Rational Choice Overload*

Lucas de Lara
Department of Economics, Columbia University
lpd2122@columbia.edu

Mark Dean
Department of Economics, Columbia University
mark.dean@columbia.edu

July 2, 2025

Abstract

We present and experimentally test a collection of search theoretic explanations for 'choice overload', the phenomena by which a default alternative is selected more often in larger choice sets. A standard search model, with constant search costs and a known distribution of item quality, cannot give rise to choice overload. If one instead assumes that either (i) the Decision Maker (DM) must learn the quality distribution (ii) search costs are increasing or (iii) the DM decides the search strategy in advance, then choice overload can occur. Unlike existing models, our approach does not require ad hoc psychological costs (decision avoidance), or for the DM to assume the choice set was selected by a profit maximizing firm (contextual inference). Data from a laboratory experiment are consistent with choice overload caused by search with learning and increasing costs, and cannot be explained by decision avoidance or contextual inference.

JEL Classifications:
Keywords: Bounded Rationality, Search, Choice Overload

*We thank the members of the Cognition and Decision Laboratory at Columbia University and seminar audiences at Chicago, Pitt, NYU, Princeton, Stanford, Berkeley Haas, UCLA, University of Queensland, Harvard, the Pan-Asian Theory Seminar, the MiddExLab Virtual Seminar, the Zurich Workshop on Economics and Psychology and the Economic Science Association for helpful comments. Also to Pietro Ortoleva and Demian Pouzo for their invaluable advice. We recognize the financial support provided by the Columbia Experimental Laboratory for Social Sciences (CELSS).

1 Introduction

The standard model of utility maximization tells us that increasing the set of available options can make the consumer no worse off, and may well improve their welfare. Since the pioneering study of Iyengar and Lepper (2000),[1] a body of work in psychology has called this assumption into question. The umbrella term "choice overload" covers a number of phenomena by which larger choice sets appear to make people worse off. By now there is a large literature investigating various aspects of choice overload; see Scheibehenne et al. (2010) and Chernev et al. (2015) for recent reviews and meta analyses.

Some of the measures used to identify choice overload are hard to interpret using the classic tools of economic analysis; examples include a reduction in ex-post reported satisfaction or lower confidence that the right choice was made.[2] Others fall very much in the realm of choice theory. In this paper we focus on the observation that larger choice sets may make people more likely to choose a "default" option, as in the famous 'jam' study of Iyengar and Lepper (2000). While previous literature has questioned the reliability of this result,[3] recent work by Dean et al. (2022) has developed more powerful tests, and established that choice overload of this type is likely to be more widespread than previously thought.

The aim of this paper is to develop a family of models that can explain choice overload[4] using an optimal search framework, and test them experimentally. We establish that in a standard search model, in which the cost per item searched is constant and the distribution of item quality is known,[5] optimal behavior can never give rise to choice overload. However, simple modifications to this basic framework can lead to the default option being chosen more often in a larger choice set - specifically (i) if the

[1]See also Reibstein et al. (1975).

[2]Though see recent work by Enke and Graeber (2023)

[3]Some direct replications of previous experiments have failed (Scheibehenne, 2008; Greifeneder et al., 2010). One recent meta-analysis concluded that the mean measured choice overload effect is zero (Scheibehenne et al., 2010). Another one (Chernev et al., 2015) concludes that whether or not choice overload exists may depend a lot on context.

[4]Despite the above discussion, from now on we will use 'choice overload' to refer specifically to the phenomenon of choosing a default option more often in larger choice sets.

[5]i.e. a model in the style of Caplin et al. (2011).

decision maker (DM) must learn about the distribution of item quality (ii) if search costs are increasing with the number of items searched or (iii) if the DM decides upon the number of items to search in advance, rather than dynamically updating their strategy.[6] We show how these models can be behaviorally distinguished, both from each other and from existing models of choice overload. Finally, we run an experiment in which we can observe subjects' search and choice behavior, and establish that search-based mechanisms are an important component of choice overload.

To our knowledge, there are currently two main classes of model used to explain choice overload. The first are models of 'contextual inference' (Kamenica, 2008; Kuksov and Villas-Boas, 2010; Nocke and Rey, 2021), in which the DM makes inferences about the nature of a set of alternatives based on its size. Typically these are driven by the assumption that the choice sets are chosen by a profit maximizing firm. The second are models based on the psychological concept of 'decision avoidance' (Beattie et al., 1994; Dean, 2008; Gerasimou, 2018), by which a decision maker avoids engaging with large choice sets because they find it aversive to do so.[7]

Our approach offers potential advantages over these frameworks. Relative to contextual inference, our model can lead to choice overload even if the alternatives are ex-ante identical in large and small choice sets. It does not require beliefs about the distribution of alternatives to vary with set size, and in particular does not require the assumption that choice sets are chosen by profit maximizing firms. This is important, as choice overload has been observed in situations in which this is clearly not the case (see for example Dean et al. (2022)). Relative to models of decision avoidance, our approach offers a rational basis for choice overload, without relying on ad hoc behavioral forces.

In order to establish the relationship between search and choice overload we begin by developing a framework that links a general model of sequential search to observables. We consider two different types of stochastic data. The first, which we term

[6] As in the model of Stigler (1961)

[7] Regret has been proposed as a driving force behind choice overload (see for example Buturak and Evren (2017)). While somewhat harder to fit into our framework, we see this as essentially working in a manner similar to decision avoidance.

the 'experimental data set' records the sequence in which alternatives are searched, the point at which search stops and the alternative that is chosen. This is the type of data we collect in our experiment, and is necessary to differentiate between models of choice overload. The second is standard stochastic choice data, which we use to define choice overload: We say that a data set exhibits choice overload if there exists a subset and a superset such that the default is chosen more often in the latter than the former.

The model we consider consists of two elements: a utility function over alternatives and a threshold, which can be a function of the search history and the number of alternatives left to search. The DM searches through alternatives one by one. After each item is searched, they compare the utility of all items seen so far to their threshold. If a previously seen item has a higher utility than the current threshold, search stops and the best item seen so far is chosen. If not, search continues and the same process is repeated after the following item has been searched.[8]

Within this general framework, we can identify different models of optimal sequential search as restrictions on the class of allowable threshold function, and so determine which can give rise to choice overload. We begin with the most straightforward of such models, in which the DM faces a fixed cost to search each alternative, and assumes that the value of each unsearched alternative is drawn from a known, fixed distribution. It has been previously shown (Weitzman, 1979; Caplin et al., 2011) that optimal behavior in this setting involves a fixed threshold which is invariant both to the size of the choice set and the number of alternatives left to search. Using our framework, we show that this model cannot give rise to choice overload.

We next consider three variants of this basic search model which can lead to the default being chosen more often in larger choice sets. First, the case in which the DM has search costs that increase the more alternatives have been searched - for example due to fatigue. This gives rise to an optimal threshold which falls the longer the DM searches, and that such a model can give rise to choice overload. Second, a model in which the DM does not engage in dynamically optimal behavior, but instead

[8] A similar set up is considered in Aguiar et al. (2016) and Aguiar and Kimya (2019), but with a threshold that is not a function of search history.

selects the number of options they will search in advance. This is the 'simultaneous search' model of Stigler (1961), and has been shown in some cases to better fit search behavior than the standard dynamic version of sequential search (De Los Santos et al., 2012). This model leads to a maximal number of alternatives the DM will search, which again can give rise to choice overload.

Perhaps the most interesting of the model variants is one in which the DM must learn about the distribution of alternatives from which they are drawing. Consider a situation in which a DM believes there to be two different types of jam shop: good and bad. If, when they begin to search, they see a number of low quality jams, they may come to believe that they are in a bad shop, meaning the value of further search is low. This may lead them to leave the store without buying any jam, even if there are, in fact, good jams available. This is the key mechanism of the learning model: search can stop because observed alternatives are *bad*, whereas in the standard search model, only high quality alternatives can stop search. This can in turn lead to choice overload if a large choice set is created from a smaller one by adding low quality alternatives.

We believe the learning channel to be of particular interest because it seems that the key feature of distributional uncertainty is common in many practical and experimental settings in which choice overload has been observed. A participant in Iyengar and Lepper's jam experiment does not know how the jams have been selected. Neither do the employees facing menus of health insurance options, as studied in Abaluck and Gruber (2023) know the rule that has determined which plans they have been offered.[9] It seems natural in such settings that a DM will learn about the quality of a choice set by sampling from it, and so may get discouraged if they see low quality alternatives. This is a feature missing from all existing models of choice overload we know of.

We show how our experimental data set can differentiate between search-based explanations for choice overload, and models of decision avoidance and contextual inference. Search models imply that choice overload is caused by people starting to

[9]The available plans have been selected for them by their employer, rather than determined by a market equilibrium, so the logic of contextual inference does not apply.

search, but stopping before they find an item which is better than the default alternative. This has a number of implications. For example, the probability of choosing the default will increase in the position of the first above-default alternative. In contrast, contextual inference and decision avoidance predict that choice overload is caused by people failing to start searching in larger choice sets, and the probability of choosing the default is independent of the position of the first above-default alternative.

The same data also allows us to test for learning, fatigue and simultaneous search as a driver of overload within the search framework. The learning model implies that, controlling for the position of the first high quality alternative, the probability of default choice should depend on the *beliefs* a Bayesian decision maker would hold about the quality of the choice set, as well as the number of alternatives left to search. The fatigue model predicts that the position of the first above-default alternative, but not beliefs, should matter. The simultaneous search model predicts the same thing, but additionally predicts that probability of stopping search should not depend on the value of alternatives so far seen.

In order to test these predictions, we run an experiment on 621 participants using the Prolific platform. The basic design is similar to that of Dean et al. (2022). Subjects must choose between alternatives which are numeric values described as sums (five addition or subtraction operations). The monetary value of the alternative is given by the value of the sum. This creates an environment in which the value of alternatives is clear to the outside observer, but challenging for the subject to work out. Each choice set consists of a default option, and either 1, 10, 15 or 20 other options. The default option always appears at the top of the screen, has the same value, and is represented by a degenerate sum.

We make two significant departures from the set up of Dean et al. (2022). First, we control the order in which the subject must search through alternatives in any given choice set, and make it observable when search has stopped, thus generating data equivalent to the 'experimental data set' described above. Second, we make explicit to the subjects the distribution from which the alternatives are being drawn. In our main treatment subjects are told that the values in each choice set are drawn from one of two distributions, one of which stochastically dominates the other. As far as we know, ours is the first experimental test of choice overload to explicitly control subject's beliefs.

While our experiment is unusual in making distribution uncertainty explicit, as we argue above we believe it likely that such uncertainty is implicit in essentially all choice overload experiments: Iyengar and Lepper (2000) did not describe the distribution from which their jams were drawn, so presumably subjects faced uncertainty about it, which could be partially resolved by looking at the quality of the jams so far searched. The same is true for all other experiments that we know of.

We make 4 key observations based on our results. First, we find evidence of choice overload, replicating the findings of Dean et al. (2022). Second, we find significant evidence for search based choice overload: most choice overload is driven by subjects who start searching, then stop before finding something better than the default. People are also more likely to choose the default if the first above-default alternative appears later in the search sequence, and to engage in 'partial' search of choice sets in which there are no above-default alternatives. All these are consistent with search based models, but not decision avoidance or contextual inference. Third, controlling for the position of the first above-default alternative, people are more likely to choose the default in rounds which should lead the subject to believe the set is of low quality, providing evidence of a learning channel. Fourth, search position, beliefs and the highest value alternative seen so far are all predictive of search termination, suggesting that both fatigue and learning play a role in choice behavior, while ruling out the simultaneous search model.

The fact that search plays a significant role in choice overload has implications for the optimal design of choice set - i.e. choice architecture. It implies that the order in which alternatives are presented matters for whether or not the default will be chosen. Moreover, the learning model implies that content of a given choice set matters, over and above ex ante beliefs. These observations can inform designs aimed at encouraging active choice - see for example Besedeš et al. (2015).

2 Relation to the Literature

There is, by now, a large empirical literature documenting the circumstances under which choice overload does and does not occur. For recent meta analyses see Scheibehenne et al. (2010) and Chernev et al. (2015). These reviews offer mixed support for choice overload. The former finds a mean effect of zero, but noted a high degree of variance. The latter identifies four variables which can increase the incidence of choice overload: decision difficulty, choice set complexity, preference uncertainty, and decision goal. Chernev et al. (2015) argue that, taking these mediators into account, there is evidence of a robust choice overload effect. The fact that overload is more likely to occur when choices are difficult and unfamiliar is consistent with an explanation rooted in bounded rationality.

Recent work by Dean et al. (2022) suggests that choice overload may be more prevalent than previously assumed. They show that previous tests are underpowered, and introduce a new approach that reveals overload where existing approaches do not.

Most previous models of choice overload fall into one of two categories: contextual inference (Kamenica, 2008; Kuksov and Villas-Boas, 2010; Nocke and Rey, 2021) and decision avoidance (Beattie et al., 1994; Anderson, 2003; Dean, 2008). The former proposes that the DM can make inferences about the quality of items in a choice set by the number of alternatives it contains. Choice overload occurs because a larger range of alternatives is assumed to signal a lower expected quality (or lower probability of fit for a consumer) for any given alternative, reducing the returns to search. This assumption is often motivated by showing that, in equilibrium, choice selections provided by profit maximizing firms will have this feature. It is worth noting that in our experiment, and many others, choice overload occurs in situations in which choice sets are clearly not chosen by profit maximizing firms.

Decision avoidance models assume that a decision maker has a choice between engaging with a given decision, and therefore potentially making an active choice, or instead avoiding the decision and sticking with the default alternative. Choice overload comes about because the cost of engagement is assumed to increase with set size. However, exactly what these costs are is typically not modeled. An exception is Natan

(2025), which proposes a model of choice overload based on rational inattention, with the assumption that attention costs are increasing in set size. Our approach instead microfounds the cost of engagement through an optimal search process.

Section 4.5 describes how our data can be used to differentiate between search, contextual inference, and decision avoidance based explanations for choice overload.

Our work also relates to the set of decision theoretic papers that have characterized how people choose from ordered lists of alternatives (Rubinstein and Salant, 2006; Horan, 2010; Aguiar et al., 2016). Perhaps the closest to our work is Manzini et al. (2019), from which we take some of our framework and notation. However, all of these papers have very different aims and results. None of them address the question of choice overload, and none aim to characterize behavior resulting from optimal search under different assumptions.

To investigate the effect of beliefs on choice overload, we introduce and solve a novel version of the sequential search model with learning. While several papers have tackled this problem (for example Rothschild (1974); Rosenfield and Shapiro (1981); Burdett and Vishwanath (1988); Koulayev (2013); De Los Santos et al. (2017); Conlon et al. (2018)), these have all used one or more assumption that make them unsuitable for our purposes: no recall, infinite horizon, or assumptions that guarantee myopic solutions (for example through the use of Dirichlet priors). Given that our aim is to study how the possibility of terminating search and choosing the default varies with the size of the choice set, we solve a model with recall, variable finite horizon and which allows for non-myopic solutions.

Finally, our work is related to the extensive empirical and smaller experimental literature on search behavior (see Honka et al. (2019) for a review of the former and Cox and Oaxaca (2008) and Charness and Kuhn (2011) for reviews of the latter). Broadly speaking, these papers provide evidence that the channels we highlight (learning, increasing search costs, simultaneous search) may be relevant in at least some settings, while not connecting to the concept of choice overload. In the field Honka and Chintagunta (2017) show that price search by consumers' behavior is well approximated by simultaneous search,[10] Hodgson and Lewis (2023) provide evidence for learning during the process of search, while Ursu et al. (2023) find data consistent with increasing search costs from fatigue.

Three experimental papers are of particular relevance to our study. Cox and Oaxaca (2000) and Casner (2021) use laboratory experiments to explicitly study search behavior in settings with distribution uncertainty. Both find evidence that uncertainty matters, in the former case through violations of the reservation utility strategy and in the latter through declining average reservation levels. Both use an environment of search without recall, meaning they cannot address the question of whether search can lead to an increase in choice overload. Brown et al. (2011) study an environment without learning, and in which the optimal strategy is for stationary reservation levels, yet find strong evidence that reservation levels are decreasing. These reservation levels 'bounce back' between rounds, making it hard to attribute them to fatigue or a misapplication of learning. The authors instead prefer an explanation based on an increasing subjective cost of waiting time during a spell. This is a potential additional source of search based choice overload which we do not study in this paper.

3 A General Framework

The theoretical section of the paper has two aims: to identify the assumptions under which search models can lead to choice overload, and to understand the testable implications of these models for our experimental data. To do so, we first define two data sets - one which matches the data that we will collect from the experiment, and a second which is closer to the standard data sets in which choice overload is identified. We show how these data sets are related, and use them to define choice overload in our setting. Next, we introduce a general class of sequential search models which explain choice behavior via a threshold rule and utility function, and show how to link this class to our data. Finally, we will consider specific versions of the optimal search model, show how they can generate choice overload, and identify testable predictions

[10] Although this runs contrary to evidence on search through choice sets in the lab, for example Caplin et al. (2011) and the data in this paper

for our experimental data.

