cash reward). If, instead, at that point they chose something that was *not* among their previously chosen options at that menu, they received that item and $I_{rev} = £1.20 < I$. Finally, if they had chosen everything at that menu originally, then they received their original endowment, I, and a randomly selected element of that menu.

3.3. *Free-Choice treatment*

Free-Choice subjects were asked to either choose one or more of the alternatives at each menu or to avoid/delay making such an active choice by selecting *"I'm not choosing now"*. What happened here if a subject made one or more active choices from their randomly selected menu during the main part of the experiment coincided with the Forced-Choice treatment's provisions in the respective cases. If a Free-Choice subject had delayed choice at that menu originally, they were asked to choose an item then. In that case, they received this item and a cash amount $I_{def} = £2.10$. Importantly, I_{def} lied strictly between I and I_{rev}.

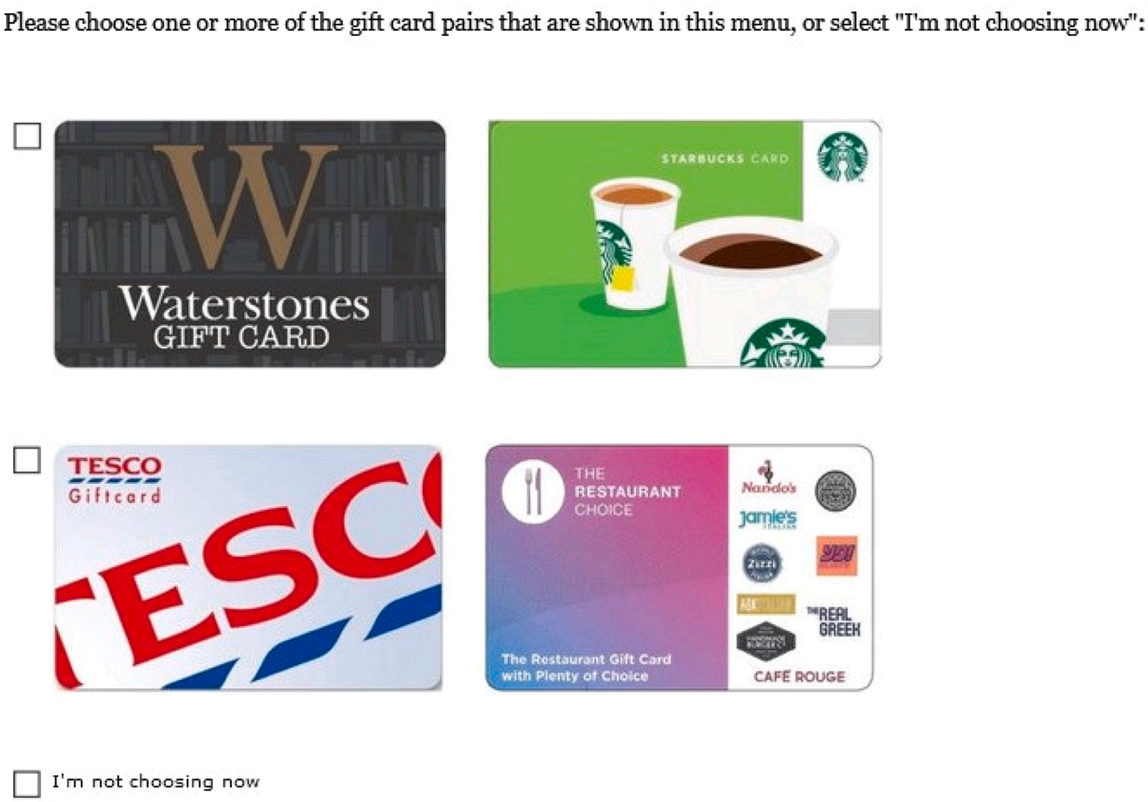

Fig. 2. Example decision problem shown in the main part of the Free-Choice treatment.

Table 3
Summary information on the two experimental treatments.

	Free-Choice treatment	Forced-Choice treatment
Original sample	141	141
Subjects excluded because they always chose everything	1	3
Subjects excluded because they always deferred	5	N/A
Subjects after all exclusions	135	138
Choice objects	6 pairs of £10 gift cards (£20 total value)	
Number of menus & decisions	50	
Reward frequency	Every subject	
Location	St Andrews lab	
Dates when the experimental sessions were conducted	17–20 Sept 2019	29 Jan 2020 (62) 22 Sep 2021 (79)

3.4. Payment

As soon as subjects finished all tasks, their randomly selected menu showed up on their screens, together with the reminder of the decision they had made at that menu. As an additional incentive for subjects to make deliberated and non-rushed decisions, they were told from the beginning that no participant would be able to receive their rewards and leave the lab in the first 50 minutes of the session. The experimenter (this author) went to each subject's desk once they were finished, and after this threshold was exceeded, asked them about their final choice from this menu, and later gave them their cash and gift-card rewards accordingly. Subjects who had previously chosen everything from that menu were invited to the experimenter's desk. After the appropriate numerical range

Table 4
Incentives in the Forced- and Free-Choice treatments.

1st decision (main stage)	2nd decision (payoff stage)	Choice reward	Cash reward	Possible interpretation
(a) Forced-Choice Treatment				
$C^1(A) = A$	none possible	random $a \in A$	initial endowment I	Revealed total indifference is costless & responded to literally
$C^1(A) \subset A$	$C^2(A) \in C^1(A)$	$C^2(A)$	I	Stable revealed preference is costless
$C^1(A) \subset A$	$C^2(A) \notin C^1(A)$	$C^2(A)$	$I_r < I$	Unstable revealed preference is costly
(b) Free-Choice Treatment				
$C^1(A) = A$	none possible	random $a \in A$	initial endowment I	Revealed total indifference is costless & responded to literally
$\emptyset \neq C^1(A) \neq A$	$C^2(A) \in C^1(A)$	$C^2(A)$	I	Stable revealed preference is costless
$\emptyset \neq C^1(A) \neq A$	$C^2(A) \notin C^1(A)$	$C^2(A)$	$I_r < I$	Unstable revealed preference is costly, and more so than
$C^1(A) = \emptyset$	$C^2(A) \in A$	$C^2(A)$	$I_d : I_r < I_d < I$	revealed indecisiveness

and numbers-to-goods assignment were specified on the random-number generating website https://random.org, a random number was generated to determine the good they would receive. The total payment per subject was approximately £22.2.

3.5. Incentives and identification

We denote by A a subject's payoff-relevant randomly selected menu and by $C^1(A)$, $C^2(A)$ the choices they made at that menu during the main and final parts, respectively. Using this notation, Table 4(a) summarizes the incentive structure in the Forced-Choice treatment.

By the very definition of indifference, a utility-maximizing Forced-Choice subject who is indifferent between all feasible alternatives is also indifferent between declaring choosable *all* these alternatives and *any* sub-collection thereof. Hence, such a subject has no *strict* incentive to either misreport their indifference or to report it truthfully. The converse is not true, however: they will declare everything to be choosable at a menu *only if* they are indifferent between all alternatives in it. This is so because, by indifference, that individual does not care which alternative they get, or whether this is decided by someone else or randomly (see also Danan (2010) for a formal elaboration of this point). Hence, it is *weakly dominant* for them to declare everything in a menu to be choosable whenever they are indifferent between all items contained in it.[10]

At the same time, if a utility-maximizing subject has one or more optimal alternative(s) at the menu, and these are strictly better than something else that's feasible, then choosing everything is dis-incentivized by the design because doing so comes with the risk of potentially receiving an inferior alternative as a reward. The design, instead, incentivizes such a subject to select those and only those options they prefer the most, as this enables them to choose one of these superior options at the end *and* receive their full cash endowment. For a utility-maximizing agent, therefore, the design in this treatment combines incentive-compatibility for truthful preference revelation in the sense of Azrieli, Chambers and Healy (2018) with a standard interpretation of multi-valued choice that is articulated in (Kreps, 2012, p.2) as follows: *"The story is that the consumer chooses one element of A. Nonetheless, we think of $c(A)$ as a subset of A, not a member or element of A. This allows for the possibility that the consumer is happy with any one of the several elements of A, in which case $c(A)$ lists all those elements. When she makes a definite choice of a single element, say x, out of A–when she says in effect, 'I want x and nothing else'–we write $c(A) = \{x\}$, or the singleton set consisting of the single element $\{x\}$. But if she says, 'I would be happy with either x or y', then $c(A) = \{x, y\}$."*

Turning to the Free-Choice treatment, we continue to denote by A a subject's randomly selected menu, and by $C^1(A)$, $C^2(A)$ the choices from that menu in main and final parts. Table 4(b) summarizes the incentive structure in that treatment. Here, $C^1(\cdot)$ is a possibly empty- and multi-valued choice correspondence: $\emptyset \subseteq C^1(S) \subseteq S$ is true for every menu S. As in the Forced-Choice

[10] This descriptive analysis and conclusion rest on the following additional assumptions:
(i) Subjects who *do* declare all feasible options to be choosable are not driven by a preference *for* randomization that may contravene—and prevail over—their preferences over these options. If they did, then we could not rule out that $a \succ b \succ c \succ d$ is true but $C(A) = A$ is observed at $A = \{a, b, c, d\}$, for example.
(ii) Subjects who *do not* declare all feasible options to be choosable are not driven by a preference *against* randomization that may contravene—and prevail over—their indifference between these options. If they did, then we could not rule out that $a \sim b \sim c \sim d$ is true but $C(A) = \{a\}$ is observed, for example.
That said, in Section 6.2 we report on the results of a method that we use in order to test for a systematic preference for randomization in the present environment.

treatment, however, $C^2(A)$ is a singleton. Moreover, for the reasons discussed in connection with the Forced-Choice treatment, a utility-maximizing Free-Choice subject is also (weakly) incentivized to choose their most preferred option(s) at every menu, with the opportunity of costly deferral being irrelevant to them.[11]

Now suppose instead that the subject is not a utility maximizer, but has incomplete preferences and cannot compare any of the alternatives at some menu. Then, depending on the individual's subjective perception of the decision's importance or cost, they may either opt to choose everything and end up with something at random, or, instead, incur the relatively small cost $I - I_{def}$ to delay making an active choice at that menu themselves. The former kind of behavior could be seen as analogous to recent findings suggesting that decision makers *prefer to randomize* when they are faced with a difficult decision repeatedly. We investigate this possibility in the context of the present experiment in Section 6.2. Importantly, though, such behavior could also be compatible with the model of Undominated Choice with Incomplete Preferences whenever every feasible option is incomparable to all others. By contrast, deferring at a small cost could be seen as a manifestation of indecisiveness-driven deferral (Tversky and Shafir, 1992; Danan and Ziegelmeyer, 2006; CCGT22) and may potentially be compatible with the model of Dominant Choice with Incomplete Preferences.[12]

We now proceed to a more formal behavioral identification analysis under this experimental design that reflects these ideas.[13] To this end, consider an agent who has a stable and transitive weak preference relation $\succsim$, complete or otherwise. Define and denote the sets of *dominant/greatest* and *undominated/maximal* elements of $\succsim$ at A by