3.1 Data

We begin with a finite grand set of alternatives X. One of these alternatives $d \in X$ is the default. A choice set is a subset of X that contains d. We use $\mathcal{A}$ to denote the set of choice sets

$$\mathcal{A} = \left\{ A \in 2^X | d \in A \right\}.$$

In our experiment, subjects will examine the alternatives in the choice set in a given order. Let $\mathcal{L}_A$ be the set of linear orders on A such that d appears first.

We denote the set of all such linear orders as $\mathcal{L}$

$$\mathcal{L} = \cup_{A \in \mathcal{A}} \mathcal{L}_A.$$

We will refer to an element $l \in \mathcal{L}$ as a list. We will sometimes abuse notation by writing l as $(xyz..)$ for the linear order $x \rhd_l y \rhd_l z...$ Further, we will use l^x to refer to the linear order l truncated at observation x.

Our experimental data will allow us to observe the list from which the DM must choose, the point at which they stopped searching through the list and the item they selected. Given that we wish to aggregate data over individuals that may have different behavior, we will consider a stochastic version of such data, and define our experimental data set as follows:

Definition 1. *Let $\mathcal{D} \subset \mathcal{L}$ be a collection of choice lists, with $A(l)$ being the collection of objects in list l (i.e. $l \in \mathcal{L}_A$). We define an* ***experimental data set*** *as $\rho : X \times X \times \mathcal{D} \to [0,1]$ such that*

1. $\rho(x, y, l) = 0$ *if* $x \notin A(l)$, $y \notin A(l)$ *or* $y \rhd_l x$
2. $\sum_{y \in X} \sum_{x \in X} \rho(x, y, l) = 1 \ \forall \ l \in \mathcal{D}$

We define the associated marginal distributions as follows

$$\begin{aligned}\hat{c}(x,l) &= \sum_{y\in X}\rho(x,y,l)\\ \hat{s}(y,l) &= \sum_{x\in X}\rho(x,y,l)\end{aligned}$$

The interpretation is that the DM faces a number of choice lists, indexed by the elements of $\mathcal{D}$, with l being a typical choice situation. From each choice list we observe the joint probability of choosing each element (the first argument in ρ) and stopping search at each element (the second argument in ρ). The restriction (1) insists that positive probability is given only to situations which are possible - i.e. that the chosen and finally searched element must both be in $A(l)$, and the chosen element must occur before the finally searched element in the list l. The distributions $\hat{s}$ and $\hat{c}$ represent the marginal stopping and choice probabilities for each list l.

While this data set matches well our experiment, choice overload is generally defined using a more standard choice data. We therefore define a standard choice data, and choice overload, as follows

Definition 2. *Let $\mathcal{C} \subset \mathcal{A}$ be a collection of choice sets. We define a* ***standard data set*** *as $p : X \times \mathcal{C} \to [0,1]$ such that $p(x,A) = 0$ if $x \notin A$ and $\sum_{x\in X} p(x,A) = 1$ for all $A \in \mathcal{C}$. We say that a data set* ***exhibits choice overload*** *if, for some $A, B \in \mathcal{C}$ such that $A \subset B$, $p(d,B) > p(d,A)$.*

Our standard choice data is essentially a classic stochastic choice data, with the added restriction that the default option d is available in all choice sets. Our definition of choice overload is similar to that used in Dean et al. (2022): A data set exhibits choice overload if there is a subset and superset such that the default is chosen more often in the latter than the former.

We can link together these two forms of data by defining the class of standard data sets which are consistent with some underlying experimental data set. In order to do so we have to consider the possible ways in which the decision maker could search through a set A. Because the models we consider below are silent on what determines search order, we will say that an experimental data set is consistent with a standard data set if there exists some probability distribution over possible search orders that lead to the same choice probabilities.

Definition 3. *We say an experimental data set $(\rho, \mathcal{D})$ is consistent with a standard data set $(p, \mathcal{C})$ if*

1. *For every $A \in \mathcal{C}$, $\mathcal{L}_A \subset \mathcal{D}$*

2. *For every $A \in \mathcal{C}$ there exists a $\pi_A \in \Delta(\mathcal{L}_A)$ such that, for every $x \in A$*

$$p(x, A) = \sum_{l \in \mathcal{L}_A} \pi_A(l) \hat{c}(x, l)$$

If this is the case, we say that $(\pi, \rho, \mathcal{D})$ generates $(p, \mathcal{C})$.

The first condition states that, for every choice set A in $\mathcal{C}$, we observe choice patterns from all possible search orders on A in $\mathcal{D}$. The second condition states that there must be a probability distribution over search orders in A such that, when combined with the choice probabilities from ρ, they match the probabilities in p.

3.2 Model

We now describe the class of models which is the focus of the paper. These are models of sequential search in which the DM's choice of whether to continue search is determined by a (potentially history dependent) threshold. Such models have two elements. The first is a one-to-one utility function $u \colon X \to \mathbb{R}$ which describes preferences over alternatives. The second is a choice threshold, which determines the DM's decision whether to stop searching or not. We allow this threshold to be a function of the alternatives that have been searched so far and the size of the choice set, which we assume to be known to the DM. Thus the threshold function is defined as $\tau : \mathcal{L} \times \mathbb{N} \to \bar{\mathbb{R}}$, with the interpretation that $\tau(l, N)$ is the threshold applied if the DM has observed the ordered alternatives l in a choice set they know to be of size

N.[11] Consistent with our desire to model stochastic choice data, we will allow for the decision maker to have a probability distribution over thresholds functions, which we denote as $T \in \Delta(\mathcal{T})$, where $\mathcal{T}$ is a finite feasible set of threshold functions.[12]

These elements can be used to define a model of sequential search with recall which gives rise to data in the form of our experimental data set. For each choice situation in $l \in \mathcal{D}$, search proceeds according to the list. The DM first draws a threshold function τ according to T. They then compare the utility of the first element they search (which will always be d) to the threshold $\tau(d, |A(l)|)$. If $u(d)$ is greater than the threshold, then search stops and d is chosen. If not, then the next item x in the ordering l will be searched. The DM then compares both $u(d)$ and $u(x)$ to $\tau(dx, |A(l)|)$. If either is above the threshold then search stops and the highest utility alternative seen so far is chosen. Otherwise search continues. We assume that, if the entire set has been searched then search stops and the highest utility item will be chosen. We can operationalize this assumption by adding the condition that $\tau(l, |A(l)|) = -\infty$.

In order to map this model to our experimental data set it is convenient to define two functions from the primitives of the model. The first is the stopping function $s_{\tau,u} : \mathcal{L} \to X$. The function $s_{\tau,u}(l)$ identifies the alternative which causes search to stop from l according to u and τ.

Definition 4. *For a utility function u and threshold function τ the stopping function $s_{\tau,u} : \mathcal{L} \to X$ is defined as $s_{\tau,u}(l) = x$ if*

- $\tau(l^x, |A(l)|) \leq u(y)$ *for some* $y \trianglerighteq_l x$
- *For all y, z such that $z \triangleright_l x$ and $y \trianglerighteq_l z$*

$$\tau(l^z, |A(l)|) > u(y)$$

[11]Note that this domain implies that the DM has a threshold for impossible situations, for example where they have seen three options yet the size of the choice set is 2. Such thresholds have no impact on choice behavior, as we shall see below.

[12]The assumption that $\mathcal{T}$ is finite is without loss of generality. Given the finite nature of X, for a utility function u there are a finite number of sets of threshold functions which give rise to different behaviors.

In words, the stopping function selects x if (i) there is something that is seen (weakly) before x that has utility above the threshold generated by the sequence of searches up to and including x and (ii) that this is not true for any z that is seen before x. Note that these conditions imply that s will be at most single valued, while the assumption that $\tau(l, |A(L)|) = -\infty$ implies that s is single valued.

Armed with the stopping function we can define the choice function $c_{\tau,u} : \mathcal{L} \to X$ which identifies the element from l that will be chosen according to u and τ.

Definition 5. *For a utility function u and threshold function τ the choice function $c_{\tau,u} : \mathcal{L} \to X$ is defined as*

$$c_{\tau,u}(l) = \arg \max_{x \trianglerighteq_l s_{\tau,u}(l)} u(x)$$

Armed with this function, we can define what it means for a data set to have a sequential search representation.

Definition 6. *For some $\mathcal{D} \subset \mathcal{L}$, an experimental data set $\rho : X \times X \times \mathcal{D} \to [0, 1]$ has a* ***sequential search representation*** *if there exist a utility function $u : X \to \mathbb{R}$, and a distribution over thresholds $T \in \Delta(\mathcal{T})$ such that, for every $l \in \mathcal{D}$*

$$\rho(x, y, l) = \sum_{\tau \in \mathcal{T}} T(\tau) 1(c_{\tau,u}(l) = x \text{ and } s_{\tau,u}(l) = y)$$

If this is the case, we say that the model $\{u, T\}$ generates the experimental data set $(\rho, \mathcal{D})$.

Next we define what it means for a model to cause choice overload

Definition 7. *We say that $\{u, T\}$ causes choice overload on $\mathcal{C} \subset \mathcal{A}$ if there exist an experimental data set $(\rho, \mathcal{D})$ and standard data set $(p, \mathcal{C})$ such that (i) $\{u, T\}$ generates $(\rho, \mathcal{D})$, (ii) $(p, \mathcal{C})$ exhibits choice overload and is consistent with $(\rho, \mathcal{D})$. We say that a model causes choice overload if it causes choice overload for some $\mathcal{C}$. We say a model causes choice overload with uniform search if the above holds, with the added condition that $(p, \mathcal{C})$ can be generated by $(\pi, \rho, \mathcal{D})$ with $\pi_A(l) = \frac{1}{|\mathcal{L}_A|}$ for all $A \in \mathcal{C}$*

and $l \in \mathcal{L}_A$. Finally, we say that a threshold distribution T causes choice overload if $\{u, T\}$ causes choice overload for some u.

We introduce the concept of 'choice overload with uniform search' because it is possible to construct situations in which the more general definition can lead to choice overload due to arbitrary and unmodelled changes in the search order between smaller and larger choice sets.[13] In the next section we will say a search process does not cause choice overload if it does not do so under the former definition, and will say it does cause choice overload if it does so under uniform search.

4 Optimal Models of Search

We now present models of optimal search behavior under various different assumptions. By characterizing the resulting threshold functions, we can then link these models to our data using the machinery from section 3. We first show that the 'standard' search model, in which the DM faces fixed search costs, optimizes dynamically and does not have to learn about the quality of a choice set, cannot generate choice overload. We then show how relaxing each of these assumptions can lead to choice overload, and demonstrate how these explanations can be differentiated in our experimental data.

4.1 The Standard Search Model

We start by considering the 'standard' sequential search problem as described by Caplin et al. (2011) and many others.

[13] As a simple example, consider a situation with choice alternatives d, x, y, z, such that

$$u(x) > u(y) > u(d) > u(z)$$

Consider also a search model in which the DM always searches the default and one other alternative. This can lead to choice overload when going from set $\{d, y, z\}$ to $\{d, x, y, z\}$ if one assumes that in $\{d, y, z\}$ y is always searched after d (and so the default is never chosen), while from $\{d, x, y, z\}$ z is always searched after d (so the default is always chosen). However, this model would not lead to choice overload in this data set when each search order is equally likely, because this would lead to $p(d, \{d, y, z\}) = \frac{1}{2}$ and $p(d, \{d, x, y, z\}) = \frac{1}{3}$.

Definition 8 (The Standard Search Problem)**.** *The decision maker faces a choice set of known size N. Initially they know the utility of the default alternative d. They assume that the utility of each alternative in the choice set is drawn independently from a probability distribution function f. The DM searches through alternatives one at a time. Following each search the DM can decide either to stop searching and choose one of the alternatives already seen, or (assuming there are still unsearched items) they can pay a fixed cost k to search one more alternative. Once search has concluded the highest utility alternative of those seen is chosen. The aim of the DM is to maximize the expected utility of the item finally chosen net of costs.*

It is well known that the optimal strategy for the standard search problem is a fixed threshold such that search stops if and only if an alternative with utility higher than that threshold is uncovered.

Remark 1. *For a given cost k, the optimal solution to the standard search problem can be characterized by a threshold function τ such that $\tau(l, N) = \tau_k$ for all non-terminal points (i.e. all l, N such that $|A(l)| \neq N$.)*

Proof. See Caplin et al. (2011). □

Our first result is that the optimal solution to the standard search problem cannot give rise to choice overload. To state this, and further results, we will say that a distribution over threshold functions T is *consistent* with an optimal model M if there exists a finite set of costs K[14] and a probability distribution P on K such that

$$T(\tau) = P(k|\tau(l, N) \text{ is optimal given costs } k \text{ and model } M)$$

In other words, T has a support which is a subset of the set of optimal threshold functions for the model M. We further say that a data set is consistent with a particular model if it can be generated by a T which is consistent with that model.

Theorem 1. *Let T be consistent with the solution to the standard search problem. Then T does not generate choice overload.*

[14] For the increasing search costs model this will be a finite set of cost *functions*.

The logic of the claim is as follows. Fixing a cost k, if a DM fails to choose the default in some small set A then it must be that they searched until they found something better. This implies (i) $u(d) < \tau_k$ and (ii) there exists some $x \in A$ such that $u(x) > u(d)$, which the DM found, either because $u(x) > \tau_k$, or because the set is completely searched. Both of these things must also be true in any superset $B \supset A$, so that means that the default also must not be chosen in the larger choice set.

We prove theorem 1 in appendix A as part of a more general result. We introduce the concept of 'search length dependent threshold functions', which depend only on the number of items that have been searched. Such functions cause choice overload if and only if they are decreasing. As the threshold functions generated by the standard search model is constant it cannot cause choice overload.

4.2 Increasing Search Costs

The first relaxation of the standard search problem we consider is one in which search costs increase over time - for example because the DM suffers from fatigue.

Definition 9 (The Search Problem with Increasing Costs)**.** *The search problem with increasing costs is identical to the standard search problem, but rather than a fixed cost k, the DM has instead a strictly increasing cost function $k : \mathbb{N} \to \mathbb{R}$, where $k(i)$ is the cost after searching i alternatives.*

First, we show that the optimal solution to the search problem with increasing costs is a threshold that decreases with the number of searches.

Lemma 1. *For a given increasing cost function $k(i)$, the optimal solution to the search problem with increasing costs can be characterized by a threshold function τ such that $\tau(l, N) = \hat{\tau}_k(|A(l)|)$ which is strictly decreasing in its argument*

See online appendix A for proofs.

The fact that the optimal threshold is decreasing as search continues guarantees that it is possible for the resulting model to generate choice overload.

Theorem 2. *Let T be consistent with the solution to the search problem with increasing costs. Then T can generate choice overload with uniform search.*

Consider a default utility $u(d)$, a cost function such that $\tau_k(1) > u(d) > \tau_k(2)$, and two alternatives x and y such that $u(x) > u(d) > u(y)$. In the set $\{d, x\}$ the default will never be chosen: because $u(d) < \tau_k(1)$ the DM will always search the first option, and so x will always be chosen over d. However, in the set $\{d, x, y\}$ the default will be chosen half the time under uniform search. Because $u(d) > \tau_k(2)$, exactly one non-default item will be searched. Half the time this will be x, in which case x will be chosen, but half the time it will be y, in which case d will be chosen.

Formally, the result follows from the fact that increasing search costs lead to decreasing search length dependent thresholds.

We can also use the sequential search representation to make predictions about what the search model with increasing costs implies for our experimental data set.

Theorem 3. *Let $\{\mathcal{D}, p\}$ be an experimental choice data which is consistent with the solution to the search problem with increasing costs. Then the following holds:*

1. *For any list $l \in \mathcal{D}$ which contains at least one alternative with utility above the default, the probability of choosing the default depends only on the position of the first such element.*

2. *For any list $l \in \mathcal{D}$ and alternative x in l, the probability that search has stopped at or before x*

$$\sum_{\tau \in \mathcal{T}} T(\tau) 1\left(s_{\tau,u}(l) \trianglerighteq_l x\right)$$

depends only on

   (a) The position of x in l

   (b) The maximal value of utility of items that occur at or before x in the list l

$$\max_{y \trianglerighteq_l x} u(y)$$

and is increasing in both values.

This theorem lays out the key markers of behavior we should see if increasing search costs are at the root of choice overload: The only determinant of whether the default is chosen is the position of the first above-default alternative in the list, and the whether or not search has stopped at a given point is determined only by how many alternatives have been searched and the maximum value that has been seen so far. This theorem will help us differentiate between increasing search costs and alternative explanations of choice overload in our experimental data.[15]

4.3 Search with Static Optimization

The standard search model assumes that DMs engage in dynamic optimization: their decision whether to continue to search or not is based on information they receive as they search. Early models of search assumed instead a model of static optimization: a DM chose in advance how many alternatives they would search through prior to beginning search (see for example Stigler, 1961). Such models are significantly easier to solve, and recent work has suggested that they may do a good job of describing search behavior in market settings (e.g. De Los Santos et al., 2012). We next define a model of search with static optimization and show that it can lead to choice overload.