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

Let 2^X be the power set of X, i.e. the collection of all subsets of X. We would like to be able to compare decision difficulty across menus and associate such difficulty with subjects' active-choice and deferring behavior. We do so by introducing a *decision cost function* $\phi : 2^X \to \mathbb{R}_+$. Given some menu A, $\phi_{\succsim}(A)$ captures the menu-specific decision cost associated with making an active choice at A. On the other hand, the menu-invariant and strictly positive cost of deferring—only relevant for subjects in the Free-Choice treatment—is captured by $\phi_{\succsim}(\emptyset)$. We further assume that, in decision problems such as those presented in this experiment, a decision is easy—hence costless—if and only if a dominant option exists. That is,

$$\begin{array}{lllll} \phi_{\succsim}(A) & = & 0 & \iff & B_{\succsim}(A) \neq \emptyset \\ \phi_{\succsim}(\emptyset) & \equiv & c_d & > & 0 \end{array}$$

Finally, we think of a subject's choice behavior at A once decision costs are also accounted for as being determined by the following rule:

$$C^1(A) = \begin{cases} B_{\succsim}(A), & \text{if } \phi_{\succsim}(A) = 0 \\ \emptyset, & \text{if } \phi_{\succsim}(A) > \phi_{\succsim}(\emptyset) \\ M_{\succsim}(A), & \text{if } 0 < \phi_{\succsim}(A) < \phi_{\succsim}(\emptyset) \end{cases} \tag{7}$$

In words: (i) the potentially multiple dominant alternatives are chosen whenever such alternatives exist (always true if $\succsim$ is complete); (ii) when no dominant alternative exists, the undominated alternatives are chosen if the decision cost associated with making an active choice is lower than the cost of deferring; otherwise choice is deferred.[14]

In the context of our experimental design it may be natural to interpret $\phi_{\succsim}(\emptyset) \equiv c_d$ in (7) as the expected cost of deferral, captured by the £0.30 deduction from subjects' initial pot when this is multiplied by the $\frac{1}{50}$ probability that this cost will actually be incurred after such a decision. For the value of $\phi_{\succsim}(A)$ on the other hand, recall first that, by assumption, making an active choice at A is costly if and only if A has no dominant alternative. When $\succsim$ is incomplete and $B_{\succsim}(A) = \emptyset$, $M_{\succsim}(A)$ consists of alternatives that are

[11] A comment may be due at this point on the discrepancy between how the design deals with situations where $C(A) = A$ or $\emptyset \neq C(A) \subset A$ at the payoff-relevant menu A. An alternative approach here would be to apply the uniform randomization rule at $C(A)$ in both situations. Introducing this dichotomy was driven by a motivation to have a design that remained comparable to the one in CCGT22. This goal is achieved: the two designs coincide in the special case where $C(A)$ is *required* to be single-valued. By contrast, an alternative design featuring a more extended application of uniform randomization would leave no role for the reduced cash payment I_{rev} that is associated with a choice reversal. The presence of this additional payment parameter, however, and its lower value compared to that of the full-payment parameter I, is potentially useful in instilling a mindset towards preference-guided choices by subjects. On the other hand, the unified choice-reward rule where randomization is applied on any non-empty $C(A)$ is immune to the possibility of subjects' choosing dominated alternatives, which in theory is left open by the dichotomous rule in the current design. While this theoretical possibility cannot be ruled out, it is worth keeping in mind that a subject who includes dominated alternatives in their first-round choice $C(A) \subset A$ not only does not benefit but, in fact, spends additional time and effort—thereby incurring a cost—in the process.

[12] We would like to draw an analogy here between the design's combination of costly deferral and costlier reversed active choice on the one hand, and some current thinking among AI specialists on how to deal with the persistent "hallucinations" problem of contemporary Large Language Models (LLMs) on the other. Specifically, in their recent working paper, Kalai et al. (2025) attribute the problem to the models' incentives in their training and evaluation stages, which *"reward guessing over acknowledging uncertainty"*. Motivated by prior literature and pre-existing test-grading systems that penalize incorrect answers to dis-incentivize guessing but do not penalize *"I don't know"*/abstaining answers, the authors propose that LLM evaluation instructions state *confidence targets* within user prompts. This paper's experimental design, as well as its predecessor (Costa-Gomes et al., 2016, 2022), also dis-incentivizes guessing in favor of choice deferral when *"I don't know"* best reflects a subject's honest response to the question *"Which alternative(s) would you choose from this menu?"*.

[13] Partly based on ideas from Section 4.1.2 in Costa-Gomes et al. (2016).

[14] Here too we note that $C^1(A)$ will not comprise alternatives that are dominated in A whenever $B_{\succsim}(A)$ is empty, as this would clearly be worse than choosing $C^1(A) = M_{\succsim}(A)$ (cf the choice-reversal cost of £1.20 to be incurred in this case if the agent does not wish to end up with an inferior option).

Table 5
Choice sizes and the corresponding relative frequencies and average response times at menus of different sizes.

	Alternatives chosen in the Forced-Choice treatment (average response times, in seconds, in parenthesis)				
	0	**1**	**2**	**3**	**4**
Menus with 2	–	83.77% (5.82)	16.23% (7.56)	–	–
Menus with 3	–	46.19% (7.31)	47.25% (8.44)	6.56% (8.88)	–
Menus with 4	–	28.50% (8.27)	41.01% (10.06)	27.29% (11.25)	3.18% (13.13)
	Alternatives chosen in the Free-Choice treatment (average response times, in seconds, in parenthesis)				
	0	**1**	**2**	**3**	**4**
Menus with 2	11.95% (6.19)	74.52% (6.01)	13.53% (7.93)	–	–
Menus with 3	6.85% (9.62)	44.67% (7.25)	43.33% (8.93)	5.15% (8.29)	–
Menus with 4	3.70% (12.71)	33.53% (8.98)	33.28% (9.93)	27.02% (11.24)	2.47% (12.92)

mutually incomparable by $\succsim$. Yet it is theoretically possible that the incomplete ordering $\succsim$ may become *less incomplete* in the interim period between the stage-1 and stage-2 decisions at A (for example, due to introspective learning).[15] Anticipating such a possibility, a forward-looking subject may associate the decision cost of choosing $M_{\succsim}(A)$ in stage 1 with the expected cost of arriving in stage 2 with a sufficiently more complete preference ordering $\succsim$ that will enable them to choose one of the $\succsim$-dominant alternatives at A. Therefore, the agent described in (7) decides between $C^1(A) = \emptyset$ and $C^1(A) = M_{\succsim}(A)$ by comparing the fixed expected cost of deferring and the menu-specific expected cost of attempting to complete their preferences until a dominant option emerges at A. This cost could be measured in units of cognitive effort or time spent on this task, and is increasing in the number of elements in $M_{\succsim}(A)$.

We conclude this analysis by observing that, for subjects in the Free-Choice treatment, (7) reduces to the model of: (i) Rational Choice/Utility Maximization if $\succsim$ is complete; (ii) Undominated Choice with Incomplete Preferences if $\succsim$ is incomplete and $\phi_{\succsim}(A) < \phi_{\succsim}(\emptyset)$ for every A; (iii) Dominant Choice with Incomplete Preferences if $\succsim$ is incomplete and $\phi_{\succsim}(A) > \phi_{\succsim}(\emptyset)$ for every A. In light of this, the reader may view the individual-level, model-based analysis of Section 5.2 as relying on this three-part assumption.

4. Aggregate-level analysis

4.1. Choice sizes

We begin our analysis by reporting on behavioral patterns at the aggregate level. To this end, we first introduce the *choice size* variable. This is simply the number of gift-card bundles that were chosen at each menu. Its values range between 0 and 4 in the Free-Choice treatment, and between 1 and 4 in the Forced-Choice treatment. However, because choice sizes 1 and 2—as well as 0 in the former treatment—are always feasible, whereas 3 and 4 are not always so, we adjust their relative frequencies accordingly. Specifically, Fig. 3 presents the distributions of menu-size adjusted choice sizes in the two treatments. These are derived once their absolute frequencies are divided by the total number of menus where these choice sizes might be observed. For example, the denominators here are $50 \times N$ and $15 \times N$ for choice sizes 1 and 4, respectively, where N is the number of subjects in the relevant treatment.

This comparison shows that the modal adjusted choice size was 1 in both treatments—at a rate of just over 50%—and was followed by sizes 2 and 3. Furthermore, the *deferral rate* in the Free-Choice treatment, defined as the relative frequency of a zero choice size, is 7.4%. This is slightly higher than the 6.9% *choose-everything* rate, defined as the relative frequency where n out of n items were chosen, for $n = 2, 3, 4$ (Table 5). The latter rate is also lower than the corresponding one in the Forced-Choice treatment (8.5%), and the difference is statistically significant ($p < .001$).[16] This difference is consistent with the intuition that, in the absence of the possibility to delay choice when faced with a potentially difficult decision, subjects are more likely to declare everything in the menu to be choosable. This, in turn, could be either a result of following a general decision rule or, in the context of our experimental design, possibly due to a preference for randomization (see Section 6.2). That said, the rate at which subjects chose all 4 alternatives, when possible, was very low in both treatments, at 3.2% and 2.5% in the Forced- and Free-Choice treatments, respectively.

Table 5 further shows how the relative frequencies of different choice sizes vary with the number of alternatives at different menus, and presents the corresponding average response times. These conditional relative frequencies are uniformly higher in the Forced-Choice treatment for choice sizes 1 and 2 in menus with 2 and 3 items. In addition, proportionally more subjects in that treatment chose 2, 3 or 4 gift-card pairs in four-element menus too. As far as deferrals in the Free-Choice treatment are concerned, these were more likely in binary menus (≈ 12%) than in those with 3 (≈ 7%) or 4 (≈ 4%) alternatives.

15 Recall that no new information about the alternatives becomes available during this period.

16 Throughout, the reported p-values are from one of the following two two-sided tests: (i) Fisher's exact test for the difference between proportions; (ii) Mann-Whitney U test for the difference between distributions.

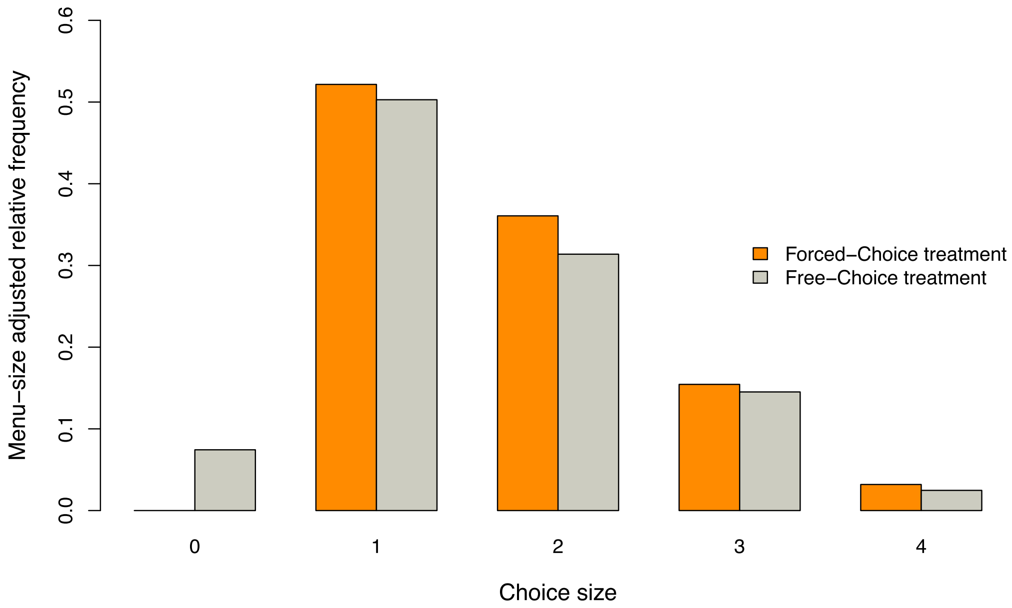

Fig. 3. Relative frequencies of the different choice sizes, adjusted for feasibility.

There is no significant difference in the distributions of subjects' average active-choice response times across treatments ($p = .612$). Notably, the average response time in both was always lowest (≈ 6 secs) when subjects chose a *single* pair of gift cards, while it increased monotonically (and statistically significantly) in similar ways as the size of choices and menus increased (Fig. 4). These novel empirical facts might be seen as intuitive evidence suggesting that the subjects' decision was easier at menus where they chose a single gift-card pair, perhaps because that was their clearly preferred one. The latter explanation would be in line with sequential-sampling and learning processes that are often mentioned in the literature of "two-alternative forced-choice decisions", such as the *drift-diffusion model* and extensions (Ratcliff and McKoon, 2008; Baldassi et al., 2020). The novelty here is that these findings point to a structure for the evolution of this process in *free-* as well as *non-binary* choices.

Yet the findings are also consistent with an alternative interpretation: subjects choosing a single item fast had to move the mouse cursor on their computer to check a single choice box on their screen, potentially mechanically resulting in shorter response times than when multiple items were selected at larger menus. This argument is less relevant in the case of deferral decisions, however, where the average response time increases monotonically from (approximately) 6 to 10 to 13 secs when the menu includes 2, 3 and 4 alternatives, respectively. This suggests that the decision to defer was not generally based on some menu-irrelevant strategy (e.g. to reach the end of the experiment quickly). Rather, it seems that it was influenced by the relevant menu's composition and, presumably, the subjects' preferences at that menu.

4.2. Choice proportions

Next, we define a subject's *choice proportion* at a menu as the number of chosen alternatives divided by the number of feasible alternatives at that menu. The distributions of average choice proportions across the two treatments are shown in Fig. 5 (top panel) and are significantly different (higher in the Forced-Choice treatment; $p = .006$). The mean (median) choice proportions across all subjects in the Forced- and Free-Choice treatments are 0.54 (0.5) and 0.49 (0.5), respectively. That is, subjects in both treatments tended to choose around half of the feasible alternatives, on average.

Additionally, 69 subjects in the Free-Choice treatment (51.1%) deferred at least once, with deferrals per subject ranging from 1 to 42 (see absolute subject frequencies in Fig. 5(b); bottom panel) and a mean (median) occurrence of 7.28 (5) menus. Moreover, 81 subjects in this treatment (60%) chose all feasible goods at least once, with such occurrences per subject ranging between 1 and 38 menus (Fig. 5(b); middle panel), and with a mean (median) occurrence of 3.42 (1). By contrast, in the Forced-Choice treatment there were 98 subjects (71%) who chose everything at least once, with occurrences per subject ranging between 1 and 45 (Fig. 5(a); middle panel), and with a mean (median) of 4.22 (2) menus. The difference in the two proportions is borderline (in)significant at the 5% level ($p = .058$). These findings lend further support to the possibility that, unable to delay their choice when faced with a hard decision, subjects are more likely to choose everything.

4.3. Choice consistency

For this analysis, which is summarized in Table 6, we compute and compare the JHM scores (see Section 2.3) of subjects' active choices under Utility Maximization while accounting for the possibility of non-trivial indifferences. That is, we compare subjects'

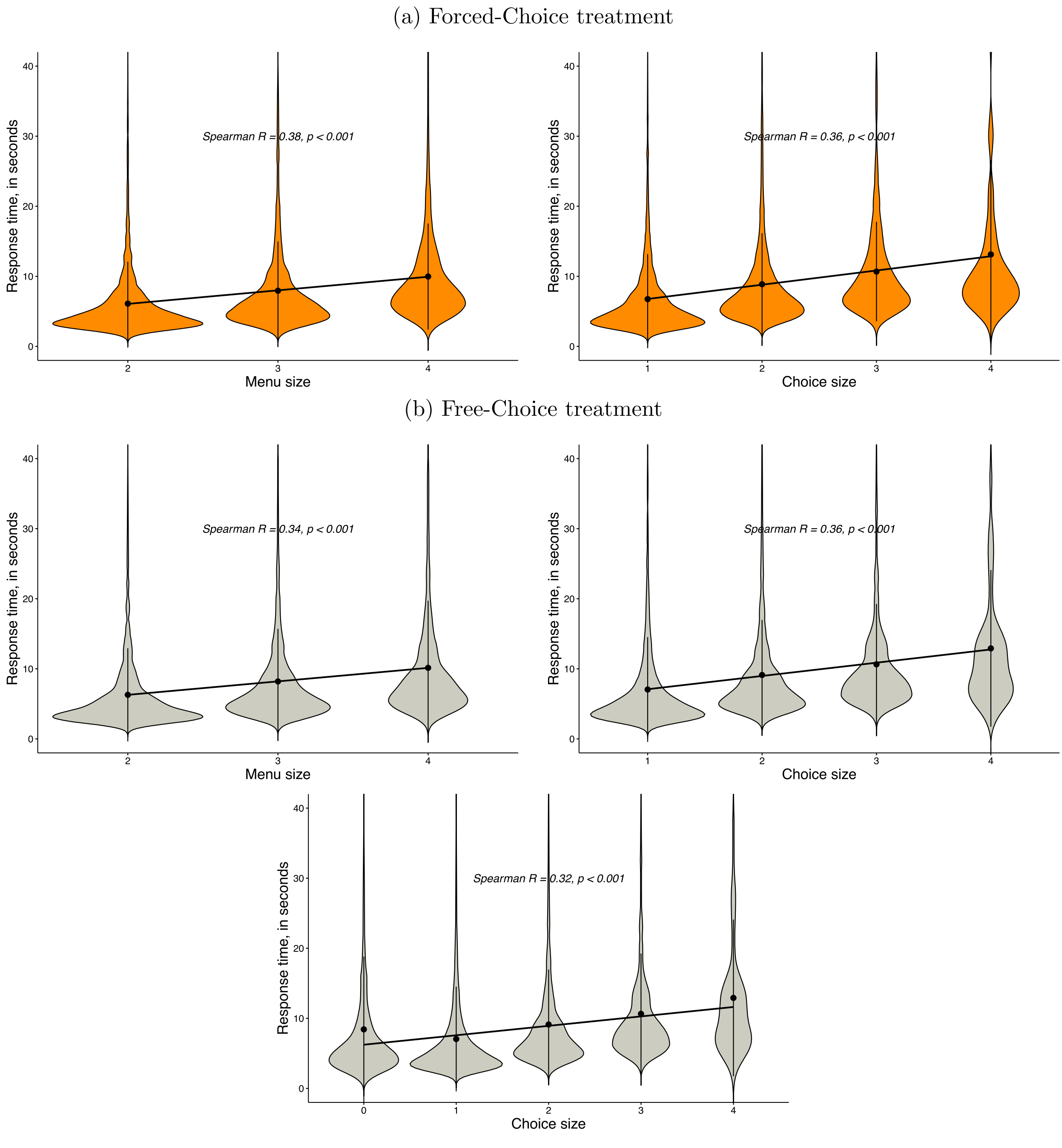

Fig. 4. Response times are positively correlated with menu and choice size in both treatments.
Note: The violin plots are thicker (thinner) in regions with more (fewer) observations, and also show the conditional average response times and 95% confidence intervals per menu/choice size.

single- and/or multi-valued choices that are compatible with indifference-permitting utility maximization, ignoring any deferral decisions. First, we focus on the distributions of the JHM scores across the two treatments. Their mean (5.38 vs 4.43), median (4.96 vs 3.50) and standard deviation (4.02 vs 3.75) are uniformly lower in the Free-Choice than in the Forced-Choice treatment, and the distributions are significantly different ($p = .045$). Thus, the number of partially or fully "mistaken"—from the point of view of the Rational Choice model—active choices is significantly smaller for subjects who were not asked to always choose some gift-card pair(s).

This is also reflected in our second test for treatment effects in active-choice consistency, where we compare the proportions of subjects who made perfectly (JHM score = 0) or, possibly, approximately (JHM score ≤ 1, 2, 3, 4 or 5) consistent active choices

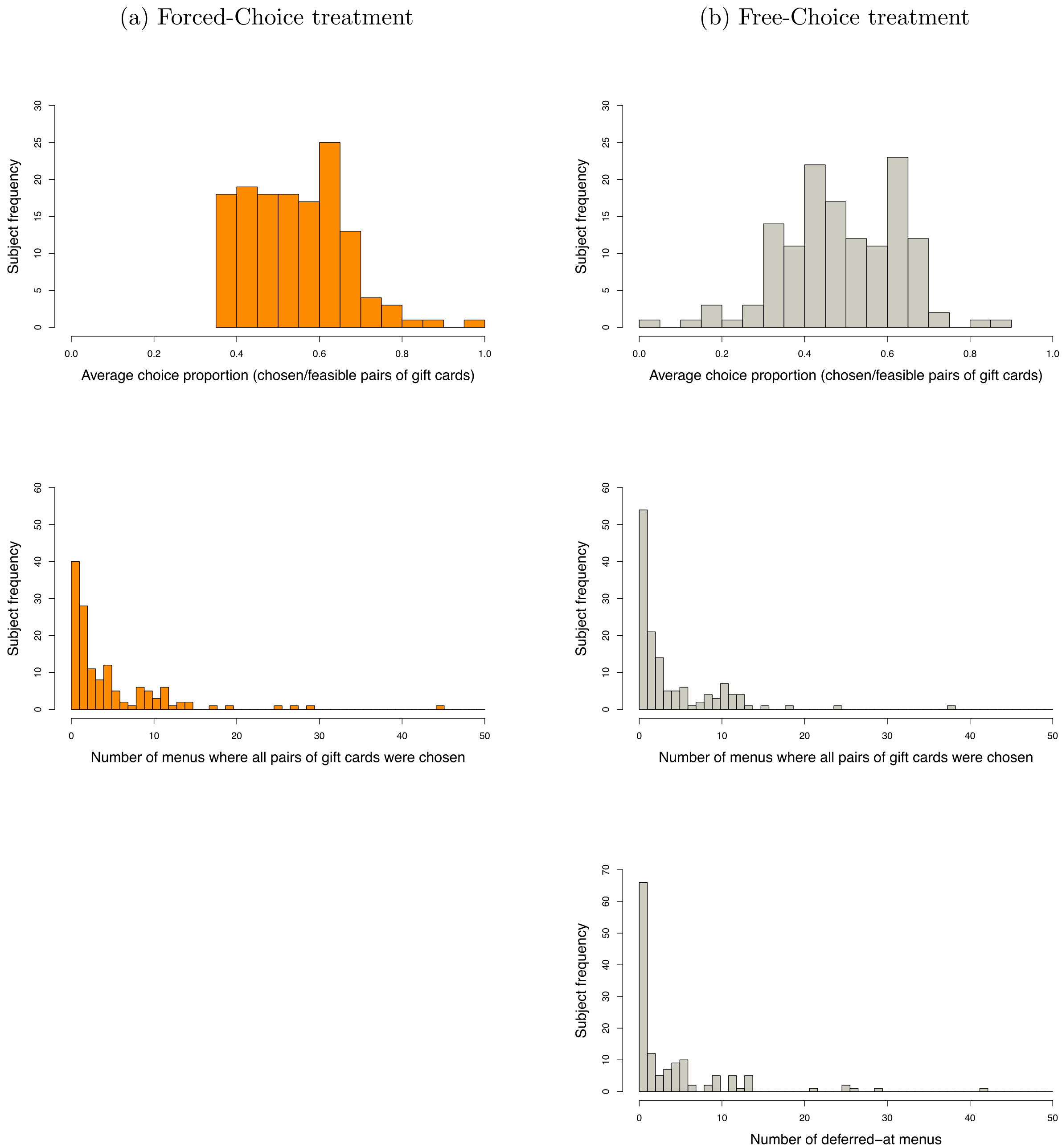

Fig. 5. Distributions of average choice proportions and menus where subjects chose everything or deferred.