Definition 10 (The Search Problem with Static Optimization)**.** *The search problem with static optimization is identical to the standard search problem, but rather than choosing a stopping function contingent on the values of the alternatives seen, the DM must choose in advance how many alternatives they wish to search.*

Any given cost level will then give rise to an optimal number of alternatives for the DM to search.

Remark 2. *For a given cost k, the optimal solution to the search problem with static optimization can be characterized by an integer n_k and a threshold function τ such*

[15] It is tempting to conclude that the increasing search cost model should lead to an increase in the probability of stopping as one proceeds down the list (rather than the cumulative probability of having stopped search). However this is not true. Imagine a cost function that rises from 0 to 1,000,000 between period 1 and 2, then to 1,000,001 in period three. Assuming a small variance in f, all search would stop in period 1, and almost none would stop in period 2.

that $\tau(l, N) = \hat{\tau}_k^s(|A(l)|)$*, with*

$$\hat{\tau}_k^s(|A(l)|) = \begin{cases} \infty, \text{ if } |A(l)| \leq n_k \\ -\infty, \text{ if } |A(l)| > n_k \end{cases}$$

for all non-terminal points (i.e. all l*,* N *such that* $|A(l)| \neq N$*.)*

Proof. See Stigler (1961). □

In turn, such a threshold function can give rise to choice overload.

Theorem 4. *Let* T *be consistent with the solution to the search problem with static optimization. Then* T *can generate choice overload with uniform search.*

The logic of the theorem is similar to that of theorem 2, while the formal result follows from the fact that static optimization leads to decreasing search length dependent thresholds.

The behavioral implications of the search problem with static optimization are similar to that of the model with increasing costs, with one important difference: the probability of search stopping should depend only on the position in the list, not the values seen so far.

Theorem 5. *Let* $\{\mathcal{D}, p\}$ *be consistent with the solution to the search problem with static optimization. Then the following holds:*

1. *For any list* $l \in \mathcal{D}$ *which contains at least one alternative with utility above the default, the probability of choosing the default depends only on the position of the first such element.*

2. *For any list* $l \in \mathcal{D}$ *and alternative* x *in* l*, the probability that search has stopped at or before* x *depends only on the position of* x *in* l*.*

4.4 Search with Learning

A third, and perhaps most interesting deviation from the standard model is the possibility that the decision maker does not know the distribution of values in a choice set before they begin searching. Imagine, for example, that a DM is searching through records in a record store. A priori, they do not know whether this is a high quality record store, with a high proportion of good records, or a low quality record store with relatively few good albums. We would expect the DM to update her beliefs about the quality of the record store as she searches, and this should affect her decision whether or not to continue searching: if they see a lot of bad records they may conclude that this is a low quality record shop, and so stop searching. As we discuss in the introduction, we think such distributional uncertainty is likely ubiquitous in situations in which choice overload has been observed.

In this section we formalize this intuition, and show how it can lead to choice overload. We do so in a relatively simple set up, in which the DM believes that the value of alternatives are being drawn from one of two possible distributions. This matches our experimental setting, and suffices to show that search with learning can cause choice overload. We also believe most of the insights we draw would generalize to more complex learning environments.

Definition 11 (The Search Problem with Learning)**.** *The search problem with learning is identical to the Standard Search Problem, but instead of believing that alternatives are drawn from a distribution f, the DM believes that the utility of all alternatives in a choice set are either drawn from distributions $f(x|\overline{\mu})$ and $f(x|\underline{\mu})$ with same support $[a, b] \subset \mathbb{R}$. They are both absolutely continuous, and satisfy the strict Monotone Likelihood Ratio property, with $f(x|\overline{\mu})$ first order stochastically dominating $f(x|\underline{\mu})$. Thus $f(x|\overline{\mu})$ is a "good" distribution, while $f(x|\underline{\mu})$ is a "bad" distribution. Initially, the DM assigns probability $\mu_0 \in (0, 1)$ that values are drawn from distribution $f(x|\overline{\mu})$.*

In the online appendix C we characterize the properties of the optimal solution to the search problem with learning, the key features of which are summarized in the following lemma.

Lemma 2. *For a given cost function k, the optimal solution to the search problem with learning can be characterized by a threshold function τ such that*

$$\tau(l, N) = \tau_k(\mu(l), n)$$

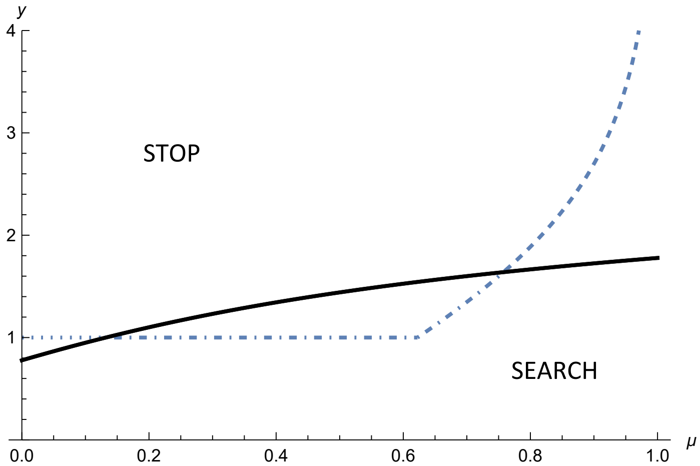


Figure 1: Optimal Strategy in the Search Problem with Learning

where $\mu(l)$ is the probability that the values in the set have been drawn from distribution $\overline{\mu}$, given sequence l and Bayes' rule, and $n = N - |A(l)|$ is the number of items remaining to be searched. $\tau_k(\mu(l), n)$ is weakly increasing in both arguments. We say a cost k guarantees search if $\tau_k(\mu, 1) > a$, the lower bound of the distribution support, for all μ. In this case, $\tau_k(\mu(l), n)$ is strictly increasing in its first argument.

The search problem with learning can generate choice overload. To see why, consider the following example. There are three alternatives d, x_1, x_2, and the problem is paramaterized such that $u(d) < \tau_k(\mu_0, 1)$, so that search begins in a set containing d and x_2. Assume further that $u(x_2) > u(d)$, implying that $p(d, \{d, x_2\}) = 0$. Now consider what happens if the DM is faced with the list dx_1x_2. Given that $\tau_k(\mu_0, 2) \geq \tau_k(\mu_0, 1)$, we know that search will continue until at least x_1. What happens at this stage will depend on both beliefs and $y = \max\{u(x_1), u(d)\}$. Figure 1 shows the optimal strategy in μ, y space for the case in which cost is $k = 1/8$ and the two distributions are normal with unit variance and means 0 and 1. The optimal strategy takes the form of a threshold, shown by the upward sloping solid black line. For combinations of μ and y that fall above the line search should stop, below the line it should continue.

Where the DM finds themselves in this space will depend on the value of $u(x_1)$. For $u(x_1) < u(d)$, increases in the value of $u(x_1)$ increase μ but not y: higher values make it more likely that values are drawn from the good distribution, but do not increase the value of the best thing seen. For $u(x_1) \geq u(d)$, increasing in the value of $u(x_1)$ increase both μ and y

The non-solid line on figure 1 shows the locus of possible points in μ, y space that can occur for different values of $u(x_1)$. It identifies three regions. For high values of $u(x_1)$, the dashed part of the line, search will stop: the DM believes it is likely that the alternatives are drawn from the good distribution, but the value of $u(x_1)$ is so high that further search is not worthwhile. For intermediate values of $u(x_1)$, the dash-dotted part of the line, search will continue: beliefs are high enough, and the value of the best thing seen low enough, that further search is worthwhile. However, for very low values of $u(x_1)$, the dotted part of the line, search again stops. Observing very low values of $u(x_1)$ leads the DM to believe it very unlikely that the values are being drawn from the good distribution. This in turn reduces the perceived value of search to the point that it is better to stop.

It is this third region that distinguishes the search problem with learning from the standard search problem, and can lead to choice overload. In the standard search problem, search can only be stopped because the DM observes a high value alternative. With learning, search can also stop because the DM observes a *low* valued alternative. In the above example, if $u(x_1)$ is low enough then the DM will stop searching because beliefs become low enough that the threshold for continuing to search falls below the utility of the default. This means that the superior alternative $u(x_2)$ is never found, resulting in choice overload.

This intuition is formalized in the following theorem.

Theorem 6. *Let T be consistent with the optimal solution to the search problem with learning with costs that guarantee search. Then T can generate choice overload with uniform search.*

Finally, we characterize the behavioral implications of the search model with learning. Ideally, we would like a behavioral characterization along the lines of theorem 3. However, unlike the search model with increasing costs, in which the threshold depends only on the number of items searched, the search model with learning implies the threshold depends on beliefs and the number of alternatives left to search. In order to characterize the behavior of this model along a single dimension, which will be convenient for our later empirical work, we define the notion of the *minimum cost* that induces search to stop by a particular point in a list.

Definition 12. *Let $l \in \mathcal{L}$ be a list and fix a utility function u. For any $x \in A(l)$, the minimum cost that induces stopping by x is defined as*

$$m_{l,u}(x) = \min\left\{k \in \mathbb{R} | s_{\tau_k,u}(l) \trianglerighteq_l x\right\}$$

where τ_k is the solution to the search problem with learning for costs k.[16]

This function combines the inputs to the stopping rule - beliefs and the number of alternatives remaining to search - into a single number. Note that, assuming u is observable and we know the process by which values are generated, we can calculate m for any l and x.

Using this definition, we can generate behavioral predictions for the search model with learning

Theorem 7. *Let $\{\mathcal{D}, p\}$ be consistent with the solution to the search problem with learning. Then the following holds:*

1. *For any list $l \in \mathcal{D}$ which contains at least one alternative with utility above the default, the probability of choosing the default depends only on $m_{l,u}(x)$ where x is the element immediately proceeding the first above-default element.*

2. *For any list $l \in \mathcal{D}$ and alternative x in l, the probability that search has stopped at or before x depends only on $m_{l,u}(x)$.*

4.5 Contextual Inference and Decision Avoidance

We have so far established that three variants of the standard search model can lead to choice overload: increasing search costs, static optimization and learning. In this

[16]That $m_{l,u}(x)$ is well defined comes from the fact that, for high enough costs, search will never start (and so the set is non empty), and assuming that, in the case of ties, search will stop (and so the set is closed).

section we contrast these search-based models of choice overload with the two current leading explanations: decision avoidance and contextual inference

Decision avoidance is a concept that has a long history in the psychology literature (see Beattie et al. (1994) and Tversky and Shafir (1992) for early examples and Anderson (2003) for a review), as well as a shorter one in economics and decision theory (see for example Dean (2008), and Gerasimou (2018, 2020)). It posits that, when faced with a difficult decision, people may exhibit a tendency to disengage with it - for example by postponing it or sticking with a default. Because large choice sets are often seen as complicated or difficult, this can lead to choice overload.

In order to capture the concept of decision avoidance within our framework, we consider a model in which the threshold function falls as the number of alternatives left to search increases. Unlike the other threshold functions we have discussed, this is not microfounded in some underlying search problem. However, we believe that it captures the idea of decision avoidance: when choice sets get larger, a DM is more willing to disengage from the decision by stopping search. In particular, this means that the DM will be more willing to stick with the default in larger choice sets.

Definition 13. *A threshold function τ exhibits **decision avoidance** if $\tau(l, N) = \tau^d(N - |l|)$ for all l, N, strictly decreasing in its argument for all non-terminal points.*

Introduced by Kamenica (2008) (see also Kuksov and Villas-Boas (2010) and Nocke and Rey (2021)), contextual inference assumes that DMs form beliefs about the quality of alternatives in a choice set based on the context, and in particular the number of alternatives. It is typically assumed that the DM forms beliefs based on the behavior of profit maximizing firms. Choice overload can occur if beliefs are such that larger choice sets are associated with lower perceived quality.

The following definition captures the concept of contextual inference by modifying the standard search problem to allow the belief distribution f to depend on the number of alternatives in the set.

Definition 14 (The Search Problem with Contextual Inference)**.** *The search problem with contextual inference is identical to the standard search problem, but instead of* *believing that alternatives are always drawn from a distribution f, the DM believes that the distribution depends on the size of the choice set, with f_n the distribution of alternatives if the set size is n. If $n < m$, then f_n strictly first order stochastically dominates f_m.*

It is easy to see that the optimal solution to the search problem with contextual inference will be to employ, in any given search problem, a fixed reservation threshold that depends on search costs and the size of the choice set.

Lemma 3. *For a given cost function k, the optimal solution to the search problem with contextual inference can be characterized by a threshold function τ such that*

$$\tau(l, N) = \tau_k^I(|N|)$$

Where τ_k^I is strictly decreasing in its argument

Our first result is that both decision avoidance and contextual inference can lead to choice overload.

Theorem 8. *Let T be consistent either with decision avoidance or contextual inference. Then T can generate choice overload with uniform search.*

While search-based explanations on the one hand and contextual inference and decision avoidance on the other both give rise to choice overload, the mechanism by which they do so is very different in a way that is observable in an experimental data set. As the following theorem makes clear, the former leads to choice overload if, in the larger choice set, search starts but stops before an alternative better than the default is found. In contrast, the latter case, choice overload only occurs because search never starts in the larger choice set.

Theorem 9. *Let the threshold function τ be consistent with either decision avoidance or contextual inference. Then for any two choice sets $A \subset B$, τ can generate choice overload only if $s_\tau(l) = d$ for every $l \in \mathcal{L}_B$. As a result, for any T that is consistent* *with either of those two models and any set A which contains an alternative better than the default*

$$p(d,B) - p(d,A) = \sum_{l \in \mathcal{L}_B} \pi_B(l)\hat{s}(d,l) - \sum_{l \in \mathcal{L}_A} \pi_A(l)\hat{s}(d,l)$$

where $\pi_A \in \Delta(\mathcal{L}_A)$ and $\pi_B \in \Delta(\mathcal{L}_B)$ are the distributions over sequences used to generate $p(.,A)$ and $p(.B)$ respectively.

Let the threshold function τ be consistent with either search with increasing costs, static optimization or learning. Then for any two choice sets $A \subset B$, τ can generate choice overload only if, for some $l \in \mathcal{L}_B$, $d \rhd_l s_\tau(l) \rhd_l x_l$ where x_l is the first above-default alternative in l. As a result, for any T that is consistent with either of these models and any set A such that $|A| = 2$ and which contains an alternative better than the default

$$p(d,B) - p(d,A) \leq \sum_{l \in \mathcal{L}_B} \pi_B(l) \sum_{y \in B | d \rhd_l y \rhd_l x_l |} \hat{s}(y,l)$$

where $\pi_B \in \Delta(\mathcal{L}_B)$ is the distribution over sequences used to generate $p(.,B)$.

A second behavioral difference between the two classes of models is whether list order is important. As previous theorems have demonstrated, under search based models, for sets that contain an above-default-alternative, the probability of choosing the default depends on the position of that alternative in the list. This is not true for models of contextual inference, as the following theorem demonstrates

Theorem 10. *Let $\{\mathcal{D}, p\}$ be an experimental choice data set generated by a model $\{u, T\}$ in which T is consistent with either decision avoidance or contextual inference. Then, for any set A which contains an above-default alternative*

$$\hat{c}(d,l) = \hat{c}(d,l') \text{ for all } l, l' \in \mathcal{L}_A$$

Specifically, the probability of choosing the default does not depend on the position of the first above-default alternative.

Finally, an obvious implication of these two models is that, if search begins in a set which contains only alternatives that are worse than the default, then it must continue until all items have been searched. This follows directly from the fact that the threshold is constant in the case of contextual inference, and weakly increasing in the case of decision avoidance.

Theorem 11. *Let $\{\mathcal{D}, p\}$ be an experimental choice data set generated by a model $\{u, T\}$ in which T is consistent with either decision avoidance or contextual inference. Then, for any set A which contains no above-default alternative, $l \in \mathcal{L}_A$, $\hat{s}(x, l) = 0$ unless x is either d or the last element in l.*

5 Experimental Design

We now describe the results of an experiment that allows us to compare search-based explanations for choice overload to decision avoidance and contextual inference, and to identify which of the various search channels are important in causing choice overload.

Our experimental design needed to satisfy a number of requirements. First, we needed choice alternatives that had a clear preference ranking if subjects fully internalized their values, yet required cognitive effort to uncover those values. Our test of choice overload relies on the former, while without the latter it would be too easy for the subject to simply choose the best option in each choice set. Second, we required a design that allowed us to control the order in which subjects searched, and observe their decision to stop searching. This is necessary to generate the experimental data set of definition 1. Finally, and unlike previous choice overload experiments, it was important for us to control the beliefs of the subject about the distributions from which values were drawn - as our theory demonstrates, these beliefs should be an important determinant of behavior.