in the two treatments. There are 4% and 13% perfectly consistent subjects in the Forced- and Free-Choice treatments, respectively ($p = .007$). The pattern is similar for possibly approximately consistent subjects, with 14.5% vs 26% making at most one fully mistaken choice ($p = .023$); 27.5% vs 37% two ($p = .092$); 36% vs 50% three ($p = .028$); 43% vs 54% four ($p = .069$); and 51% vs 59% up to five ($p = .181$).

To account for the fact that deferring at a menu can never decrease a decision maker's active-choice consistency, we also carry out subject-specific simulations to find how likely it would be for a possibly deferring subject to attain their JHM score given that they made their active choices only at those particular menus where they did so. More specifically, in line with Selten (1991), Beatty and Crawford (2011), Bouacida and Martin (2021) and other studies, we compute in each treatment the Selten measure of predictive success of the model of Rational Choice on subjects' active choices. We do so as follows: (1) for every experimental subject, create 10,000 datasets from as many artificial subjects who were restricted to make uniform-random active choices only at those menus

Table 6
Active-choice consistency in the two treatments.

	Forced Choice	Free Choice	p-value
Mean JHM	5.38	4.43	
Median JHM	4.96	3.50	0.045
St.Dev JHM	4.02	3.75	
JHM = 0	5	17	0.007
JHM ≤ 1	20	35	0.023
JHM ≤ 2	38	51	0.092
JHM ≤ 3	50	67	0.028
JHM ≤ 4	59	73	0.069
JHM ≤ 5	70	80	0.181
N	138	135	

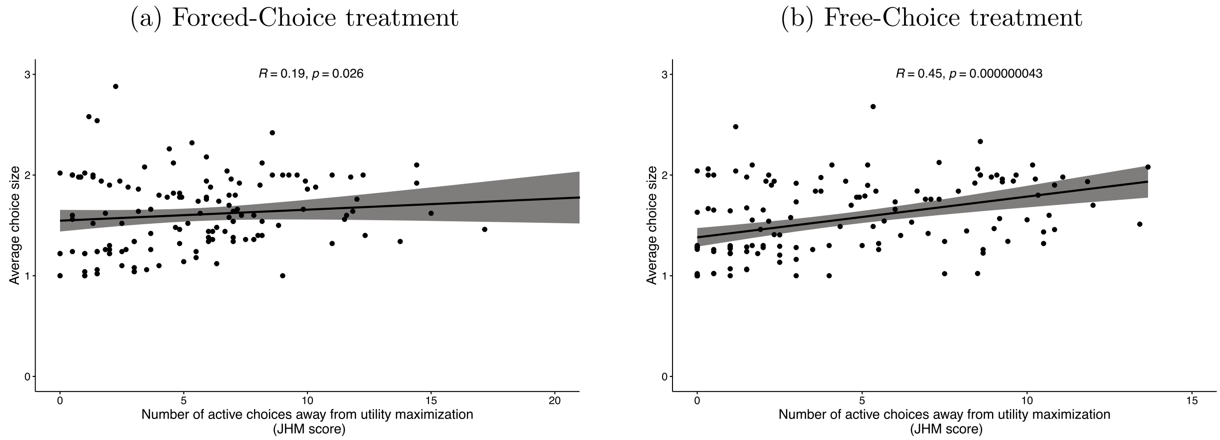

Fig. 6. Choice consistency is negatively correlated with the average choice size.
Note: R is the Spearman coefficient and p is the p-value. Shaded areas indicate 95% confidence intervals.

where the experimental subject did so; (2) find the proportion of such subject-specific simulated datasets within this block that have a zero JHM score when utility maximization under either strict or weak preferences is accounted for. Then, from the proportion, p_i, of perfectly consistent human subjects in treatment i, subtract the average proportion, a_i, of artificial subjects who are also perfectly consistent. The closer the difference $m_i := p_i - a_i$ is to 1, the higher the proportion of subjects whose active choices were consistent with utility maximization, and the more likely it is that this could not have happened randomly. Low but positive values on the other hand could arise either because: (i) relatively many experimental subjects were consistent, but probably due to chance; or (ii) relatively few subjects were consistent, and consistency in this environment was unlikely to occur by chance.

The latter case turns out to apply to the data from both our treatments, where we find

$$\begin{aligned} m_{forced} &= p_{forced} - a_{forced} &\approx 0.036 - 0 &= 0.036 \\ m_{free} &= p_{free} - a_{free} &\approx 0.126 - 0 &= 0.126 \end{aligned}$$

In line with the comparisons presented in Table 6, however, one may extend this analysis further to the case of *approximate* active-choice compliance with utility maximization. Taking as our approximation threshold the JHM score of 5 that corresponds to 10% of the maximum value that this score can take here, we find that a similar gap remains in the predictive “approximate” success of utility maximization between treatments:

$$\begin{aligned} m^{10\%}_{forced} &= p^{10\%}_{forced} - a^{10\%}_{forced} &\approx 0.507 - 0 &= 0.507 \\ m^{10\%}_{free} &= p^{10\%}_{free} - a^{10\%}_{free} &\approx 0.592 - 0.009 &= 0.583 \end{aligned}$$

Fig. 6 further shows that there is a significant negative correlation between subjects’ active-choice consistency and their average choice sizes in each of the two treatments, with this relationship being more than twice as pronounced for Free-Choice subjects ($R = 0.44$ vs $R = 0.19$ on the JHM index). In line with our preceding findings and discussion, an intuitive explanation for this difference is that, unlike Free-Choice subjects, their Forced-Choice counterparts could not defer at menus where they might perhaps have wished to do so, opting instead to choose more alternatives per menu. But while delaying choice when confronted with a difficult problem safeguards the consistency of one’s behavior, choosing more—possibly all—alternatives could do the opposite because it opens up more possibilities for choice reversals/cycles to emerge.

Some additional support for this explanation is obtained by also comparing the active-choice consistency of Free-Choice subjects who deferred at least once and those who did not. More specifically, 14 of the 69 deferring subjects (20%) were perfectly consistent in this sense, while only 3 of the 66 non-deferring ones (4.5%) were so (equivalently, 14 out of the 17 perfectly consistent in this

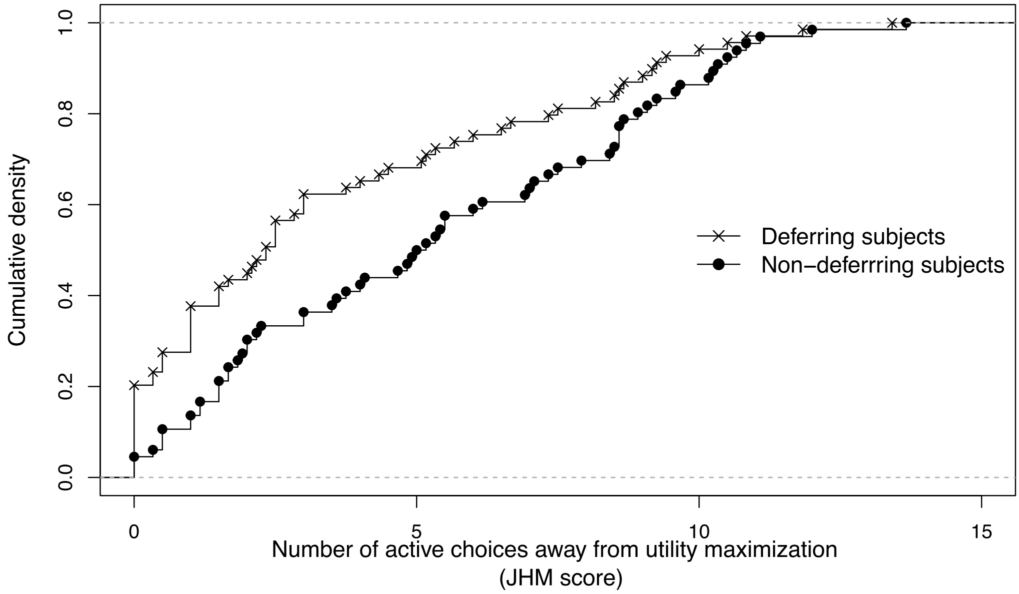

Fig. 7. In the free-choice treatment, deferring subjects are more consistent than non-deferring ones.

treatment deferred at least once). The difference between these two proportions is significant ($p = .008$). Moreover, the two groups had an average (median) JHM score of 3.58 (2.33) and 5.31 (5.08), respectively, and the difference in the two distributions is significant ($p = .003$). In fact, as shown in Fig. 7, deferring subjects are uniformly more consistent than non-deferring ones in the sense that the distribution of their JHM scores first-order stochastically dominates the corresponding distribution of the latter subjects. This within-treatment effect further highlights the important mediating role of deferrals for consistency.

The results from all these tests document a clear negative effect that forcing choice exerts on subjects' choice consistency. This complements in some important ways the respective finding that was originally documented in CCGT22. In particular, our results show that forced choices are less consistent than free choices even when: (i) subjects have to decide from nearly twice as many menus (50 vs 26); and (ii) the experimental design allows subjects to make multi-valued choices, which in turn enables the analyst to test for the possibility that their behavior is well-approximated by utility maximization *with or without indifferences*.

We ask, finally, how a subject's choice consistency, as captured by the JHM index, is related to their response times. Fig. 8 (top panel) shows that there is a significant—and very similar in size, with a Spearman coefficient of 0.35 and 0.45—*positive* correlation, in both the Forced- and Free-Choice treatments, between subjects' active-choice JHM scores and their average response times. That is, choice consistency is *negatively* correlated with the time it takes for subjects to make their active choices. This interesting finding is broadly consistent with the important prediction of sequential sampling and introspective learning models mentioned earlier. These predict that shorter response times are more likely when a clearly preferred alternative exists, which would imply in turn more consistent active-choice behavior. The novelty here, again, is that this is found in a rich dataset that comprises both binary and non-binary menus.

A distinct possible explanation, however, is that subjects with a higher cognitive ability—an unmeasured variable in this study—are both more consistent and faster than subjects with a lower cognitive ability. Yet another possibility is that spending more time before deciding is more likely a characteristic of people who are prone to second thoughts, and therefore more likely to be involved in choice reversals when presented with a series of decision problems. Conversely, it is also possible that subjects who had neither a stable preference relation nor a clear decision rule ended up spending more time on average at each menu, but without this extra time ultimately alleviating the effects of their ambivalence.

5. Individual-level analysis

5.1. Non-parametric model fitting

We now turn to the tests of our possibly multi-valued and empty-valued experimental choice data for their consistency with each of the three models introduced in Section 2. These tests were performed by applying the model-specific JHM method that was introduced in Section 2.3.[17] A brute-force combinatorial-optimization algorithm was implemented for these computations. This

[17] The model-based extension of HM for single-valued (forced or free) choices was originally introduced in Costa-Gomes et al. (2016, 2022). Since Gerasimou and Tejiščák (2018), this has been extended and made freely available online in the free and open-source tool Prest . With retainment of its original focus on the model of Rational Choice/Utility Maximization with strict preferences, the HM approach has been extended in another

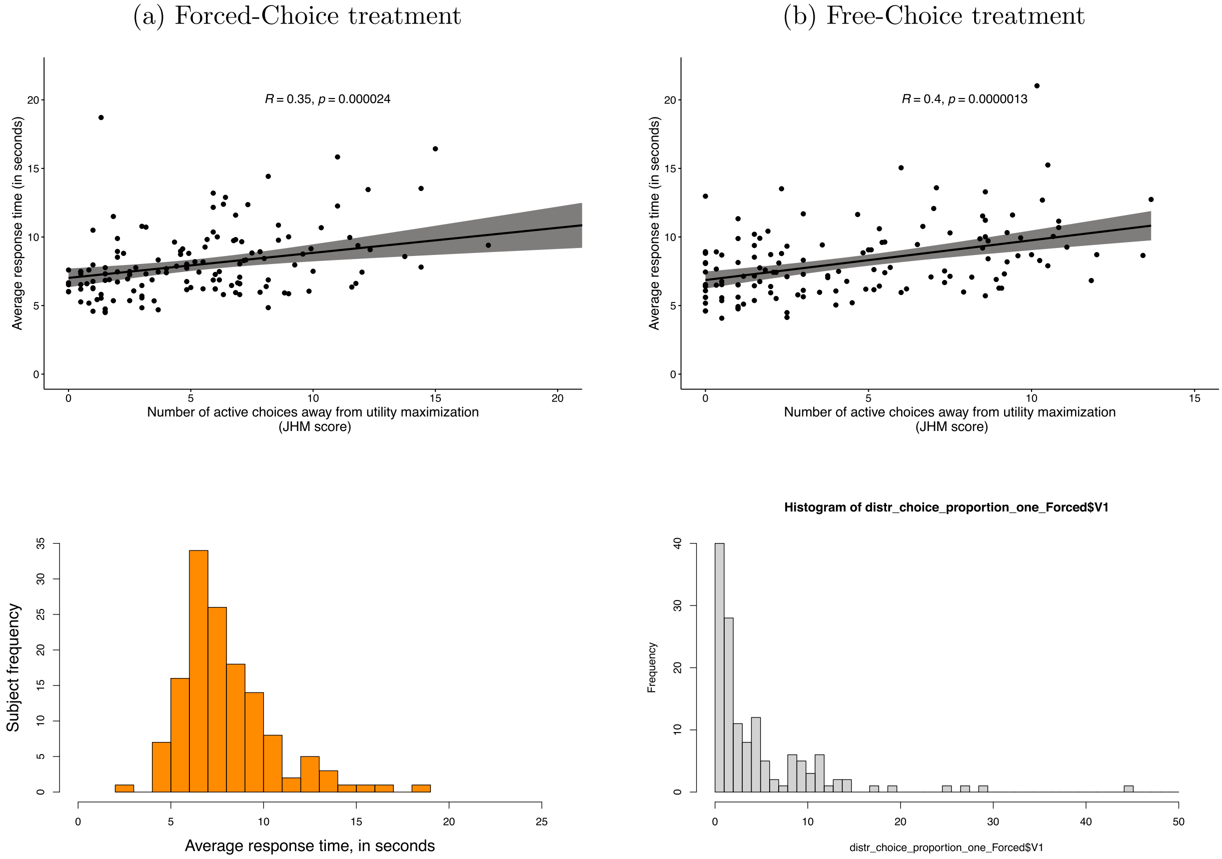

Fig. 8. Choice consistency is negatively correlated with average response times in both treatments.
Notes: R is the Spearman coefficient and p is the p-value. Shaded areas indicate 95% confidence intervals. There is no significant difference in the average active-choice response times between treatments ($p = .550$).

involved the production of all choice datasets that are generated by all possible instances of every model, and comparing each such dataset against every subject's own dataset in order to find the model(s) and instance(s) that are closest to it in the minimum JHM score sense. This method therefore detects perfect as well as *approximate* model fits, and in the latter case it quantifies the approximation in an intuitive way that may be interpreted as the number of "mistakes" made by an agent who might be portrayed as following a particular decision rule.

Despite the large numbers of weak orders (4,683), partial orders (130,023) and, especially, incomplete preorders (209,527) that are defined on a set of 6 alternatives (Online Encyclopedia of Integer Sequences, 2021), the above tool makes such an *exact* computation possible very quickly—in less than 12 seconds—for all subjects in each treatment, and for each of the three models. Furthermore, although the brute-force algorithm is linear in the number of subjects but exponential in the number of alternatives, the model-rich distance-score method itself is scalable and can be extended to analyze datasets that are derived from much larger sets of alternatives by employing powerful open-source constraint solvers.

Table 7 and Fig. 9 present summaries of this goodness-of-fit analysis for both Forced- and Free-Choice subjects. In line with standard practices, whereby one also wishes to understand the extent to which a certain behavior or model fit could have occurred randomly (Bronars, 1987), we also applied our method on simulated datasets of artificial uniform-random behaving subjects under both a Forced- and a Free-Choice configuration (138,000 and 135,000 such subjects, respectively).[18] Their JHM score distributions for each of the three models are juxtaposed in Fig. 9 with those of human subjects. For comparison, the minima from the simulated-subject JHM score distributions are presented in Table 7.

intuitive direction by Apesteguia and Ballester (2015). However, although the "Swaps" measure of Apesteguia and Ballester (2015) that pertains to general choice environments is also readily computable in Prest, it is currently unclear how it should be computed when subjects choose multiple items from the same menu, or how it should be extended—if at all—to models that generalize utility maximization.

[18] Details on the simulations' structure are available in Online Appendix E.

Table 7
Classification of subjects who are perfectly/approximately explainable by one of the three models.

	Utility Maximization	Undominated Choice with Incomplete Preferences	Dominant Choice with Incomplete Preferences	All
	Forced-Choice treatment ($N = 138$)			
Subjects with JHM = 0	5 (4%)	0	–	5 (4%)
Subjects with JHM ≤ 5	61 (44%)	16 (11.5%)	–	77 (56%)
Mean (median) best score (JHM ≤ 5)	2.01 (1.83)	3.56 (4)	–	2.34 (2)
Mean (median) best-model preference orderings (JHM ≤ 5)	1.11 (1)	4.75 (2)*	–	1.87 (1)
Minimum JHM score in simulations	13.25	12.33	–	–
	Free-Choice treatment ($N = 135$)			
Subjects with JHM = 0	3 (2%)	0	7 (5%)	10 (7%)
Subjects with JHM ≤ 5	30 (22%)	7 (5%)	37 (27%)	74 (55%)
Mean (median) best score (JHM ≤ 5)	2.12 (1.75)	2.37 (2.08)	1.93 (2)	2.05 (2)
Mean (median) best-model preference orderings (JHM ≤ 5)	1 (1)	4.14 (2)*	1.11 (1)	1.35 (1)
Minimum JHM score in simulations	16.92	15.75	16	–

Model-score ties were always broken in favor of Rational Choice/Utility Maximization (no other ties emerged).
*Across the indifference-permitting and indifference-excluding compatible instances (these are observationally equivalent).

For human subjects, Rational Choice often tied with one—but never both—of the other two models. The classification presented in Table 7 broke all such ties in favor of the textbook model. Under this tie-breaking assumption, the two exhibits show—separately for each of the two experimental treatments—the number, proportion and relative-frequency distributions of distinct subjects that are on average no more than 10% away from being explainable perfectly by a model. We chose this conservative approximation range for two reasons: (i) simulations suggest that a JHM score of 5 for any of the three models is extremely unlikely to occur randomly in this decision environment; (ii) for subjects whose score does not exceed this threshold there is typically a *unique* best-matching preference ordering that explains their behavior under the respective subject-optimal model, particularly for Utility Maximization and Dominant Choice with Incomplete Preferences.

The proportion of Forced-Choice subjects with a perfect or 10%-approximate model fit was 4% (all under Rational Choice) and 56%, respectively, with 44% and 11.5% of those in the latter group classified under Rational Choice and Undominated Choice with Incomplete Preferences. Furthermore, the proportion of Free-Choice subjects with a perfect or 10%-approximate fit was 7% and 55%, respectively. 22%, 5% and 27% of all subjects in this treatment were categorized under Rational Choice, Undominated, and Dominant Choice with Incomplete Preferences, respectively. In addition, 3 and 7 of these subjects were *perfectly* compatible with the first and third model, respectively. One subject in the former group and five in the latter revealed indifferences in their weakly ordered and weakly preordered preferences, respectively.

The model fit is slightly better on average in the Free-Choice treatment, with a mean (median) best JHM score of 2.05 (2) vs 2.34 (2), although the distributions are not significantly different. Similarly, and in line with the results from the consistency analysis of Section 4, the proportion of subjects who were best-matched by one of the two models of consistent active choices (i.e. Rational Choice and Dominant Choice with Incomplete Preferences) is higher in the Free-Choice treatment (49% vs 44%). Additionally, the proportion of subjects who are best-matched by the third model (Undominated Choice with Incomplete Preferences), which generally predicts inconsistent active choices, is lower in that treatment (5% vs 11.5%). When Rational Choice is not the best-matching model, moreover, its distance score is on average 1.4 and 5.7 JHM score units higher in the Forced- and Free-Choice treatments, respectively. Finally, the highest JHM score of 5 for experimental subjects in this approximate model-fitting analysis is far below the corresponding minimum scores of simulated random-behaving subjects, which ranged between 12.3 and 16.9.