In the following sections we explain how our design fulfills these criteria.

5.1 The Experimental Environment

The basic choice problem is similar to that in Caplin et al. (2011) and Dean et al. (2022): Subjects choose between options which represent different amounts of

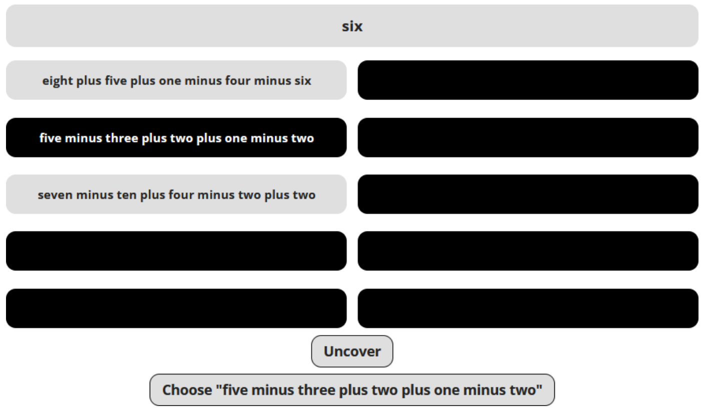


Figure 2: Typical Screenshot

experimental points. The value of each option is represented by a sum (written out in words) In each choice problem, the subject chooses between a default option (which has a value of 6 experimental points) and between 1 and 20 other alternatives. The value of the default option is known: it is the same in each choice problem, and it is displayed at the top of the screen as a degenerate sum. The value of the other alternatives are not initially known.

In order to allow us to identify which alternatives the subject has searched, all non-default alternatives are initially obscured. By pressing a button, the subject can uncover each alternative in a pre-determined order. We therefore know the order of search. The set of uncovered alternatives allows us to approximate the set of searched alternatives. Specifically, it provides an upper bound on the set. We can be sure that any alternative that was not uncovered has not been searched by the subject. It is possible that subjects uncover alternatives, but do not internalize their values (for example, they could keep clicking until they had uncovered all the alternatives, then process their values). In order to reduce this possibility, we introduce a short delay of about 3 seconds after each reveal before the next alternative can be uncovered.

When subjects are ready to make their choice, they can click on their preferred alternative and press 'choose'. Figure 2 shows a screenshot from a typical choice environment, in which some alternatives have already been uncovered.

At the end of the experiment, one choice problem was selected at random for each subject, and they were paid based on the option they chose in that situation. Experimental points were converted to money at a rate of \$0.50:1. This payment was added to a \$3 participation fee.

A full set of instructions is available in the supplemental materials.

5.2 Item Values and Beliefs

A key input into the theory from section 4 is the beliefs of the decision maker over the distribution of values from which choice alternatives are drawn. In most choice overload experiments these parameters of this process are not made explicit to the subject, so it is hard to know what these beliefs might be. In order to provide a controlled test of our theory we tell subjects exactly how the values of the alternatives they are choosing between are generated.

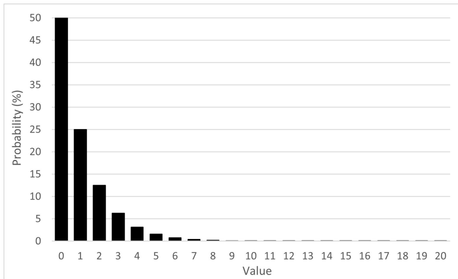
In section 4 we introduced three possible mechanisms by which search could generate choice overload - distribution uncertainty, increasing search costs and static optimization. In our baseline treatment we consider an environment in which subjects do in fact have uncertainty about the distribution from which values are drawn. This provides a situation in which all three of our identified channels could, in principle be a factor in causing choice overload. Given that previous demonstrations of choice overload have not provided information on how values were selected, we would also argue that decision uncertainty is a ubiquitous feature of such experiments. Including it in our study therefore makes it more comparable to previous work.

For completeness, we also run a second treatment in which we turn off distribution uncertainty, the results of which are reported in section 6.1.

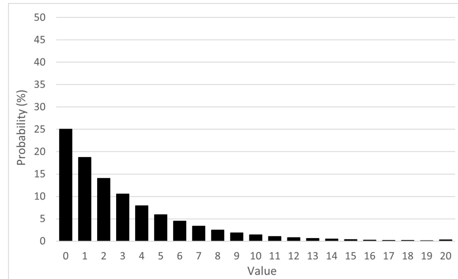
5.2.1 Baseline Treatment

In the baseline treatment subjects are told that the values of the alternatives are drawn from one of two distributions. Each distribution is geometric (truncated at 20). The 'good' distribution has $\lambda = 0.25$, while the 'bad' distribution has $\lambda = 0.5$.[17] The PDF of each of these distributions is shown in Figure 3. The same figures were

[17]In the instructions the distributions were referred to as A and B, rather than 'good' and 'bad'.



(a) ‘Bad’ Distribution



(b) ‘Good’ Distribution

Figure 3: The two possible distributions of values for the learning treatment

also included in the experimental instructions. Subjects were also made to draw values from the two distributions as part of the instructions.

Prior to each choice problem the subject was reminded that the values they faced in that choice problem would be drawn from one of these distributions, and a priori each was equally likely. The instructions made it explicit that all alternatives in each choice problem were drawn from the same distribution, but that different choice problems may come from different distributions. They also made it clear that subjects could update their beliefs about which distribution was more likely by observing the values they uncovered - i.e. that there was an opportunity for learning.

Because we are explicit about the procedure by which values were drawn, we can calculate the beliefs that a Bayesian subject should have following any observed sequence of draws. We make heavy use of this fact in the analysis that follows.

Prior to the start of the experiment, 100 values were drawn from each distribution. Sums were generated to equal each of these values using the procedure of Caplin et al. (2011). Each subject faced 10 choice problems: 4 with 2 options (the default plus one other alternative), 2 with 11 options, 2 with 16 options and 2 with 21 options. In each choice problem it was equally likely that the good distribution or bad distribution would be used. The relevant number of options was then drawn randomly from the chosen distribution, and presented in a random order. The order in which choice sets were presented was also randomized. Randomization occurred at the subject level, meaning each subject faced a different set of choice problems. This random assignment of choice problems will be important in later analysis.

5.2.2 No-Learning Control

The no-learning control proceeded exactly as the learning treatment. The only difference is that subjects were told that in each decision problem they would face values drawn from a single distribution which was the average of the two distributions A and B from the learning treatment. They were also told explicitly that this meant that the values they had so far uncovered would not help them predict the values of other alternatives in the choice set.

5.3 Implementation

The experiment was run on the Prolific platform. Subjects are restricted to those located in the USA and fluent in English.[18]

307 subjects in the baseline treatment and 314 in the no-learning control successfully finished the experiment. During the course of the instructions, subjects were asked to complete 5 comprehension questions. If they got any question wrong more than 2 times then they were routed away from the experiment and could not finish it. This happened to 71 and 72 subjects in the baseline and no-learning treatments respectively. Subjects also faced two practice rounds before the 10 'live' rounds that constitute the experiment.

6 Results

We use the data from the above experiment to answer six questions, and so tease out the role of the various different drivers of choice overload identified in section 4.

First we establish that there is indeed choice overload in our baseline environment, and so we have a suitable test bed to differentiate between different theories of choice overload.

[18]The experiment and subsequent analysis was pre-registered using the Open Science Framework osf.io/3jfpn

Table 1: Default Choice by Set Size

Set Size	**% Default Choice**	*N*
2	**9**	*53*
11	**38**	*211*
16	**41**	*273*
21	**44**	*274*

Percentage of sets that contained at least one item better than the default in which the default was chosen.

Second, we test whether the increase in choice of default in larger choice sets is matched by an increase in the frequency of observations in which subjects do no searching - a prediction of 'classic' models of choice overload (i.e. contextual inference and decision avoidance).

Third, we examine whether, in choice sets with no above-default alternative, there is significant partial search - i.e. some, but not all alternatives are searched - which is inconsistent with classic models of choice overload.

Fourth, we test for whether the position of the first above-default alternative is predictive in determining whether the default is chosen. Search based models allow for this behavior while classic models do not.

Fifth, we test for the impact of distribution uncertainty on default choice while controlling for the effect of fatigue through search order position.

Sixth, we use a behavioral model to estimate whether learning, the number of options searched, and the highest value alternative seen so far are determinants of the decision to stop searching, allowing us to differentiate between the three models of search proposed in section 4.

Choice Overload First, we determine the extent of choice overload in our baseline treatment. In order to do so, we build on the approach of Dean et al. (2022), and look for specific subsets and supersets such that default choice is higher in the latter than in the former.

Given that we know what the rankings of alternatives should be, we focus our attention on sets which contain an option which is better than the default. Sets which have no such option should have high default choice even in small choice sets, and so have little opportunity for choice overload. In table 1 we show the fraction of sets which contain an alternative better than the default in which the default is chosen. In lemma 5 in the online appendix D we show that, if a stochastic choice function does not exhibit choice overload then default choice in such sets should be non-increasing in set size.

Table 1 shows that default choice is between 29 and 35 percentage points (pp) higher in sets of size greater than 2 than it is in sets of size 2. In order to confirm statistical significance, table 2 reports the results of the regression

$$chose\_default_{i,j} = \alpha + \sum_{k \in \{11,16,21\}} \beta_k 1(size_{i,j} = k) + \varepsilon_{i,j}$$

where i is the subject, j is the round, $chose\_default_{i,j}$ is a variable that takes the value 1 if the default was chosen in that observation, $size_{i,j}$ is the size of the choice set in that observation and 1 is the indicator function. Again, we run the regression only on sets that contained an above-default alternative. Standard errors are clustered at the subject level. We report results from Ordinary Least Squares (OLS) regressions, though a logit specification gives equivalent results.

Table 1 confirms that we observe choice overload in our baseline specification. Default choice is higher in the size 11, 16 and 21 choice sets relative to the size 2 choice set at the 1% level. Tests of linear restriction reveal no difference in default choice probability between the size 11, 16 or 21 choice sets.

Choice Overload and Failure to Search A prediction of classic models of choice overload is that it is possible for above-default alternatives to not be chosen when they are available, but only if the subject fails to begin searching in a choice set. This means that the increase in default choice reported in table 1 should be matched by an increase in the fraction of choice sets in which search does not begin

Table 3 shows that this is not the case. In the first column it again reports the

Table 2: Regression Results

	Default Choice
S=11	0.285***
	(0.050)
S=16	0.316***
	(0.050)
S=21	0.347***
	(0.052)
Constant	0.094***
	(0.041)
R^2	0.028
N	811

OLS regression of default choice on set size dummies. Standard errors clustered at the subject level. * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

fraction of sets in which a better than default alternative was available but not chosen, broken down by choice set size. In the second it reports the fraction of such choice sets in which search does not begin. The increase in the former is not matched by an increase in the latter. Compared to sets of size 2, the increase in probability that no search occurs for larger sets is only 4%, while the increase in default choice is 32%.

Partial Search As demonstrated in theorem 11, classic models imply that, in sets with no above-default alternative, search should either never start, or be complete. Figure 4 shows the distribution of search lengths in such sets of size 11, 16 and 21. While in each case there is a mass of people who do search all the way through the choice set, there is a majority of observations in which subjects search some, but not all of the set: 59% for size 11, 64% for size 16 and 70% for size 21. These observations are consistent with search costs, static optimization or learning, but not with contextual inference or decision avoidance.

Table 3: Categorizing Choice Overload

Size	Choose Default	Search Never Starts	N
2	9%	2%	*53*
11	38%	7%	*211*
16	41%	5%	*273*
21	44%	5%	*274*
$s \geq 2 - s = 2$	32%	4%	

For each choice set size, reports the fraction of sets that contain an above-default option in which the default was chosen (Choose Default), and the fraction of sets in which search never starts.

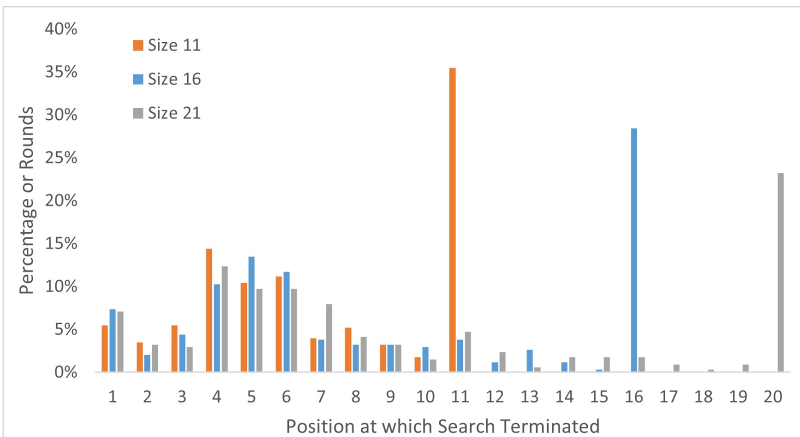


Figure 4: Histogram of search termination position for choice sets of size 11, 16 and 21

Search order effects As we show in theorem 10, search order matters for the choice of default in search models, but not classic models of overload. We can test this prediction due to the fact that each choice set in the experiment is generated randomly, creating variation in the order in which the first above-default alternative appears. Figure 5 shows the probability that the default is chosen, grouping sets by the position of the first above-default alternative. It shows a clear upward trend - default choice is more likely in sets in which the first above-default alternative appears later. This pattern is confirmed by regression analysis. An OLS regression of default choice on the position of the first above-default alternative, with standard errors clustered at the subject level, returns a coefficient significant at the 0.1% level

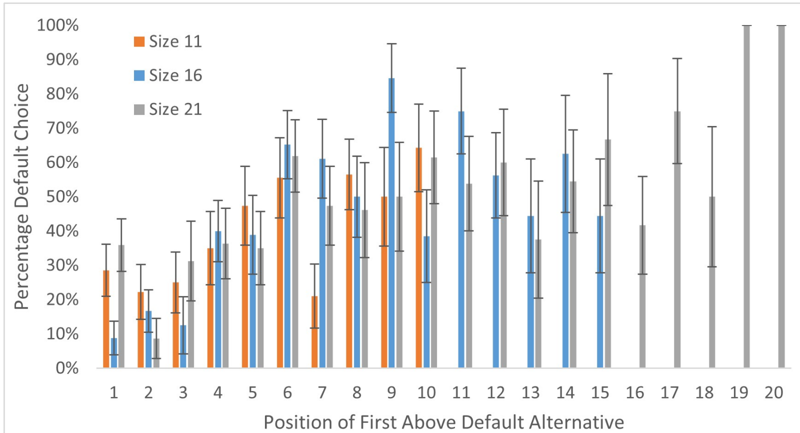


Figure 5: Effect of position of first above-default alternative on default choice. Bars show the probability of default being chosen. Error bars represent standard errors.

for all choice set sizes.

Separating Distribution Uncertainty from Increasing Costs The results so far indicate that search channels are important in determining default choice. However, they do not allow us to understand which features of search are important: both the increasing search costs and distribution uncertainty would predict that sets where the first above-default alternative appears later should see more default choice.

We address this issue by again making use of random variation in the search sets that subject's face. Consider two sets in which the first above-default alternative appears after position k, but which contain different values prior to that. Set A contains lower values in these early positions than does set B. As per theorem 3, the increasing search cost model would imply that, on average, default choice should be the same in these two sets - the only thing that determines default choice is the position of the first above-default alternative. In contrast, the distribution uncertainty model would imply that default choice should be more likely in set A than set B, as the subjects would have more pessimistic beliefs by position k and so be more likely to quit searching.

In order to test for such effects, we run a regression of the form

$$chose\_default_{i,j} = \alpha + \beta min\_cost^k_{i,j} + \varepsilon_{i,j}$$

for $k = 1...15$. For each of these regressions, we use data on choice sets of size greater than K, and include all observations in our baseline treatment for which there is (a) an alternative that is above the default value that appears after position k and (b) the choice set is generated using values from the high quality distribution. $min\_cost^k_{i,j}$ is the minimum cost variable introduced in definition 12, which records the lowest cost such that an optimal decision maker would stop searching at or before position k. As before we estimate this model using OLS with errors clustered at the subject level.[19]

Recall that a high value of the minimum cost variable means that only people with high search costs should stop searching before this point. Thus, we would anticipate a *negative* relationship between minimum costs and default choice.

The minimum cost variable combines information on both beliefs and the number of items remaining to be searched derived through an optimizing model. For transparency we repeat our analysis with a 'minimum beliefs' variable, which records the lowest level of belief that values are being drawn from the high quality distribution prior to point k. Again, we would expect to see a negative relationship between minimum beliefs and default choice.

Table 4 reports the results of these regressions. It shows that both minimum beliefs and minimum costs are significantly related to default choice, predominantly in later rounds.

Determinants of Stopping Behavior The analysis of section 4 makes it clear that, to differentiate between the three variants of the search model, it is useful to know the determinants of the decision to stop searching. In order to perform such

[19]In the preregistration plan we mistakenly reported that we were intending to use a discretized version of the minimum cost variable. This provides similar results. For completeness we have included these results in table 6 in the online appendix E.