For the 37 subjects who are best-explained by Dominant Choice with Incomplete Preferences we conducted an additional robustness check. Specifically, by analogy to the way in which Selten (1991) predictive success measure of the Rational Choice model was evaluated in Section 4.3, we generated a set of 10,000 random-behaving simulated subjects for each of the 37 human experimental participants who were best-matched by the Dominant-Choice model. In every such subject's simulations block, moreover, the menus at which the underlying human subject deferred were held fixed. This allows for evaluating the model's predictive success in this environment within a 10%-approximation range. Fig. 10 shows the histogram of differences between the minimum JHM score in each subject-specific 10k-simulations block and the relevant subject's actual JHM score. For all 37 subjects the difference is positive and typically much greater than 10 JHM score units (here, again, pertaining to either active-choice or deferral decisions). This analysis

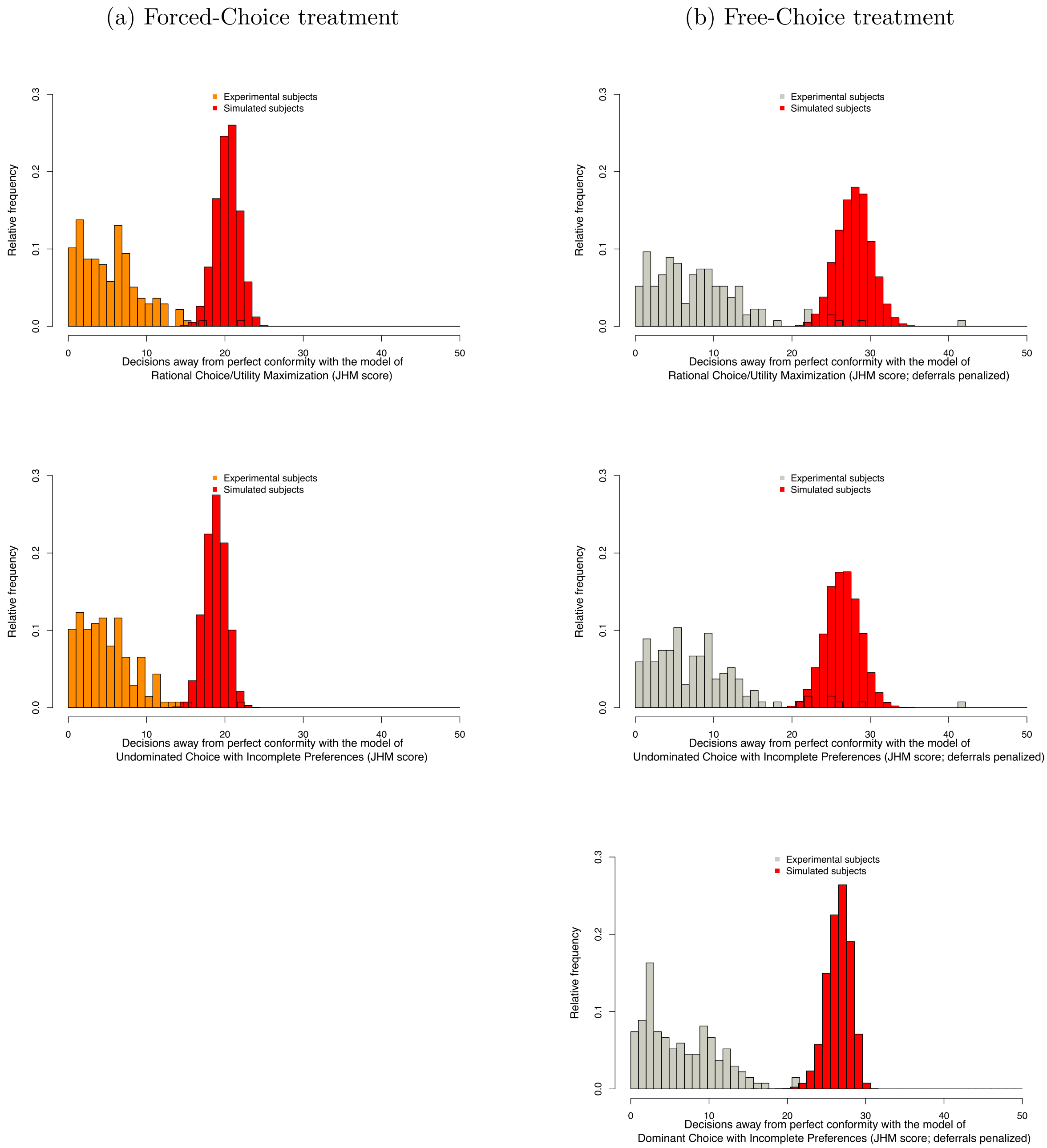

Fig. 9. Distributions of Jaccard-Houtman-Maks distance scores for each of the three models..

suggests that the Dominant-Choice model is not excessively permissive in this environment, and that the fits it is associated with could not plausibly have occurred under random behavior.[19]

[19] We note that 30 of these 37 subjects, as well as 26 of the 61 Free-Choice subjects who are best-matched by Utility Maximization, are approximated equally well (but not better) by Utility Maximization with an Outside Option. This model, which is analyzed choice-theoretically in Gerasimou (2018a, Section 3), assumes that "choosing nothing" is an additional alternative that can be added to the set of 6 gift-card bundles, and occupies a position in the subject's weak ordering over this expanded set. The model then predicts that choice is deferred at a menu if and only if the outside option is ranked above every gift-card bundle in that menu. This decision rule is not incentive-compatible in the context of our experimental design. The reason is that, even if none of the £20-worth gift-card bundles in a menu was considered to be sufficiently appealing, deferring would be

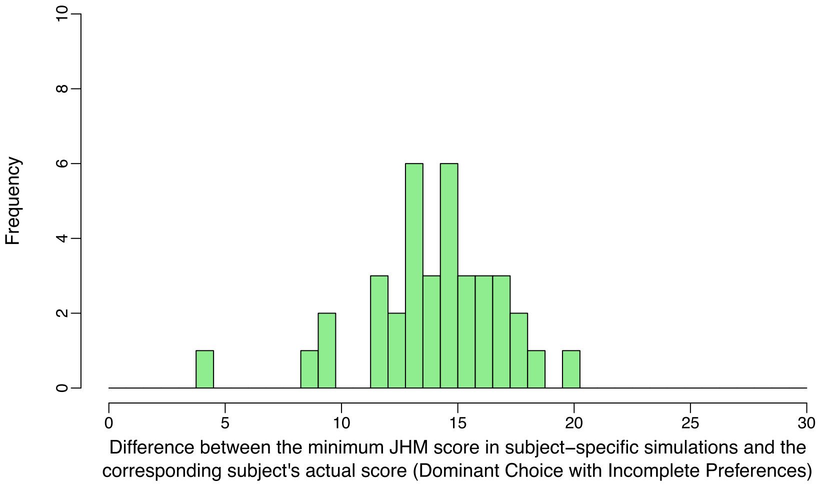

Fig. 10. All 37 subjects that were best-matched by Dominant Choice with incomplete preferences were unlikely to have achieved this classification at random given the menus where they deferred.

Table 8
Revealed indifferences conditional on each subject's best-fitting model.

	Indifference-revealing subjects		Mean (median) indifferences	
Utility Maximization	70	(80%)	3.00	(2)
Dominant Choice with Incomplete Preferences	30	(81%)	2.29	(2)
Undominated Choice with Incomplete Preferences	16	(69.5%)	0.69*	(0.5)*
Overall	119	(79%)	2.47	(1.2)

*Across the indifference-permitting and indifference-excluding compatible instances (these are observationally equivalent).

5.2. Recovery of strict preferences, indifference and indecisiveness

As was highlighted earlier, the collected choice data and the non-parametric method applied on them jointly allow us to separate the strict part from the potential indifference and/or incomparability parts of each subject's revealed preferences, conditional on the model that best explains the subject's overall behavior. For the vast majority of subjects that are perfectly or approximately matched by some model, this analysis reveals a unique (weak or strict; complete or incomplete) preference ordering under that model. Indeed, the mean number of such orderings is 1.87 and 1.35 in the Forced- and Free-Choice treatments, respectively, with a median of 1 in both cases. Thus, despite the relatively large number of 50 distinct menus and decisions in the experiment, and notwithstanding that this collection is still smaller than the "ideal" full-domain collection of 63 menus, this model-rich analysis generates sharp and subject-specific preference estimates. The directed graphs of the subject-optimal preference orderings that were . recovered with this method are shown in Online Appendix D for all relevant subjects.[20]

We now turn to the important task of distinguishing between indifference and incomparability/indecisiveness. To this end, it is perhaps worth recalling first that indifferences are generically non-existent in a large class of the benchmark Bewley (2002) class of incomplete preference relations under uncertainty, which are defined on a space of bundles of continuous commodities or uncertain acts (Gerasimou, 2018b). By contrast, however, this is very much not the case when such preferences are over finitely many discrete options, as in the choice environment studied in this paper. Indeed, our analysis shows that out of the 60 subjects (22% of the total) that the two models of incomplete-preference maximization explain optimally across the two treatments, it is possible to separate—in the ways discussed in Section 2.2–the strict, indifference and incomparability/indecisiveness components in the preferences of 46 (77%) of them. The empirical documentation of this theory-based separation of the three binary relations that are possible when preferences are incomplete is a novel contribution of this paper. Fig. 11b and c show the directed graphs of two subjects' incomplete weak/strict preferences that were optimally recovered by these models; Fig. 11a does so for a utility-maximizing subject.

Looking across the three models and two experimental treatments, Table 8 summarizes the key revealed-indifference information that is derived from each subject's best-fitting model, and demonstrates the descriptive relevance of indifference in these data. In

dominated by choosing the most preferred feasible bundle: if that menu was randomly selected for payment, the subject would choose such a bundle in any case; but if they had deferred at that menu previously, then they would receive a lower cash reward than if they had not.

[20] The graphs for each recovered preference ordering were visualized with the GraphViz open-source tool (Gansner and North, 2000) from within the Prest tool.