Table 4: The Effect of Distribution Uncertainty

	Min. Beliefs			Min. Costs		
k	**Coeff**	**s.e.**	**N**	**Coeff**	**s.e.**	**N**
1	-0.14	0.10	650	0.61	0.67	650
2	-0.37*	0.20	564	-0.67	0.48	564
3	-0.51	0.32	508	-0.59	0.40	508
4	-0.61	0.45	436	-0.44	0.39	436
5	-0.35	0.60	379	0.00	0.38	379
6	-0.49	0.80	317	0.04	0.39	317
7	-1.37	1.06	261	-0.29	0.38	261
8	-2.92**	1.22	207	-0.59	0.39	207
9	-3.58**	1.50	172	-0.84**	0.40	172
10	-3.85*	1.97	132	-1.14***	0.43	132
11	-5.22**	2.32	107	-1.65***	0.46	107
12	-8.99***	2.50	81	-1.78***	0.51	81
13	-8.10*	4.50	64	-1.66***	0.54	64
14	-11.37*	6.07	45	-1.98***	0.58	45
15	-11.03	6.82	30	-1.61**	0.72	30

Each line reports the coefficient of a regression of default choice on minimum beliefs (left hand panel) and minimum costs (right hand panel) measured at period k, looking only at sets in which the first above-default alternative appears after position k. * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

analysis, we will reformulate our data. Recall, that, for each choice set, we observe the sequence of alternatives that the subject could see, and the point at which they stopped searching. We will now take an observation to be stopping behavior when faced with the kth alternative in the search order shown to subject i in round j.

In order to understand the determinants of the subject's decision to stop searching we run the following regression

$$stop_{i,j}^{k} = \alpha_i + \beta_1 above_{i,j}^{k} + \beta_2 k + \beta_3 min\_cost_{i,j}^{k} + \varepsilon_{i,j}^{k} \tag{1}$$

where $stop_{i,j}^{k}$ is a dummy that takes the value 1 if the subject has stopped searching at or before observation k,[20] $above^k_{i,j}$ is a dummy that takes the value 1 if the DM has seen an alternative above the value of the default at or before alternative k and $min\_cost^k_{i,j}$ is the minimum cost variable calculated for alternative k in choice set j for person i. We estimate the regression using OLS, including subject fixed effects and standard errors clustered at the subject level, using data from our baseline treatment. We restrict to observations which are not the last one in the search sequence (in which case the subject must de facto stop searching), and choice sets in which the subjects began the search process.

For transparency, we also consider a regression specification in which the determinants of the minimum cost variable - minimum beliefs and the number of items left to search - are included separately in the regression.

Distribution uncertainty would imply that the minimum cost variable (or minimum beliefs and items remaining to search) is negatively related to stopping. The increasing costs model would imply that k should be significantly and positively related to stopping, while the simultaneous search model implies that the 'above' variable should *not* be predictive of stopping behavior.

[20]The theory developed in section 4 tells us that 'stopping at or before k' is the correct dependent variable, not 'stopping at k'. To see why, consider the prediction of the increasing costs model. It is true that the likelihood of having stopped by k is increasing in k, controlling for whether something better than the default has been seen. However, it is not true that the likelihood of stopping at k is increasing in k. For example consider a cost function that induces a reservation level of infinity up to period t, and zero afterwards. This should induce everyone to stop in period t, and so the probability of stopping in this period is higher than it is in (say) $t + 1$.

Table 5: Determinants of Stopping Search

	Search Stopped	
	(1)	**(2)**
Min. Cost	-0.713***	
	(0.072)	
Min. Beliefs		-0.397***
		(0.041)
k	0.031***	0.029***
	(0.001)	*(0.002)*
Items Remaining		-0.007***
		(0.002)
Above	0.060***	0.084***
	(0.017)	*(0.017)*
Const	0.329***	0.322***
	(0.018)	*(0.026)*
R^2	0.542	0.537
N	24,182	24,182
Subject f.e.	Yes	Yes

Dependent variable is whether or not search has stopped by given observation. OLS regression with standard errors clustered at the subject level * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

Table 5 shows the results for both specifications. All of the dependent variables are highly significant, offering support for the distribution uncertainty and increasing costs channel, but not the simultaneous search model.

6.1 No Learning Control

In the online appendix E.1 we repeat the analysis above for the No Learning control treatment in which we attempt to switch off distributional uncertainty. One challenge in doing so is that a subject who has internalized our instructions should have beliefs that are invariant to the values they have seen. In order to provide a more interesting test, we construct minimum belief and minimum cost variables as if the subject's believed they were in the Baseline treatment. This allows us to test whether subjects are more likely to stop searching and choose the default if they have seen low value alternatives.

The results are broadly in line with what we would expect: overload still exists in the No Learning control treatment, but it is smaller, and appears less related to search and distribution uncertainty.

Tables 7 and 8 show that there is still choice overload in the No Learning control, though less than in the baseline treatment. Regression analysis confirms that there is significantly less overload in the control treatment, albeit due to higher default choice in the size 2 treatment. Table 9 shows that there is a higher fraction of rounds in which subjects do not start searching in sets of size greater than 2 compared to sets of size 2. This increase is larger than in the baseline treatment but smaller than the increase in default choice. Figure 7 shows that there is again a relationship between the position of the first above-default alternative and default choice, albeit weaker in the baseline treatment, while figure 6 shows a significant fraction of partial search, though again smaller than in the baseline treatment . Table 10 shows either no relationship or a positive relationship between minimum beliefs or minimum costs and choice of default, once the position of the first above-default alternative is controlled for. Table 11 shows that minimum beliefs and the number of alternatives remaining are not predictive of stopping search. Minimum costs are still predictive, but the relationship is much weaker than the baseline treatment.

6.2 Discussion and Conclusion

We have demonstrated, both theoretically and experimentally, that search theoretic concerns can drive choice overload. Both increasing search costs and uncertainty mean that, in a choice set with many mediocre options, a decision maker will fail to find a high quality alternative. While our results do not allow us to rule out heterogeneity in the population, we can show that the number of subjects who do not simply do not engage in larger choice sets - as predicted by existing models of choice overload - is relatively small.

These results lead to two further questions: How important are each of these forces in choice situations outside the laboratory, and what does this mean for optimal 'choice architecture' - or the design of the decision making environment? We believe that both are important topics for future research. Our expectation is that both fatigue and uncertainty should, if anything, be more important for 'real world' decisions. Take, for example, the job of selecting a retirement savings plan from a collection offered by an employer. It seems likely that users would have a great deal of uncertainty about how these plans had been selected, particularly those newly arrived at the company. Moreover each plan would require significant effort to understand. Thus we would expect the two channels we have identified to play an important role in their decision making.

A model incorporating the two elements we have identified could also be used to evaluate the welfare effects of different choice environments. At a basic level, it could be used to determine which and how many alternatives should be included in a choice set. Assuming that there is uncertainty or heterogeneity in the tastes of the consumer, then there is a trade off when adding options which may be good for some people, but bad for many. A model incorporating the forces we identify will give new answers to these questions. The same model could be used to evaluate more exotic options, such as ranking the alternatives (and therefore giving a free signal about their quality), or presenting alternatives in smaller groups (for example as in Besedeš et al. (2015)). Unique amongst the models we (and others) have considered, distribution uncertainty means that the composition of the choice set, and the order in which they are seen, can affect stopping behavior. This gives a richer set of tools for policy makers to consider when designing the choice environment.

References

Abaluck, Jason and Jonathan Gruber, "When less is more: Improving choices in health insurance markets," *The Review of Economic Studies*, 2023, *90* (3), 1011–1040.

Aguiar, Victor H and Mert Kimya, "Adaptive stochastic search," *Journal of Mathematical Economics*, 2019, *81*, 74–83.

Aguiar, Victor, Maria Jose Boccardi Chalela, and Mark Dean, "Satisficing and stochastic choice," *Journal of Economic Theory*, 2016, *166* (C), 445–482.

Anderson, Christopher J, "The psychology of doing nothing: forms of decision avoidance result from reason and emotion.," *Psychological bulletin*, 2003, *129* (1), 139.

Beattie, Jane, Jonathan Baron, John C Hershey, and Mark D Spranca, "Psychological determinants of decision attitude," *Journal of Behavioral Decision Making*, 1994, *7* (2), 129–144.

Besedeš, Tibor, Cary Deck, Sudipta Sarangi, and Mikhael Shor, "Reducing choice overload without reducing choices," *Review of Economics and Statistics*, 2015, *97* (4), 793–802.

Brown, Meta, Christopher J. Flinn, and Andrew Schotter, "Real-Time Search in the Laboratory and the Market," *American Economic Review*, April 2011, *101* (2), 948–74.

Burdett, Kenneth and Tara Vishwanath, "Declining Reservation Wages and Learning," *The Review of Economic Studies*, 1988, *55*, 655–666.

Buturak, Gökhan and Özgür Evren, "Choice overload and asymmetric regret," *Theoretical Economics*, 2017, *12* (3), 1029–1056.

Caplin, Andrew, Mark Dean, and Daniel Martin, "Search and satisficing," *American Economic Review*, 2011, *101* (7), 2899–2922.

Casner, Ben, "Learning while shopping: an experimental investigation into the effect of learning on consumer search," *Experimental Economics*, 2021, *24* (1), 238–273.

Charness, Gary and Peter Kuhn, "Lab labor: What can labor economists learn from the lab?," *Handbook of labor economics*, 2011, *4*, 229–330.

Cheng, Steve, "Differentiation Under the Integral Sign with Weak Derivatives," *CiteSeerX 10.1.1.525.2529*, 2010.

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman, "Choice overload: A conceptual review and meta-analysis," *Journal of Consumer Psychology*, 2015, *25* (2), 333–358.

Conlon, John J., Laura Pilossoph, Matthew Wiswall, and Basit Zafar, "Labor Market Search with Imperfect Information and Learning," *NBER Working Paper 24988*, 2018.

Cox, James C and Ronald L Oaxaca, "Good news and bad news: Search from unknown wage offer distributions," *Experimental Economics*, 2000, *2* (3), 197–225.

_ and _ , "Laboratory tests of job search models," *Handbook of Experimental Economics Results*, 2008, *1*, 311–318.

Dean, Mark, "Status quo bias in large and small choice sets," *Unpublished working paper*, 2008.

_ , Dilip Ravindran, and Jörg Stoye, "A Better Test of Choice Overload," 2022.

Enke, Benjamin and Thomas Graeber, "Cognitive uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

Gerasimou, Georgios, "Indecisiveness, undesirability and overload revealed through rational choice deferral," *The Economic Journal*, 2018, *128* (614), 2450–2479.

_ , "The Decision-Conflict Logit," *arXiv preprint arXiv:2008.04229*, 2020.

Greifeneder, Rainer, Benjamin Scheibehenne, and Nina Kleber, "Less may be more when choosing is difficult: Choice complexity and too much choice," *Acta psychologica*, 2010, *133* (1), 45–50.

Hodgson, Charles and Gregory Lewis, "You can lead a horse to water: Spatial learning and path dependence in consumer search," Technical Report, National Bureau of Economic Research 2023.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest, "Empirical search and consideration sets," in "Handbook of the Economics of Marketing," Vol. 1, Elsevier, 2019, pp. 193–257.

_ **and Pradeep Chintagunta**, "Simultaneous or sequential? Search strategies in the US auto insurance industry," *Marketing Science*, 2017, *36* (1), 21–42.

Horan, Sean, "Sequential search and choice from lists," *Unpublished paper.[716]*, 2010, *1.*

Iyengar, Sheena S and Mark R Lepper, "When choice is demotivating: Can one desire too much of a good thing?," *Journal of personality and social psychology*, 2000, *79* (6), 995.

Jones, D.S., *The Theory of Generalised Functions*, Cambridge: Cambridge University Press, 1982.

Kamenica, Emir, "Contextual inference in markets: On the informational content of product lines," *American Economic Review*, 2008, *98* (5), 2127–49.

Koulayev, Sergei, "Search With Dirichlet Priors: Estimation and Implications for Consumer Demand," *Journal of Business and Economic Statistics*, 2013, *31* (2), 226–239.

Kuksov, Dmitri and J. Miguel Villas-Boas, "When More Alternatives Lead to Less Choice," *Marketing Science*, 2010, *29* (3), 507–524.

Manzini, Paola, Marco Mariotti, and Levent Ulku, "Sequential Approval: A Model of "Likes", Paper Downloads and Other Forms of Click Behaviour.," 2019.

Natan, Olivia R., "Choice Frictions in Large Assortments," *Marketing Science*, 2025, *44* (3), 491–732.

Nocke, Volker and Patrick Rey, "Consumer Search and Choice Overload," 2021.

Reibstein, David J, Stuart A Youngblood, and Howard L Fromkin, "Number of choices and perceived decision freedom as a determinant of satisfaction and consumer behavior.," *Journal of Applied Psychology*, 1975, *60* (4), 434.

Rosenfield, Donald B. and Roy D. Shapiro, "Optimal Adaptive Price Search," *Journal of Economic Theory*, 1981, *25* (1), 1–20.

Rothschild, Michael, "Searching for the Lowest Price When the Distribution of Prices Is Unknown," *Journal of Political Economy*, 1974, *82* (4), 689–711.

Rubinstein, Ariel and Yuval Salant, "A model of choice from lists," *Theoretical Economics*, 2006, *1* (1), 3–17.

Santos, Babur De Los, Ali Hortaçsu, and Matthijs R. Wildenbeest, "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, May 2012, *102* (6), 2955–80.

_ , _ , and _ , "Search with Learning for Differentiated Products: Evidence from E-Commerce," *Journal of Business and Economic Statistics*, 2017, *35* (4), 626–641.

Scheibehenne, Benjamin, "The effect of having too much choice," 2008.

_ , Rainer Greifeneder, and Peter M Todd, "Can there ever be too many options? A meta-analytic review of choice overload," *Journal of consumer research*, 2010, *37* (3), 409–425.

Stigler, George J., "The Economics of Information," *Journal of Political Economy*, 1961, *69*, 213–213.

Tversky, Amos and Eldar Shafir, "Choice under conflict: The dynamics of deferred decision," *Psychological science*, 1992, *3* (6), 358–361.

Ursu, Raluca M, Qianyun Zhang, and Elisabeth Honka, "Search gaps and consumer fatigue," *Marketing Science*, 2023, *42* (1), 110–136.

Weitzman, Martin L, "Optimal Search for the Best Alternative," *Econometrica: Journal of the Econometric Society*, 1979, pp. 641–654.

A Search Length Dependent Threshold Functions and the Proofs of Theorems 1, 2 and 4

We bring together the proof of theorems 1, 2 and 4 because they can all be seen as manifestations of the same underlying principle. In each case, the optimal threshold strategies are what we term **search length dependent** - the threshold depends only on the number of alternatives that have been searched - not the value of those alternatives, nor the size of the choice set. It is possible to characterize what types of search length dependent threshold functions can lead to choice overload.

Definition 15. *A threshold function τ is **Search Length Dependent** if $\tau(l, N) = \tau(l', M)$ if $|A(l)| = |A(l')|$, $|A(l)| \neq N$ and $|A(l')| \neq M$. In such cases the the threshold is a function only of the number of alternatives that have been searched, and we use the notation $\hat{\tau}(n)$ for the threshold when n alternatives have been searched. We say that a Search Length Dependent threshold function is*

- ***Increasing*** *if $\hat{\tau}(n) \geq \hat{\tau}(m)$ for $n > m$ with the inequality strict for some $m > 1$ (assuming all alternatives have not been searched)*
- ***Decreasing*** *if $\hat{\tau}(n) \leq \hat{\tau}(m)$ for $n > m$ with the inequality strict for some $m > 1$*
- ***Constant*** *if $\hat{\tau}(n) = \hat{\tau}(m)$ for $n < m$ (assuming all alternatives have not been searched).*

We say a threshold distribution T is increasing (decreasing, constant) search length dependent if its support contains only increasing (decreasing, constant) search length dependent threshold functions.

To prove the relevant theorems, we make use of the following lemma, which states that search length dependent thresholds can generate choice overload if and only if they are decreasing. As the standard search problem gives rise to a constant search length dependent model, while search with increasing costs and search with static optimization give rise to decreasing search length dependent models, theorems 1, 2 and 4 follow immediately.

Lemma 4. *A search length dependent threshold T can generate choice overload with uniform search if it is decreasing. Otherwise it cannot generate choice overload.*

Proof. First, let T be decreasing. Let $m > 1$ be the smallest integer such that $\hat{\tau}^*(m+1) < \hat{\tau}^*(m)$ for some $\hat{\tau}^*$ in the support of T. This means, that for all $\hat{\tau}$ in the support of T the thresholds are constant for points 1 to m (i.e. $\hat{\tau}(1) = ... = \hat{\tau}(m)$) and either constant or strictly decreasing between m and $m+1$. Now construct a set B which contains d, y and $x_1, ..., x_m$. Select the utility function such that $u(y) > \hat{\tau}^*(m) > u(d) > \hat{\tau}^*(m+1) > u(x_1) > .... > u(x_m)$. Consider the set $A = \{d, y\}$. There is only one possible list ordering in A, which is dy. Thus, the probability of choosing the default is

$$p(d, A) = \sum_{\tau \in \mathcal{T}} T(\hat{\tau}) 1(\hat{\tau}(1) < u(d))$$

Specifically, the default will not be chosen under $\hat{\tau}^*$.

Now consider the set B. Note that any τ that led to the choice of the default in set A will also lead to the choice of the default in set B under any list ordering, as for such thresholds $\hat{\tau}(1) < u(d)$, and so only the default will be searched. Furthermore, with some probability the threshold $\hat{\tau}^*$ will lead to the choice of the default. Consider the list $l = dx_1....x_m y$. From this list, because $u(d) > \hat{\tau}^*(m+1)$, $s_{u,\tau}(l) = x_m$ and $c_{u,\tau}(l) = d$. Thus we have

$$\begin{aligned} p(d, B) &\geq \\ \sum_{\tau \in \mathcal{T}} T(\hat{\tau}) 1(\hat{\tau}(1) < u(d)) + T(\hat{\tau}^*) \frac{1}{|\mathcal{L}_B|} & \\ &> \\ \sum_{\tau \in \mathcal{T}} T(\hat{\tau}) 1(\hat{\tau}(1) < u(d)) & \\ &= p(d, A) \end{aligned}$$

Thus demonstrating choice overload with uniform search.

Next assume that T is either increasing or constant and fix a utility function u. We show that, if an increasing or constant threshold function $\hat{\tau}$ leads to a choice other than the default for some list order in a set A, then it must lead to a choice other than the default for any set $B \supset A$ and list $l' \in \mathcal{L}_B$. This is enough to establish the lemma, as it implies that the probability of choice of default in set A must be lower than in set B. To see this claim, note that, if d is not chosen from $l \in \mathcal{L}_A$ according to $\hat{\tau}$, it must be that, for some $x \in A$, $u(x) > u(d)$ and $x \trianglerighteq_l s_{u,\hat{\tau}}(l)$. This means (i) that $\hat{\tau}(1) > u(d)$ and (ii) that there exists an $x \in A(L)$ such that $u(x) > u(d)$. Now consider any $l' \in \mathcal{L}_B$. Note that, because $\hat{\tau}(1) > u(d)$ and $\hat{\tau}(k) \geq \hat{\tau}(1)$ for all $|B| > k \geq 1$, search will either stop because (a) some y is found at search point k such that $u(y) \geq \hat{t}(k) > u(d)$, or (b) because the entire set has been searched, in which case x has been seen. In either case there exists an element z such that $u(z) > u(d)$ and $z \trianglerighteq_l s_{u,\hat{\tau}}(l')$, meaning that $c_{u,\hat{\tau}}(l') \neq d$. $\square$

B Proofs

B.1 Proof of Lemma 1

Proof. First, define for each number i the myopic optimal strategy which is the optimal strategy if there were only one more alternative to search, and the search cost is $k(i)$. This is a threshold strategy $\bar{t}(i)$ where this value solves

$$k(i) = \int_{\bar{t}(i)}^{\infty} (y - \bar{t}(i)) f(y) dy \tag{2}$$

Note that, as $k(i)$ is increasing in i, it must be the case that $\bar{t}(i)$ is decreasing in i. Let $\hat{V}(x)$ be the expected value of searching one more period if the utility of the best seen alternative is x – *i.e.*

$$\hat{V}(x) = \int_{-\infty}^{x} x f(y) dy + \int_{x}^{\infty} y f(y) d(y)$$

Note that

$$\begin{aligned} \hat{V}(x) - k(i) &> x \text{ for } x < \bar{t}(i) \\ \hat{V}(x) - k(i) &< x \text{ for } x > \bar{t}(i) \end{aligned} \tag{3}$$

We claim that the threshold strategy $\bar{t}(i)$ is in fact the fully optimal strategy $\hat{\tau}_k(i)$. We show this for a search for a set of arbitrary size M using induction on j, the number of items left to search. Definitionally the claim is true for $j = 1$, so assume it is true of j and consider $j+1$. Consider a DM who has searched $M-j-1$ items and the best item seen so far has value $y > \bar{t}(M-j-1)$. The search should continue if

$$y < k(M-j-1) + V_{M-j-1}(y)$$

Where $V_{M-i-j-1}(y)$ is the value of continuing to search assuming optimal behavior in the future. But as $y > \bar{t}(M-j-1) > \bar{t}(M-j) = \tau_k(M-j)$ (where the last equality follows from the inductive assumption), we know that $V_{M-j-1}(y) = \hat{V}(y)$, as search will stop next period. By inequality (3) we therefore have that $y > k(M-j-1) + \hat{V}(y) = k(M-j-1) + V_{M-j-1}(y)$ so it is optimal to stop searching.

Now consider the case in which $y < \bar{t}(M-j-1)$. Now note that $V_{M-j-1}(y) \geq \hat{V}(y)$, because $V_{M-j-1}(y)$ is the maximal value of continuing to search across all feasible search strategies, while $\hat{V}(y)$ is the value for a particular feasible strategy. Then we have that

$$y < k(M-j-1) + \hat{V}(y) \leq k(M-j-1) + V_{M-j-1}(y)$$

And so it is optimal for search to continue.

We have therefore established that the optimal threshold strategy is to set $\tau(i) = \bar{t}(i)$ for all i. Because $\bar{t}(i)$ is decreasing in i this completes the proof. $\square$

B.2 Proof of Theorem 3

Proof (Part 1). Let x be the first item in l that has utility above d. Fix a threshold function τ. Then $c_{\tau,u}(l) = d$ if and only if search stops before an alternative with utility above d is found - i.e. $s_{\tau,u}(l) \rhd_l x$. This will occur if and only if, for some $y \rhd_l x$,

$$u(d) > \tau(l^y, |A(l)|) = \hat{\tau}_k(|A(l^y)|)$$

Let y^* be the element that immediately precedes x in l. Then, as $\hat{\tau}_k$ is strictly decreasing, the above condition will hold if and only if

$$u(d) > \hat{\tau}_k(|A(l^{y^*})|)$$

As u and is fixed and $\tau = \hat{\tau}_k$ is fixed, whether or not this condition holds depends only on $|A(l^{y*})| = F$. Thus, for any τ, $1\,(c_{\tau,u}(l) = d)$ depends only on F, and therefore so does

$$\hat{c}(d,l) = \sum_{\tau \in \mathcal{T}} T(\tau) 1(c_{\tau,u}(l) = d)$$

completing the proof. □

Proof (Part 2). For any u, l $\tau = \hat{\tau}_k$ and x

$$s_{\tau,u}(l) \trianglerighteq_l x$$

If and only if, for some $y \trianglerighteq_l x$

$$\max_{z \trianglerighteq_l y} u(z) > \tau(l^y, |A(l)|) = \hat{\tau}_k(|A(l^y)|).$$

As, for all $y \trianglerighteq_l x$, $\max_{z \trianglerighteq_l y} u(z) \leq \max_{z \trianglerighteq_l x} u(z)$ and $\hat{\tau}_k(|A(l^y)|) \geq \hat{\tau}_k(|A(l^x)|)$, this condition will hold if and only if

$$\max_{z \trianglerighteq_l x} u(z) > \hat{\tau}_k(|A(l^x)|)$$

The left hand side depends only on the maximal value of utility that occurs at of before x, while the right hand side depends only on $|A(l^x)|$ - i.e. the position of x in l, thus completing the proof. □

B.3 Proof of Theorem 5

Proof (Part 1). Let x be the first item in l that has utility above d. Fix a threshold function $\hat{\tau}_k^s$. Then $c_{\hat{\tau}_k^s,u}(l) = d$ if and only if search stops before an alternative with utility above d is found - i.e. $s_{\hat{\tau}_k^s,u}(l) \rhd_l x$. Given that search will definitely stop at n_k (assuming that this is less than the length of list l), then we have $c_{\tau,u}(l) = d$ if and only if $n_k \leq F$ - the position of the first above-default alternative. Thus the probability that the default is chosen depends only on the probability $P(k|n_k \leq F)$. □

Proof (Part 2). For any u, l $\tau = \hat{\tau}_k^s$ and x

$$s_{\tau,u}(l) \unrhd_l x$$

if and only if $|A(l^x)| \geq n_k$ i.e. the position of x in l. Thus the probability of stopping at or before x is just $P(k|n_k \leq |A(l^x)|)$. □

B.4 Proof of Theorem 6

Proof. Let $\hat{\tau}_k$ be the threshold function in the support of T that is consistent with the highest cost k. This implies that $\hat{\tau}_k(\mu, n) \leq \tau(\mu, n)$ for every τ in the support of T, beliefs μ and items remaining n.

Note that it must be the case that observing a (the worst possible alternative) increases the likelihood that the choice set is of low quality, so $\mu(da) < \mu(d) = 0.5$. As k guarantees learning, this means that $\hat{\tau}_k(\mu(da), 1) < \hat{\tau}_k(\mu(d), 1)$.

Consider three alternatives d, a and x and a utility function u such that $u(a) < u(d) < u(x)$ and $\hat{\tau}_k(\mu(da), 1) < u(d) < \hat{\tau}_k(\mu(d), 1)$.

We will show that this will cause choice overload when comparing the sets $A = \{d, x\}$ and $B = \{d, a, x\}$. Note first, that the default will never be chosen in set A. This is because $u(d) < \hat{\tau}_k(\mu(d), 1) \leq \tau(\mu(d), 1)$ for every τ in the support of T.

Thus x will always be searched in choice set A, meaning that it will be chosen, as $u(x) > u(d)$. We can therefore conclude that $p(d, A) = 0$.

In order to show that T can generate choice overload with uniform search, it is therefore enough to find one threshold in the support of T and one list order from B in which d is chosen. This will be the case for the order dax ad the threshold $\hat{\tau}_k$. This is because, by assumption, $\hat{\tau}_k(\mu(da), 1) < u(d)$, and so search stops after the sequence da is observed. As $u(d) > u(a)$, d will be chosen, implying

$$p(d, B) \geq T(\hat{\tau}_k)\frac{1}{2} > 0 = p(d, A)$$

□

B.5 Proof of Theorem 7

Proof (Part 1). Let y be the first item in l that has utility above d. Fix a threshold function $\hat{\tau}_k$. Then $c_{\hat{\tau}_k,u}(l) = d$ if and only if search stops before an alternative with utility above d is found - i.e. $s_{\hat{\tau}_k,u}(l) \rhd_l y$. Let x be the alternative immediately proceeding y, then by definition

$$s_{\hat{\tau}_k,u}(l) \rhd_l y \text{ if and only if } k \geq m_{l,u}(x)$$

Thus

$$\begin{aligned} \hat{c}(d, l) &= \sum_{\tau \in T} T(\tau) 1 \left(s_{\tau,u}(l) \rhd_l y \right) \\ &= \sum_{k \in K} P(k) 1(k \geq m_{l,u}(x)) \end{aligned}$$

□

Proof (Part 2). This follows directly from the definition of the minimum cost that induces stopping

$$\begin{aligned}
& \sum_{y \trianglerighteq_l x} \hat{s}(y,l) \\
= & \sum_{\tau \in T} T(\tau) 1\left(s_{\tau,u}(l) \trianglerighteq_l x\right) \\
= & \sum_{k \in K} P(k) 1(k \geq m_{l,u}(x))
\end{aligned}$$

□

B.6 Proof of Lemma 3

Proof. By Remark 1, we know that for a given k and $|N|$ the solution to the optimal search problem is a fixed reservation level. Furthermore the solution is given by the solution to the miopic search problem - i.e. the x^* that solves

$$k = \int_{x^*}^{\infty} (x - x^*) f_{|N|} dx$$

For a fixed x^*, the right hand side of this expression is decreasing in $|N|$, as $|M| > |N|$, implies that $f_{|N|}$ strictly first order stochastically dominates $f_{|M|}$. Thus, for a fixed cost, an increase in set size from $|N|$ to $|M|$ must be matched by a decrease in x^* to ensure that the right hand side of the expression equals the left hand side. □

B.7 Proof of Theorem 8

Proof. First we show that decision avoidance is consistent with choice overload. Pick some $\hat{\tau}^{d*}$ in the support of T. By assumption, $\tau^{d*}(1) > \tau^{d*}(2)$. Pick 3 alternatives, d, x and y, and a utility function u such that

$$\tau^{d*}(1) > u(y) > u(d) > u(x) > \tau^{d*}(2)$$

Now consider the set $A = \{d, y\}$. Here, there is one possible list - dy. The probability of choosing the default in this list is equal to the probability of threshold functions such that $\tau^d(1) < u(d)$

$$p(d, A) = \sum T(\tau)1(\tau^d(1) < u(d))$$

Note that τ^{d*} is not part of this set.

Now consider a set $B = \{d, x, y\}$ and note that, for any list order l, τ^{d*} will lead to the choice of the default: $u(d) > \tau^{d*}(2)$ means that $s_{\tau^{d*},u}(l) = d$ and so $c_{\tau^{d*},u}(d)$. Note also that any threshold that lead to the choice of d in A will also lead to the same choice in B for any search order, as the fact that d was chosen in A implies that $\tau^d(1) < u(d)$, and the fact that τ is strictly decreasing implies that $\tau^d(2) < \tau^d(1)$. Thus we have

$$\begin{aligned} p(d, B) &\geq \sum T(\tau)1(\tau^d(1) < u(d)) + T(\tau^{d*}) \\ &> \sum T(\tau)1(\tau^d(1) < u(d)) \\ &= p(d, A) \end{aligned}$$

Next we show that contextual inference is consistent with choice overload. to see this, let k^* be the highest cost such that there exists a $\tau^I_{k^*}(x)$ in the support of T. By assumption, we know that $\tau^I_{k^*}(2) > \tau^I_{k^*}(3)$. Pick three alternatives, d, x and y such that

$$\tau^I_{k^*}(2) > u(y) > u(d) > u(x) > \tau^I_{k^*}(3)$$

Note that, by construction $\tau^I_k(2) \geq \tau^I_{k^*}(2) > u(d)$ for every $\tau^I_k(1) \in T$, and so in the set $A = \{d, y\}$ we have that, for all such threshold functions the set will be completely searched, and because $u(y) > u(d)$, $u(y)$ will be chosen. Thus $p(d, A) = 0$, and so to demonstrate choice overload with uniform search, we only need to find a single search order from the set $B = \{d, x, y\}$ and single threshold in T such that the default is chosen. Note that $u(d) > \tau^I_{k^*}(3)$, and so $s_{\tau^I_{k^*},u}(l) = d$ for any l such that $|A(l)| = 3$. As a result, $c_{\tau^I_{k^*},u}(l) = d$ for any search order derived from B, completing the proof. $\square$

B.8 Proof of Theorem 9

Proof. We first deal with the case in which τ is consistent with either decision avoidance or the search problem with contextual inference. In order to cause choice overload, it must be that, with some probability, d was not chosen from the set A. This means that there must be some $x \in A$ such that $u(x) > u(d)$. A necessary condition for choice overload is that for some $l \in \mathcal{L}_B$ $c_{\tau,u}(l) = d$. We therefore complete this part of the proof by first showing

$$c_{\tau,u}(l) = d \Rightarrow s_{\tau,u}(l) = d$$

We do so by showing that $s_{\tau,u}(l) \neq d$ then $c_{\tau,u}(l) \neq d$. As τ is consistent with either decision avoidance or contextual inference, we know that $\tau^d(N)$ is weakly decreasing in N, the number of items remaining to search,[21]. If $s_{\tau,u}(l) \neq d$ it must be the case that that $\tau(N-1) > u(d)$, and so $\tau(i) > u(d)$ for all $i < N-1$, as $\tau(N-1) \leq \tau(i)$. Thus search will only stop if, for some x searched at position j, $u(x) > \tau(N-j) > u(d)$, or all items have been searched. In either case, (given that there is an alternative x in l such that $u(x) > u(d)$), we have $u(x) > u(d)$ for some $x \trianglerighteq_l s_{\tau,u}(l)$ and so $c_{\tau,u}(l) \neq d$.

This confirms that $s_\tau(l) = d$ for some $l \in \mathcal{L}_B$. However, as decision avoidance and contextual inference stopping rules depend only on the number of alternatives, the fact that $s_\tau(l) = d$ for some $l \in \mathcal{L}_B$ implies $s_\tau(l) = d$ for all $l \in \mathcal{L}_B$.

We next show that, for any $x \in A \subset B$ where x is an above-default alternative

$$p(d,B) - p(d,A) = \sum_{l \in \mathcal{L}_B} \pi_B(l)\hat{s}(d,l) - \sum_{l \in \mathcal{L}_A} \pi_A(l)\hat{s}(d,l)$$

As, for any set S,

$$p(d,S) = \sum_{l \in \mathcal{L}_S} \pi_S(l)\hat{c}(d,l)$$

[21] It is strictly decreasing if τ is consistent with decision avoidance, constant if it is consistent with contextual inference.