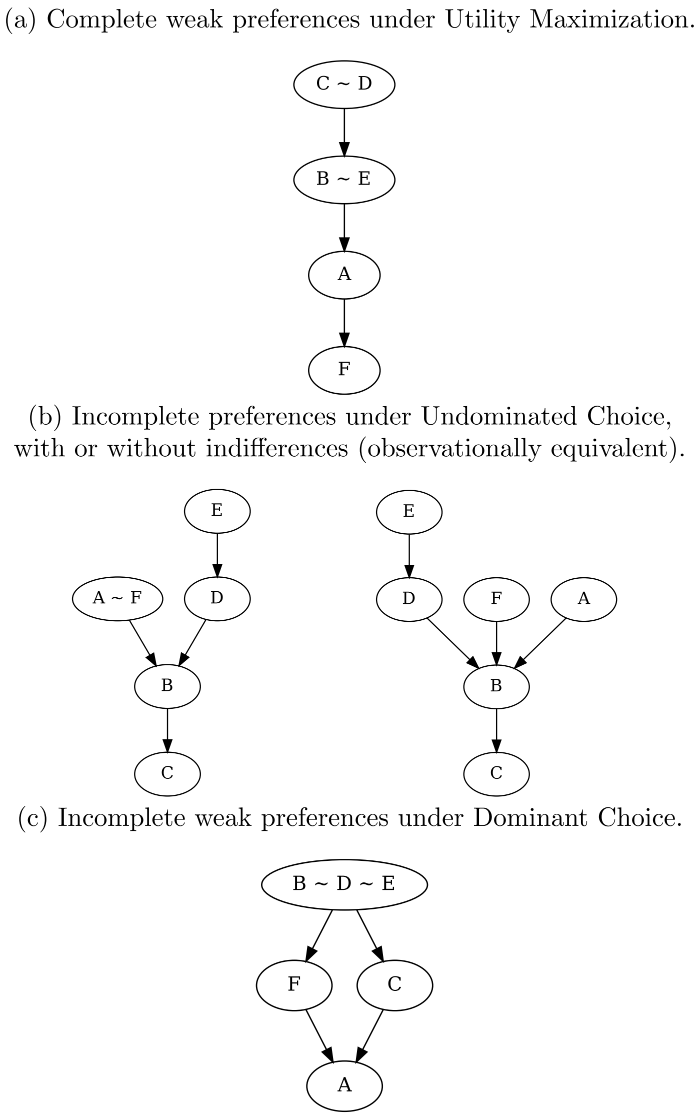

Fig. 11. Graphs of three subjects' preference relations recovered optimally by the three models.

particular, 80%, 81% and 69.5% of all subjects who are best-explained by Utility Maximization, Dominant Choice and Undominated Choice with Incomplete Preferences, respectively, have a model-compatible preference relation that reveals at least one indifference between distinct gift-card bundles. Across all models, subjects and their respective model-compatible preference relations, the average (median) number of indifference comparisons between distinct gift-card pairs is 2.5 (1.2). Considering that there are 15 possible binary comparisons on a set of 6 elements, this means that the indifference relation accounts for 17% of subjects' preference comparisons, on average. These averages and proportions are highest for Utility Maximization (3; 19%), followed by Dominant (2.3; 15%) and Undominated Choice with Incomplete Preferences (0.69; 5%).

5.3. Assessing the robustness of model-fitting estimates

We report, finally, on a test that aimed to assess the robustness of the goodness-of-fit estimates presented earlier. In this test, we split each subject's 50 decisions into those that appeared in the first and second half of the experiment[21] and conducted the same non-parametric analysis separately on each half. We then counted the subjects for whom the same model was found to be optimal in each of the first and second 25 decisions, as well as in all 50. This is similar in spirit to conventional parametric cross-validation exercises, but adapted to account for this study's non-parametric environment and discreteness of the JHM goodness-of-fit measure.

[21] Recall that each subject had a randomly different menu-presentation order.

In the Forced-Choice treatment, 53 of the 77 subjects (69%) were best-matched by the same model in both halves and overall, possibly in distinct modes (i.e. with or without indifferences). Moreover, 14 of the remaining subjects (18%) were best-matched by Undominated Choice with Incomplete Preferences in the first half, and by Rational Choice in the second. Furthermore, for seven subjects (9%) the transition was in the other direction. Analogous findings pertain to the Free-Choice treatment, despite the fact that all three models have meaningful explanatory power gives rise to additional possibilities. In particular, 44 of the 74 subjects in this treatment (59%) were best-matched by the same model in the first and second halves and overall. Finally, 11 individuals (15%) who were best-matched by a model of incomplete-preference maximization in the first half behaved as utility maximizers in the second half, while the opposite was true for another 12 subjects (16%). In summary, a sizeable proportion of subjects in both treatments are best-matched by the same model throughout, and this is more pronounced in the Forced-Choice treatment.

6. Testing for the presence of other decision models and heuristics

6.1. Satisficing

Simon (1956) famously coined the term *satisficing* to describe resource-constrained agents who, instead of searching through all available alternatives in order to find the best, only search until they find one that meets an acceptability threshold. Two recent forced-choice studies that tested the satisficing hypothesis in economics are Reutskaja, Nagel, Camerer and Rangel (2011) and Caplin, Dean and Martin (2011). The former found weak evidence for such behavior in choice from lists of food snacks under intense time pressure, whereas the latter found strong evidence in choice from lists of monetary amounts that were described verbally through a series of additions and subtractions, without (or with limited) time pressure. The fact that all 50 menus in both treatments of our experiment were presented vertically as unnumbered lists allows us to complement the preceding model-based analysis with a test for satisficing in the present time-unconstrained and possibly free-choice environment where multiple gift-card bundles could be selected at each menu.

To this end, we first consider a relatively narrow but potentially informative test by focusing on the frequency with which subjects opted to choose only the *first* alternative that appeared in the menus they saw. We ask, then, whether this frequency can be viewed as being above and beyond what might be reasonably interpretable differently. We find that 7 and 15 subjects (8%) in the Forced- and Free-Choice treatments, respectively, who are not approximately explainable by one of the three models discussed earlier, chose only the first option at frequencies that strictly exceeded the 97.5% cut-off values of 0.28 and 0.22 that are derived from the relevant simulations.[22] When the 95th percentile cut-off values of 0.26 and 0.2 are used as thresholds instead, these numbers rise to 8 and 18, respectively (9.5%).

In addition to this satisficing criterion, as a potential indicator of such behavior we also use the average position of each subject's chosen item(s) in the 50 menus' list orderings. For the 5 and 1 (2.2%) subjects in the Forced- and Free-Choice treatments, respectively, who were not classified as approximately explainable by one of the three models, we find an average position of their chosen item(s) strictly below the 2.5th percentile cut-off value of 1.83 that is derived from simulated uniform-random forced choices (common in both treatments). When the 5th-percentile cut-off average position of 1.86 is used instead, those numbers rise to 6 and 3, respectively (3.3%). Accounting, finally, for subject overlaps across this categorization criterion and the one preceding it leads to 8 and 15 (8.5%) or 10 and 19 (10.6%) distinct subjects in the Forced- and Free-Choice treatments who, under the high and low-confidence statistical analyses, respectively, may be viewed as exhibiting a systematic satisficing behavior.

Interestingly, a cross-treatment comparison of how the average positions of all chosen alternatives are distributed shows that Forced-Choice subjects are more likely to choose items that appear higher up in the menu list ($p = .009$; Fig. 12). An intuitive explanation is that, for individuals who are either insufficiently motivated to make 50 careful decisions over 2–4 pairs of gift cards in an experimental lab or find the task to be cognitively challenging, being unable to avoid/defer the decision at a small cost could make the use of a non-compensatory decision heuristic such as satisficing more likely.

6.2. Preference for randomization

Agranov and Ortoleva (2017, 2025), Dwenger, Kübler and Weizsäcker (2018), Cettolin and Riedl (2019) and Agranov, Healy and Nielsen (2023), among others, have recently documented a *preference for randomization* in repeated choices from binary menus of money lotteries.[23] Such a preference refers to subjects' frequent tendency to change their choices from one occurrence of a menu to the next, and their willingness to also incur a small cost in order to have their choice determined randomly. The aspect of our experimental design whereby subjects are rewarded with a random alternative at a menu if they chose everything at that menu allows us to test for the existence of a similar preference for randomization.

As with our analysis of satisficing, to proceed with this investigation we regard a subject as potentially exhibiting such a preference if both of the following are true: (i) they are not approximately explainable by one of the three models of preference maximization;

[22] These cut-off values were obtained as follows. First, for each synthetic dataset we found the number of menus where only the first alternative in the menu was chosen (all within-menu orderings here replicated those in the experiment). Then, we found the 97.5th percentile value in the distribution that results from these 10,000 datasets.

[23] In addition, Ong and Qiu (2023) report such a preference in an ultimatum-bargaining setting where receivers are willing to incur a cost in order to randomize between acceptance and rejection of an offer.

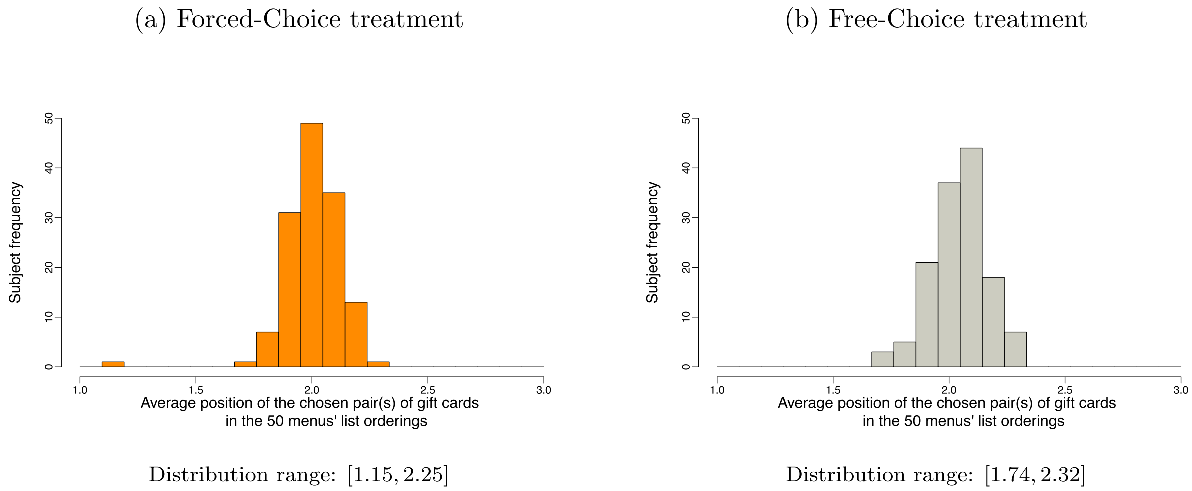

Fig. 12. Forced-Choice subjects are more likely to select items that are higher up in the menu list.

(ii) the number of menus were they chose everything exceeds the 97.5th (95th) percentile simulations-based cut-off values[24] of 14 and 11 (13 and 10) for the Forced- and Free-Choice treatments, respectively. Under either level of significance, the two criteria are simultaneously satisfied by 2 and 5 subjects (2.5%).[25] It is worth noting, however, that these estimates are conservative and may be better seen as lower bounds. This is so because the experimental design allows subjects who might have exhibited such a systematic preference for randomization to reveal it only when they were faced with difficulty deciding between *all* feasible alternatives at a given menu, not from a proper subset thereof. Despite this and the fact that our experiment was not designed with such tests in mind, the finding reported here adds to the literature on preference for randomization by showing that such a preference could also be manifested: (i) in binary as well as non-binary menus: (ii) with riskless alternatives; (iii) when deferral is feasible and acts as another obvious way for an individual to deal with a difficult decision; and (iv) in non-repeated-choice environments.

7. Concluding remarks

This paper proposes and implements theory-guided methods of data collection and analysis that aim to contribute towards eliciting a decision maker's possibly weak or incomplete preferences and, where relevant, distinguishing between their indifference and indecisiveness parts. On the data-collection side, the paper contributes an incentivized experimental design with multi-valued forced- and free-choice treatments. On the data-analytic side, it deploys a model-rich, non-parametric method that allows for recovering an individual's possibly weak or incomplete preferences from their multi-valued choices in a model-optimal way. Using the collected experimental data, this method was fruitfully applied to the textbook model of riskless utility maximization with potential indifferences, and to two models of incomplete-preference maximization. The latter models allow for—and may distinguish between—indifference and/or incompleteness. The method is general and can also be applied to other models of bounded-rational general choice.