It is enough to show that, for any set in which there is an above alternative

$$\hat{c}(d,l)=\hat{s}(d,l)$$

Given that we have already shown that

$$c_{\tau,u}(l)=d\Rightarrow s_{\tau,u}(l)=d$$

and as it is clearly true that $s_{\tau,u}(l)=d\Rightarrow c_{\tau,u}(l)=d$, we can conclude that

$$\hat{c}(d,l)=\sum_{\tau\in\mathcal{T}}T(\tau)1(s_{\tau,u}(l)=d)=\hat{s}(d,l)$$

We next consider the case in which the threshold function τ is consistent with either search with increasing costs, static optimization or learning. We first need to show that τ can generate choice overload only if, for some $l\in\mathcal{L}_B$, $d\rhd_l s_\tau(l)\rhd_l x_l$ where x_l is the first above-default alternative in l. To see this, first note that, if $s_\tau(l)=d$ for all $l\in\mathcal{L}_B$, it means that $u(d)>\tau(d,|B|)$. All the threshold functions currently being considered are either invariant to $|B|$ (increasing costs or static optimization) or increasing in $|B|$ (learning). So, as $|A|<|B|$, this would imply that $u(d)>\tau(d,|A|)$ and so $s_\tau(l)=d$ for all $l\in\mathcal{L}_A$. This would in turn imply $c_\tau(l)=d$ for all $l\in\mathcal{L}_A$, and so for any induced data set, $p(d,A)=1$, making choice overload impossible. This implies that, for choice overload to occur, a necessary condition is that, for some $l\in\mathcal{L}_B$, $d\rhd_l s_\tau(l)$. Next, note that, $x_l\trianglerighteq_l s_\tau(l)$, implies that $c_\tau(l)\neq d$, as by construction $u(x_l)>u(d)$. Thus, it must be the case that, for some l such that $d\rhd_l s_\tau(l)$, it must also be that $s_\tau(l)\rhd_l x_l$, otherwise every list order that leads to the choice of default in B would also lead to the choice of default in A and so we would have $p(d,B)\leq p(d,A)$.

Finally we show that, for any T that is consistent with either of these models and any set A such that $|A|=2$ and which contains an alternative better than the default

$$p(d,B)-p(d,A)\leq\sum_{l\in\mathcal{L}_B}\pi_B(l)\sum_{y\in B|d\rhd_l y\rhd_l x_l}\hat{s}(y,l)$$

First, note that, for any list l with associated first above-default alternative x_l the default is chosen if and only if search stops before x_l. Thus it is generally true for any set S that contains an above-default alternative that

$$\begin{aligned} p(d,S) &= \sum_{l\in\mathcal{L}_B}\pi_S(l)\sum_{y\in B|d\trianglerighteq_l y\triangleright_l x_l}\hat{s}(y,l) \\ &= \sum_{l\in\mathcal{L}_B}\pi_S(l)\hat{s}(d,l)+\sum_{l\in\mathcal{L}_B}\pi_B(l)\sum_{y\in B|d\triangleright_l y\triangleright_l x_l}\hat{s}(y,l) \end{aligned}$$

Furthermore,

$$\begin{aligned} \hat{s}(d,l) &= \sum_{\tau\in\mathcal{T}}T(\tau)1(s_{\tau,u}(l)=d) \\ &= .\sum_{\tau\in\mathcal{T}}T(\tau)1\left(u(d)>\tau(d,|A(l)|)\right) \end{aligned}$$

- i.e. search stops at d if and only if a threshold function is drawn such that the utility of d is above the threshold.

Finally, note that, as $|A|=2$, it is always the case that the only alternative before the first above-default alternative is d itself, so

$$\begin{aligned} p(d,A) &= \sum_{l\in\mathcal{L}_B}\pi_S(l)\hat{s}(d,l) \\ &= \sum_{\tau\in\mathcal{T}}T(\tau)1\left(u(d)>\tau(d,|A(l)|)\right) \end{aligned}$$

This implies that

$$\begin{aligned} & p(d,B)-p(d,A) \\ =\ & \sum_{\tau\in\mathcal{T}}T(\tau)1\left(u(d)>\tau(d,|B(l)|)\right)+\sum_{l\in\mathcal{L}_B}\pi_B(l)\sum_{y\in B|d\triangleright_l y\triangleright_l x_l}\hat{s}(y,l) \\ & -\sum_{\tau\in\mathcal{T}}T(\tau)1\left(u(d)>\tau(d,|A(l)|)\right) \end{aligned}$$

As argued above, for threshold functions in the classes considered, $u(d)>\tau(d,|B(l)|)$

implies that $u(d) > \tau(d, |A(l)|)$, and so

$$\sum_{\tau \in \mathcal{T}} T(\tau) 1\left(u(d) > \tau(d, |B(l)|)\right) - \sum_{\tau \in \mathcal{T}} T(\tau) 1\left(u(d) > \tau(d, |A(l)|)\right) \leq 0$$

Providing the required results. $\square$

B.9 Proof of Theorem 10

Proof. First note that, as shown in the proof of theorem 9, for models in this class, we have that the default can only be chosen if it is the only alternative searched - i.e.

$$c_{\tau,u}(l) = d \Rightarrow s_{\tau,u}(l) = d$$

This implies that

$$\begin{aligned} \hat{c}(d, l) &= \sum_{\tau \in \mathcal{T}} T(\tau) 1\left(u(d) > \tau(d, |A(l)|)\right) \\ &= \sum_{\tau \in \mathcal{T}} T(\tau) 1\left(u(d) > \tau(d, |A(l')|)\right) \\ &= \hat{c}(d, l') \end{aligned}$$

$\square$

B.10 Proof of Theorem 11

Proof. Consider any τ consistent with either decision avoidance or the standard search problem with contextual inference. Either $u(d) \geq \tau(d, |A|)$ in which case $s_{\tau,u}(l) = d$, or $u(d) < \tau(d, |A|)$, in which case $u(x) < \tau(l^y, |A|)$ for any $x, y \in A$ where y is not the last element in l. This follows from the fact that (i) $u(x) \leq u(d)$ $\forall$ $x \in A$ (by assumption) and (ii) $\tau(l^y, |A|) \geq \tau(d, |A|)$ as τ is either decreasing or constant. Thus, for any τ in the support of T, $s_{\tau,u}(l)$ is either d or the terminal element in l, and so the probability of stopping at any other element is zero. $\square$

C Sequential Search with Learning

A decision maker faces a choice set with N options of scalar monetary rewards. These options have been drawn from one of two probability distributions $F(x|\overline{\mu})$ or $F(x|\underline{\mu})$ with same support $[a,b] \subset \mathbb{R}$. $F(x|\overline{\mu})$ is a "good" distribution, while $F(x|\underline{\mu})$ is a "bad" distribution. They are both absolutely continuous, with respective probability density functions $f(x|\overline{\mu})$ and $f(x|\underline{\mu})$, which satisfy a strict Monotone Likelihood Ratio property. This also implies that $F(x|\overline{\mu})$ strictly first order stochastically dominates $F(x|\underline{\mu})$, i.e., $F(x|\overline{\mu}) < F(x|\underline{\mu})$ for all $x \in (a,b)$.[22]

The decision maker has a belief $\mu \in (0,1)$ over the probability that the choice set they're facing was drawn from $F(x|\overline{\mu})$, and initially holds a prior $\mu_0 \in (0,1)$. Every belief μ defines a new probability distribution $F(x|\mu) = \mu F(x|\overline{\mu}) + (1-\mu)F(x|\underline{\mu})$, with $f(x|\mu)$ analogously defined. After looking at option $x \in [a,b]$ when holding belief $\mu \in (0,1)$, the decision maker updates belief to $\mu(x)$ according to Bayes' Rule:

$$\mu(x) = \frac{1}{1 + \frac{(1-\mu)}{\mu}\frac{f(x|\underline{\mu})}{f(x|\overline{\mu})}}.$$

This defines a function $(x,\mu) \mapsto \mu(x)$ from options x and current beliefs μ to a new updated belief $\mu(x)$. Notice that this resulting function $\mu(x)$ is increasing in μ and in x due to the strict Monotone Likelihood Ratio assumption.[23]

The DM sequentially searches through the available options. All search options remain available to choose - i.e. search with recall. Every period they have a value $y \in [a,b]$ defining the highest monetary reward encountered thus far in the search process and a current belief $\mu \in (0,1)$. They can either *stop* and pick the best option found so far, terminating the problem, or *search* and look at a new option. Upon drawing a value x, they update their belief to $\mu(x)$ as described above and update their best available option if $x > y$ or not otherwise, and move on to the next period.

[22]The assumption of strict monotonicity and stochastic dominance are used to obtain the strict monotonicity results in what follows, and which are used in the main text. For weak monotonicity, the weaker and more usual forms of these assumptions suffice.

[23]The following results depend only on these monotonicity properties of the updating process, and not specifically on Bayesian updating. Any updating rule satisfying them will imply the same results.

The period variable n defines how far along the search process the agent is. Specifically it marks how many options in the choice set are *left to be searched.* This defines the following important periods. The agent starts at period $n = N$, with all options still unsearched, and only the default option being seen. $n = 1$ is the *final period*, in which there is only a single option left to search. $n = 0$ is the *terminal period*, in which all options have been searched and the problem terminates. Each time the agent searches, the period counter decreases by one.

This process defines a collection of value functions $V_N, V_{N-1}, ..., V_1, V_0 : [a, b] \times (0, 1) \to \mathbb{R}$. The value function $V_n(y, \mu)$ gives the value of reaching period n while holding belief μ and having encountered a best option with value y thus far (i.e. after the value drawn in period n has been revealed). These value functions are recursively connected to each other as follows. By assumption, in the terminal period the process must terminate, and therefore:

$$V_0(y, \mu) = y \text{ for all } y \text{ and } \mu.$$

And for $n = 1, ..., N$, the decision maker can either stop, and take option value y, or search, in which case they might draw a value that is above or below their current best one, while updating their beliefs accordingly. Therefore, we have the Bellman equation:

$$V_n(y, \mu) = \max\left\{y, -k + \int_a^y V_{n-1}(y, \mu(x))dF(x|\mu) + \int_y^b V_{n-1}(x, \mu(x))dF(x|\mu)\right\}.$$

In particular, given V_0 as assumed, in the last period $n = 1$:

$$V_1(y, \mu) = \max\left\{y, -k + yF(y|\mu) + \int_y^b x dF(x|\mu)\right\}.$$

That is, in this last period, beliefs μ only affect the value function through the distribution of values in period 0, hence the problem takes a typical form in which the learning aspect can be ignored. Given a belief μ, the solution takes the form of a *threshold* $\tau_1(\mu) \in [a, b)$ such that there's search if $y < \tau_1(\mu)$ and stopping otherwise.

It's obtained by setting $y = \tau_1(\mu)$ such that the two terms inside the max operator in V_1 are equal. Applying integration by parts on $\int_y^b x dF(x|\mu)$ then yields the standard equation defining $\tau_1(\mu)$:

$$k = \int_{\tau_1(\mu)}^{b} 1 - F(x|\mu)dx,$$

with $\tau(\mu) = a$ if $k \geq \int_a^b 1 - F(x|\mu)dx$.

It turns out that in earlier periods, despite the additional element of learning, the main properties of the problem and its solution stay the same. Denoting $\pi_n^*(y, \mu) \in \{\text{stop}, \text{search}\}$ the optimal policy function, the next proposition summarizes these properties.

Proposition 1. *For any period n the following hold*

1. *$V_n(y, \mu)$ is a continuous function.*

2. *Given a belief μ, $V_n(y, \mu)$ is an increasing convex function of y.*

3. *Given a belief μ, $V_n(y, \mu)$ is absolutely continuous in y.*[24]

4. *$\frac{\partial}{\partial y} V_n(y, \mu) \leq 1$ for all y and belief μ such that it exists, with*

    i. *$\frac{\partial}{\partial y} V_n(y, \mu) = 1$ if $\pi_n^*(y, \mu) = stop$.*

    ii. *$\frac{\partial}{\partial y} V_n(y, \mu) < 1$ if $\pi_n^*(y, \mu) = search$.*

5. *For each belief μ, there's a threshold $\tau_n(\mu) \in [a, b)$ such that*

    i. *$\pi_n^*(y, \mu) = stop$ if $y \geq \tau_n(\mu)$, $\pi_n^*(y, \mu) = search$ otherwise.*

    ii. *the resulting function $\tau_n(\mu)$ is non-decreasing in μ.*

    iii. *for every belief μ there's $k_n(\mu)$, increasing in μ, such that if the cost of searching is $k < k_n(\mu)$, then $\tau_n(\mu) > a$. If $\mu < \mu'$ are such that $\tau_n(\mu) > a$, then $\tau_n(\mu) < \tau_n(\mu')$.*

6. *Given y, $V_n(y, \mu)$ is non-decreasing in beliefs μ. And, given $\mu < \mu'$, if $y < \tau_n(\mu)$, then $V_n(y, \mu) < V_n(y, \mu')$.*

[24]Note that this in turn implies that it is almost everywhere differentiable in y.

Proof. The proof is by induction over the period n, starting from $n = 1$.

Before starting, define for each $(x, y) \in [a, b]^2$ the continuous function φ given by $\varphi(x, y) = \mathbb{I}\{x < y\}y + \mathbb{I}\{x \geq y\}x$, with $\mathbb{I}$ being the indicator function. Define then, for any $y \in [a, b]$ and $\mu \in (0, 1)$ and non-terminal period n, the *value of searching*:

$$S_n(y, \mu) = \int_a^b V_{n-1}(\varphi(x, y), \mu(x))dF(x|\mu),$$

with this being well-defined by the assumptions that will hold on V_{n-1} during the proof, namely continuity.

First, we establish (1)-(6) for V_1.

That (1)-(5) hold for V_1 follows from the discussion preceding the proposition. It might only be necessary to notice that if $\pi_1^*(y, \mu) = \text{search}$, $\frac{\partial}{\partial y}V_1(y, \mu) = \frac{\partial}{\partial y}S_1(y, \mu) = F(y|\mu) < 1$ if $y < b$, and is increasing in y. This derivative can be obtained by applying the Leibniz integral rule. Given this, absolute continuity also follows. In fact, $V_1(y, \mu)$ is Lipschitz continuous with constant 1. When checking for Lipschitz continuity, the only case that might not be immediate is if y_1 and y_2 are such that $y_1 \leq \tau_1(\mu) < y_2$. But then, by the definition of $\tau_1(\mu)$, it follows that

$$y_1 \leq V_1(y_1, \mu) = S_1(y_1, \mu) - k \leq \tau_1(\mu) = S_1(\tau_1(\mu), \mu) - k \leq S(y_2, \mu) - k < y_2 = V_1(y_2, \mu).$$

To see (6), notice that, for any y and any μ, if $\pi_1^*(y, \mu) = \text{search}$,

$$V_1(y, \mu) = S_1(y, \mu) - k = -k + \int_a^b \varphi(x, y)dF(x|\mu).$$

Therefore, V_1 is the maximum of y, constant in μ, and $S_1 - k$. We establish now that S_1 is non-decreasing in μ, hence so is V_1, and if $y < b$, S_1 is increasing in μ. Fix y and take μ and μ' such that $\mu' > \mu$. Notice that $F(\cdot|\mu')$ first order stochastically dominates $F(\cdot|\mu)$ and that, for a fixed y, φ is a non-decreasing function of x to obtain that

$$S_1(y, \mu) = \int_a^b \varphi(x, y)dF(x|\mu) \leq \int_a^b \varphi(x, y)dF(x|\mu') = S_1(y, \mu').$$

If $y < b$, then this inequality is strict, since for $x \geq y$, $\varphi(y, x) = x$ is increasing and strict stochastic dominance holds. Furthermore, then, given $\mu < \mu'$, if $y < \tau_1(\mu)$, then $y < \tau_1(\mu) < \tau_1(\mu') < b$, hence $V_1(y, \mu) = S_1(y, \mu) < S_1(y, \mu') = V_1(y, \mu)$.

We now proceed to the next step of the induction. Assume (1)-(6) hold for V_{n-1} and consider V_n. We start with some preliminary observations. Notice that

$$V_n(y, \mu) = \max\{y, S_n(y, \mu) - k\}.$$

From the inductive assumption, V_{n-1} is continuous, and since $\varphi(x, y)$ and $\mu(x)$ are continuous functions, $S_n(y, \mu)$ is continuous. Therefore, so is $V_n(y, \mu)$, it being the maximum of two continuous functions. Thus (1) is established. The above allows us to establish the following which will be of use later: $\pi^*_n(y, \mu) =$ search if $S_n(y, \mu) > y + k$ and $\pi^*_n(y, \mu) =$ stop otherwise. In particular, $\pi^*_n(b, \mu) =$ stop for all beliefs μ. Indeed, $\varphi(x, b) = b$ for all $x \in [a, b]$, and since $\tau_{n-1}(\mu) < b$ by the inductive assumption, $S_n(b, \mu) = b < b + k$.