Despite the relatively large number of decisions from 50 distinct menus, the behavior of more than 55% of all subjects in our sample is either perfectly or approximately matched by one of the simple but richly structured deterministic model of complete or incomplete preference maximization that we considered. Furthermore, in the vast majority of cases these fits correspond to a unique preference ordering. Utility maximization accounts for the behavior of over 60% of all subjects in this group (33% of the total). Importantly, the complete or incomplete preference relation that is optimally recovered conditional on a subject's best-matching model typically features a non-trivial indifference relation that may encompass up to 19% of all possible comparisons. This highlights the potential importance of accounting for indifferences in revealed-preference analyses on discrete choice environments.

Finally, we conducted additional investigations on the 45% of subjects who lied outwith the above categorization, to understand if some might be best viewed as behaving according to another model or heuristic. Focusing on preference for randomization and two potential modes of satisficing behavior, we found that up to an additional 11%-13% of all subjects could be described in one of these ways. This raises to nearly 70% the total proportion of subjects whose behavior we can explain. While this proportion can be increased, mechanically, by increasing the approximation threshold in the model-based analysis of Section 5 (e.g. from 10% to 20%),

[24] These cut-off values were obtained as follows. First, for each synthetic dataset we found the number of menus where all alternatives were chosen. Then, we found the 97.5th and 95th percentile values in the distribution that results from these 10,000 datasets. An alternative to this approach might be to focus on the aggregate synthetic data; find the relative frequency of menus where everything is chosen; then the number of menus—out of 50—where this happens in expectation; and finally use this as the (non-)randomness benchmark for this analysis. These frequencies are 0.377 and 0.231 in the Forced- and Free-Choice synthetic data (97.5th percentile values), leading to such alternative cut-off values of 18.85 and 11.55, respectively. Following this approach does not alter the set of subjects who may be viewed as systematic randomizers.

[25] None of these subjects are "satisficers" under the analysis of the previous subsection.

such explanatory gains would be mitigated by the ensuing indeterminacy in the preferences that would be compatible with such less good fits.

Overall, the paper's findings point to the usefulness of free- and forced-choice experimental designs that allow for multi-valued choices toward eliciting subjects' indifference and/or indecisiveness relations alongside their strict preferences. It also highlights the importance of methodologically pluralistic methods that enable analysts to recover preferences or choice rules by analyzing observable behavioral data through the lens of rational choice/utility maximization as well as other models of imperfectly rational choice.

Data availability

The full replication package is available here: https://www.openicpsr.org/openicpsr/project/248549/version/V1/view.

Declaration of competing interest

I, Georgios Gerasimou, have no material or financial interests to disclose in relation to this submission.

Funding

I cordially thank Carlos Cueva, Miguel Costa-Gomes and Matúš Tejiščák for our previous joint work and for their comments on an earlier version of this paper. For helpful feedback I also thank Elias Bouacida, referees and audiences at SAET 2021 (online), SPUDM 2021 (online), DETC 2022 (Durham), FUR 2022 (Ghent), SAET 2022 (online) and EEA-ESEM 2022 (Milan). Nathan Jones, Iravati Kashyap, Kanishk Khanna and Jacob Beaddie provided excellent research assistance with computations (2020-21). The project was supported by British Academy and University of St Andrews funds, and was approved by the University of St Andrews Research Ethics Committee (EC9805/EC16707). This is a significantly revised version of a paper that was originally circulated in November 2021 (arXiv:2111.14431). Any errors are my own.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.geb.2026.05.011.

References

Agranov, M., Healy, P.J., Nielsen, K., 2023. Stable randomization. Econ. J. 133, 2553–2579.
Agranov, M., Ortoleva, P., 2017. Stochastic choice and preferences for randomization. J. Polit. Econ. 125, 40–68.
Agranov, M., Ortoleva, P., 2025. Ranges of randomization. Rev. Econ. Stat. 107, 1702–1713.
Apesteguia, J., Ballester, M., 2015. A measure of rationality and welfare. J. Polit. Econ. , 1278–1310.
Arvanitis, S., Scaillet, O., Topaloglou, N., 2023. Spanning analysis of stock market anomalies under prospect stochastic dominance. Manag. Sci. 70, 6002–6025.
Azrieli, Y., Chambers, C.P., Healy, P.J., 2018. Incentives in experiments: a theoretical analysis. J. Polit. Econ. 126, 1472–1503.
Baldassi, C., Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., Pirazzini, M., 2020. A behavioral characterization of the drift diffusion model and its multi-alternative extension for choice under time pressure. Manag. Sci. 66, 5075–5093.
Beatty, T. K.M., Crawford, I.A., 2011. How demanding is the revealed preference approach to demand. Am. Econ. Rev. 101, 2782–2795.
Bewley, T.F., 2002. Knightian decision theory. part i. Decis. Econ. Finance 25, 79–110.
Bossert, W., Sprumont, Y., Suzumura, K., 2005. Maximal-element rationalizability. Theory Decis. 58, 325–350.
Bossert, W., Suzumura, K., 2010. Consistency, Choice and Rationality. Cambridge, MA: Harvard University Press.
Bouacida, E., 2021. Identifying choice correspondences: a general method and an experimental implementation. University of Lancaster, Working Paper 2021/006.
Bouacida, E., Martin, D., 2021. Predictive power in behavioral welfare economics. J. Eur. Econ. Assoc. 19, 1556–1591.
Bronars, S.G., 1987. The power of non-parametric tests of preference maximization. Econometrica 55, 693–698.
Caplin, A., Dean, M., Martin, D., 2011. Search and satisficing. Am. Econ. Rev. 101, 2899–2922.
Cettolin, E., Riedl, A., 2019. Revealed preferences under uncertainty: incomplete preferences and preferences for randomization. J. Econ. Theory 181, 547–585.
Costa-Gomes, M., Cueva, C., Gerasimou, G., Tejiščák, M., 2016. Choice, deferral and consistency. University of St Andrews Working Paper 1416. Available at https://georgiosgerasimou.com/_static/paper-CDC-2016.12.26.pdf.
Costa-Gomes, M., Cueva, C., Gerasimou, G., Tejiščák, M., 2022. Choice, deferral and consistency. Quant. Econ. 13, 1297–1318.
Cubitt, R., Navarro-Martinez, D., Starmer, C., 2015. On preference imprecision. J. Risk Uncertain. 50, 1–34.
Danan, E., 2010. Randomization vs selection: how to choose in the absence of preference? Manag. Sci. 56, 503–518.
Danan, E., Ziegelmeyer, A., 2006. Are preferences complete? An experimental measurement of indecisiveness under risk. Working Paper.
Dean, M., 2008. Status quo bias in large and small choice sets. Working Paper.
Dwenger, N., Kübler, D., Weizsäcker, G., 2018. Flipping a coin: evidence from university applications. J. Public Econ. 167, 240–250.
Eliaz, K., Ok, E.A., 2006. Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. Games Econ. Behav. 56, 61–86.
Gansner, E.R., North, S.C., 2000. An open graph visualization system and its applications to software engineering. Softw. Pract. Exp. 30, 1203–1233.
Gent, I.P., Jefferson, C., Miguel, I., 2006. Minion: a fast, scalable, constraint solver. In: Proceedings of the 17th European Conference on Artificial Intelligence. IOS Press. Available at https://constraintmodelling.org/minion/download/.
Gerasimou, G., 2018a. Indecisiveness, undesirability and overload revealed through rational choice deferral. Econ. J. 128, 2450–2479.
Gerasimou, G., 2018b. On the indifference relation in Bewley preferences. Econ. Lett. 164, 24–26.
Gerasimou, G., 2024. Characterization of the Jaccard dissimilarity metric and a generalization. Discrete Appl. Math. 355, 57–61.
Gerasimou, G., Tejiščák, M., 2018. Prest: open-source software for computational revealed preference analysis. J. Open Source Softw. 3. Available at https://github.com/prestsoftware/prest/releases.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1, 114–125.
Hitt, M.P., 2019. Inconsistency and Indecision in the United States Supreme Court. Ann Arbor: Michigan University Press.
Houtman, M., Maks, J.A., 1985. Determining all maximal data subsets consistent with revealed preference. Kwant. Methoden 19, 89–104.

Jaccard, P., 1901. Étude Comparative de la distribution Florale dans une Portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37, 547–579.
Kalai, A.T., Nachun, O., Vempala, S.S., Zhang, E., 2025. Why language models hallucinate. arXiv Working Paper arXiv:2509.04664.
Kreps, D.M., 2012. Microeconomic Foundations I: Choice and Competitive Markets. Princeton: Princeton University Press.
Levandowsky, M., Winter, D., 1971. Distance between sets. Nature 234, 34–35.
Levy, H., 2016. Stochastic Dominance: Investment Decision Making under Uncertainty. Heidelberg: Springer. 3rd edition.
Mandler, M., 2009. Indifference and incompleteness distinguished by rational trade. Games Econ. Behav. 67, 300–314.
Marczewski, E., Steinhaus, H., 1958. On a certain distance of sets and the corresponding distance of functions. Colloq. Math. 6, 319–327.
Nightingale, P., Akgün, O., Gent, I.P., Jefferson, C., Miguel, I., 2014. Automatically improving constraint models in Savile row through associative-commutative common subexpression elimination. In: Proceedings of the 20th International Conference on Principles and Practice of Constraint Programming (CP 2014). Springer, pp. 590–605.
Ong, Q., Qiu, J., 2023. Paying for randomization and indecisiveness. J. Risk Uncertain. 67, 45–72.
Online Encyclopedia of Integer Sequences, 2021. Entries A000670, A001035 and A000798. Available at https://oeis.org.
Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. 20, 873–922.
Reutskaja, E., Nagel, R., Camerer, C.F., Rangel, A., 2011. Search dynamics in consumer choice under time pressure: an eye-tracking study. Am. Econ. Rev. 101, 900–926.
Richter, M.K., 1966. Revealed preference theory. Econometrica 34, 635–645.
Schmeidler, D., 1969. Competitive equilibria in markets with a continuum of traders and incomplete preferences. Econometrica 37, 578–585.
Schwartz, T., 1976. Choice functions, “rationality” conditions, and variations of the weak axiom of revealed preference. J. Econ. Theory 13, 414–427.
Selten, R., 1991. Properties of a measure of predictive success. Math. Soc. Sci. 21, 153–167.
Sen, A., 1971. Choice functions and revealed preference. Rev. Econ. Stud. 38, 307–317.
Sen, A., 1997. Maximization and the act of choice. Econometrica 65, 745–779.
Simon, H.A., 1956. Rational choice and the structure of the environment. Psychol. Rev. 63, 129–138.
Stoye, J., 2015. Choice theory when agents can randomize. J. Econ. Theory 155, 131–151.
Tversky, A., Shafir, E., 1992. Choice under conflict: the dynamics of deferred decision. Psychol. Sci. 3, 358–361.