Now take a fixed belief μ and consider

$$\frac{\partial}{\partial y} S_n(y, \mu) = \frac{\partial}{\partial y} \int_a^b V_{n-1}(\varphi(x, y), \mu(x)) dF(x|\mu).$$

From the inductive assumptions on V_{n-1}, the Leibniz integral rule applies to S_n, hence it's absolutely continuous and we can switch the signs of differentiation and integration to obtain the derivative, which exists almost everywhere:[25]

$$\frac{\partial}{\partial y} S_n(y, \mu) = \int_a^b \frac{\partial}{\partial y} V_{n-1}(\varphi(x, y), \mu(x)) dF(x|\mu) = \int_a^y \frac{\partial}{\partial y} V_{n-1}(y, \mu(x)) dF(x|\mu),$$

with the last equality following by separating the integral over $[a, y]$ and $(y, b]$, the latter upon which it is zero, since $\varphi(x, y) = x$.

[25] Specifically, $V_{n-1}(\varphi(x, y), \mu(x))$ is continuous and absolutely continuous in y, $\varphi(x, y)$ being absolutely continuous and bounded, with a bounded derivative. Under these conditions, it's known that the Leibniz integral rule still holds, hence S_n will be absolutely continuous with its derivative given by the switching of the signs of differentiation and integration. See, for example, Theorem 3 in `https://planetmath.org/differentiationundertheintegralsign`. In fact, it can be shown that the Leibniz integral rule still holds under even weaker conditions by using generalized functions, as in Theorem 7.40 of Jones (1982, p.263). Proofs relating these more general results to more specific cases such as ours can be found in Cheng (2010).

We now establish that $S_n(y,\mu)$ is non-decreasing in beliefs μ, and if $y < b$ it is increasing. This follows similarly to S_1. Fix y and consider beliefs μ, μ' with $\mu' > \mu$. From (6) on V_{n-1}, $V_{n-1}(\varphi(x,y),\mu(x)) \leq V_{n-1}(\varphi(x,y),\mu'(x))$ for all $x \in [a,b]$, and thus

$$S_n(y,\mu) = \int_a^b V_{n-1}(\varphi(x,y),\mu(x))dF(x|\mu) \leq \int_a^b V_{n-1}(\varphi(x,y),\mu'(x))dF(x|\mu).$$

Now first order stochastic dominance and $V_{n-1}(\varphi(x,y),\mu'(x))$ being non-decreasing in x, from the inductive assumption and $\varphi(x,y)$ and $\mu'(x)$ being non-decreasing in x, lead to

$$S_n(y,\mu) \leq \int_a^b V_{n-1}(\varphi(x,y),\mu'(x))dF(x|\mu) \leq \int_a^b V_{n-1}(\varphi(x,y),\mu'(x))dF(x|\mu') = S_n(y,\mu').$$

And if $y < b$, then this last inequality is strict. This is because of strict stochastic dominance and the fact that for $x > y$, $V_{n-1}(\varphi(y,x),\mu(x)) = V_{n-1}(x,\mu(x))$ is increasing in x.

(2)-(6) will now follow from these previous observations and the inductive assumptions on V_{n-1}. We first register an intermediate result:

Claim. If $\pi_n^*(y,\mu) = \text{stop}$, then $\pi_n^*(y',\mu) = \text{stop}$ at any $y' > y$ and if $\pi_n^*(y,\mu) = \text{search}$ then $\pi_n^*(y',\mu) = \text{search}$ at any $y' < y$.

Proof of Claim. By inductive assumption, $\frac{\partial}{\partial y}V_{n-1}(y,\mu) \leq 1$ for all y,μ, thus

$$\frac{\partial}{\partial y}S_n(y,\mu) = \int_a^y \frac{\partial}{\partial y}V_{n-1}(y,\mu(x))dF(x|\mu) \leq F(y|\mu) < 1, \text{ if } y < b.$$

That is, $\frac{\partial}{\partial y}S_n(y,\mu) < 1$ for all $y \in [a,b)$. Therefore, since $\frac{\partial}{\partial y}S_n(y,\mu) < 1$ while $\frac{\partial}{\partial y}y = 1$, the claim follows.

□

What the claim establishes is that, holding μ constant, there is, at most, one switch from searching to stopping as one increases the value of the best option y. What remains to obtain (5.i) is to show that, if $\pi_n^*(y,\mu) = \text{search}$ for some y, then a switch to $\pi_n^*(y', \mu) =$ stop happens at some $y' < b$. That is, while it's possible to always stop, in particular if the cost k is too high, it's not possible to always search. There is always a non-degenerate interval $[\tau_n(\mu), b]$ at which there's stopping.

That is, from the claim and the fact that for any period n, given μ, $\pi_n^*(b, \mu) =$ stop, it already follows that there's $\tau_n(\mu) \in [a, b]$ such that $\pi_n^*(y, \mu) =$ stop if $y \geq \tau_n(\mu)$ and $\pi_n^*(y, \mu) =$ search otherwise. It remains to see that, in fact, $\tau_n(\mu) < b$. But notice that if $\tau_n(\mu) > a$ then $S_n(y, \mu) > y + k$ for some y, while $S_n(b, \mu) < b + k$, so $\tau_n(\mu) < b$ follows from continuity. Thus (5.i) is established.

To establish (5.ii), notice that for any period n, given any y, if $\pi_n^*(y, \mu) =$ stop then $\pi_n^*(y, \mu') =$ stop for all $\mu' < \mu$. Indeed, if $S_n(y, \mu) \leq y + k$, since $S_n(y, \mu)$ is non-decreasing in beliefs, then given $\mu' < \mu$, $S_n(y, \mu') \leq S_n(y, \mu) \leq y + k$.

For (5.iii), start by noticing that

$$S_n(a, \mu) = \int_a^b V_{n-1}(x, \mu(x)) dF(x|\mu) \geq \int_a^b x dF(x|\mu) > a.$$

Hence $k_n(\mu) = S_n(a, \mu) - a$, increasing in μ since $S_n(a, \mu)$ is, is such that if $k < k_n(\mu)$, then $S_n(a, \mu) > a + k$ and thus $\tau_n(\mu) > a$.

Now, if $\mu < \mu'$ are such that $\tau_n(\mu) > a$, then $\tau_n(\mu) + k = S_n(\tau_n(\mu), \mu) < S_n(\tau_n(\mu), \mu')$, because $\tau_n(\mu) < b$. Therefore, $\pi^*(\tau_n(\mu), \mu') =$ search, which implies that $\tau_n(\mu) < \tau_n(\mu')$.

With the threshold functions $\tau_n(\mu)$ in hand, we can say that

$$V_n(y, \mu) = \begin{cases} S_n(y, \mu) - k, \text{ if } y < \tau_n(\mu) \\ y, \text{ otherwise} \end{cases}.$$

The remaining properties can now be more directly established.

For (6), $V_n(y, \mu)$ is constant in μ when $\tau_n(\mu) \leq y$, and non-decreasing in μ when $\tau_n(\mu) > y$. Therefore, V_n is the maximum of two non-decreasing functions of μ, and so is non-decreasing in μ. Now, the second part of (6) follows as in $n = 1$. If $\mu < \mu'$ are such that $y < \tau_n(\mu)$, then $y < \tau_n(\mu) < \tau_n(\mu') < b$, hence $V_n(y, \mu) = S_n(y, \mu) < S_n(y, \mu') = V_n(y, \mu)$.

For (3) and (4), $V_n(y, \mu)$ is absolutely continuous in y, being the maximum of two absolutely continuous functions of y. In $(\tau_n(\mu), b]$, $\frac{\partial}{\partial y} V_n(y, \mu) = 1$, and in $[a, \tau_n(\mu))$, $\frac{\partial}{\partial y} V_n(y, \mu) = \frac{\partial}{\partial y} S_n(y, \mu) < 1$ for $a \leq y < \tau_n(\mu) < b$ as already established.

Finally, for (2), $V_n(y, \mu)$ is linear and increasing in y in $y \geq \tau_n(\mu)$, and in $y < \tau_n(\mu)$, $V_n(y, \mu) = S_n(y, \mu)$ is an increasing convex function. Indeed, notice first that $\frac{\partial}{\partial y} V_n(y, \mu) = \frac{\partial}{\partial y} S_n(y, \mu) > 0$ from the inductive hypothesis that $\frac{\partial}{\partial y} V_{n-1}(y, \mu) > 0$. Second, for the convexity of S_n, notice that, for all x, $\varphi(x, y)$ is a convex function of y. Therefore, since by the inductive assumption $V_{n-1}(y, \mu)$ is increasing and convex in y for all μ, we have that $V_{n-1}(\varphi(x, y), \mu(x))$ is a convex function of y for all x. Hence, $S_n(y, \mu)$ is convex in y. $V_n(y, \mu)$ is, then, the maximum of two convex increasing functions of y, establishing (2).

□

Beyond these properties for any given period, the value functions and the threshold functions also have a relationship between them across periods. As might be expected, a decrease in the number of options left to search decreases value, and therefore decreases the thresholds. This captures an exploration-exploitation trade-off that exists because this model has a finite horizon. As more and more options are searched, the opportunity for learning decreases, and so it becomes more profitable to stop searching and choose.

Proposition 2. *1. For any y, μ, $V_{n-1}(y, \mu) \leq V_n(y, \mu)$ for any period n.*

2. For any μ, $\tau_{n-1}(\mu) \leq \tau_n(\mu)$ for any period n.

Proof. The proof is again by induction over n.

Since the search process terminates in $n = 0$, $V_1(y, \mu) \geq y = V_0(y, \mu)$ for any y and μ, and we can take $\tau_0(\mu) = a$ for all μ, hence $\tau_1(\mu) \geq \tau_0(\mu)$.

Now assume (1) and (2) hold for n and consider $V_{n+1}(y, \mu)$ and $\tau_{n+1}(\mu)$ for some y, μ. We will show that $V_{n+1}(y, \mu) \geq V_n(y, \mu)$, establishing (1), and in the process will also show that it cannot be that $\pi^*_n(y, \mu) = \text{search}$ but $\pi^*_{n+1}(y, \mu) = \text{stop}$, hence $\tau_{n+1}(\mu) \geq \tau_n(\mu)$, which establishes (2).

If $\pi_n^*(y,\mu) =$ stop and $\pi_{n+1}^*(y,\mu) =$ stop, $V_{n+1}(y,\mu) = y = V_n(y,\mu)$. If $\pi_n^*(y,\mu) =$ stop and $\pi_{n+1}^*(y,\mu) =$ search, $V_n(y,\mu) = y \leq V_{n+1}(y,\mu)$ by definition. If $\pi_n^*(y,\mu) =$ search and $\pi_{n+1}^*(y,\mu) =$ search, then

$$V_{n+1}(y,\mu) = S_{n+1}(y,\mu) - k \geq S_n(y,\mu) - k = V_n(y,\mu)$$

since $S_{n+1}(y,\mu) \geq S_n(y,\mu)$ by the inductive assumption that $V_n(y,\mu) \geq V_{n-1}(y,\mu)$ for all y, μ.

And it cannot be that $\pi_n^*(y,\mu) =$ search but $\pi_{n+1}^*(y,\mu) =$ stop. Indeed, this would imply that

$$S_n(y,\mu) - k = V_n(y,\mu) > y = V_{n+1}(y,\mu) \geq S_{n+1}(y,\mu) - k$$

which, as mentioned above, would contradict the inductive assumption (1) on V_n and V_{n-1}.

□

In the main text, lemma 2 connects the results from propositions 1 and 2 to our general sequential search framework. Since in our general framework we allow for varying costs of search, we further index the threshold functions obtained in the propositions by the cost k. It is immediate from the definition of these thresholds that they are lower the higher the costs.

Notice that for thresholds to be strictly interior for all periods and beliefs, it's sufficient that $\lim_{\mu \to 0} \tau_1(\mu) > a$ because of statements 5.ii from proposition 1 and 2 from proposition 2. To see that such costs guaranteeing search exist, define $k^* = \int_a^b x dF(x|\underline{\mu}) - a > 0$ and notice that $k_n(\mu) = \int_a^b V_{n-1}(x,\mu(x))dF(x|\mu) - a \geq k^*$ for all μ and n. Therefore, any $k < k^*$ is such that it satisfies the desired properties. We then define that a cost k *guarantees search* if the threshold for that cost is strictly interior for period 1 and all beliefs, in which case they are strictly increasing in beliefs from statement 5.iii in proposition 1.

D Additional Lemmata

D.1 Testing for Choice Overload

Lemma 5. *Let B and A be two sets such that $A \subset B$. Both B and A have been constructed by drawing items uniformly (without replacement) from a grand set X. Let $|A| = 2$, and $d \in A$. Let T be a subset of X. Then, if the stochastic choice function p does not exhibit choice overload, then*

$$E(p(d, A)|A \cap T \neq 0) \geq E(p(d, B)|B \cap T \neq 0)$$

Proof. Enumerate $T = \{t_1 ... t_n\}$. Note that

$$\begin{aligned} E(p(d, A)|A \cap T &\neq 0) \\ &= \sum_{j \in T} \frac{1}{n} p(d, \{d, t_j\}) \end{aligned}$$

Now let $F(t, j)$ be the event that B contains t, and that the number of elements of t in B in total is j - i.e. $t \in B$ and j$|B \cap T| = 2$

We can then write

$$\begin{aligned} E(p(d, B)|B \cap T &\neq 0) \\ &= \left[\sum_{i=1}^{n} \frac{1}{i} P(F(t_1, i)) E\left[p(d, B)|F(t_1, i)\right]\right] \\ &\quad + \left[\sum_{i=1}^{n} \frac{1}{i} P(F(t_2, i)) E\left[p(d, B)|F(t_1, i)\right]\right] \\ &\quad + ... \\ &\quad + \left[\sum_{i=1}^{n} \frac{1}{i} P(F(t_n, i)) E\left[p(d, B)|F(t_1, i)\right]\right] \end{aligned}$$

Now, by symmetry, it must be the case that $\frac{1}{i} P(F(t_1, i)) = \frac{1}{n}$, so we can rewrite the above as

$$
\begin{aligned}
E(p(d,B)|B\cap T &\neq 0) \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}nP(F(t_1,i))E\left[p(d,B)|F(t_1,i)\right]\right] \\
&\quad + \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}nP(F(t_2,i))E\left[p(d,B)|F(t_1,i)\right]\right] \\
&\quad + ... \\
&\quad + \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}nP(F(t_n,i))E\left[p(d,B)|F(t_1,i)\right]\right]
\end{aligned}
$$

Finally note that, for each j, if p does not exhibit choice overload, then $p(d,\{d,t_j\}) \geq p(d,B)$ for every B in $F(t_1,i)$. Therefore as $\sum_{i=1}^{n}\frac{1}{i}nP(F(t_1,i)) = 1$, we have

$$
p(d,\{d,t_j\}) \geq \sum_{i=1}^{n}\frac{1}{i}nP(F(t_1,i))E\left[p(d,B)|F(t_1,i)\right]
$$

And so we have the desired result $\square$

E Additional Results

Table 6 repeats the analysis of table 4, but replacing the independent variables with discretized versions of themselves.

E.1 No Learning Control

The following tables and figures replicate the main analysis using data from the no learning control treatment.

Table 6: The Effect of Distribution Uncertainty

	Min. Beliefs			Min. Costs		
k	**Coeff**	**s.e.**	**N**	**Coeff**	**s.e.**	**N**
1	-0.04	0.04	650	0.00	0.04	650
2	-0.09**	0.04	564	-0.04	0.04	564
3	-0.05	0.04	508	-0.03	0.05	508
4	-0.04	0.05	436	-0.04	0.05	436
5	0.01	0.05	379	-0.01	0.05	379
6	0.02	0.06	317	-0.02	0.55	317
7	-0.05	0.06	261	-0.08	0.06	261
8	-0.08	0.06	207	-0.14**	0.06	207
9	-0.05	0.07	172	-0.16**	0.07	172
10	-0.08	0.08	132	-0.18**	0.09	132
11	-0.08	0.09	107	-0.25**	0.10	107
12	-0.23**	0.11	81	-0.33***	0.11	81
13	-0.22*	0.11	64	-0.22*	0.13	64
14	-0.33**	0.14	45	-0.24*	0.14	45
15	-0.19	0.17	30	-0.27	0.17	30

Each line reports the coefficient of a regression of default choice on a dummy variable for whether minimum beliefs (left hand panel) and minimum costs (right hand panel) measured at period k are above the median, looking only at sets in which the first above-default alternative appears after position k. * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

Table 7: Default Choice by Set Size - No Learning Control

Set Size	% Default Choice	*N*
2	**26**	*47*
11	**39**	*218*
16	**41**	*310*
21	**43**	*379*

Percentage of sets that contained at least one item better than the default in which the default was chosen. Data from No Learning control treatment.

Table 8: Regression Results- No Learning Control

	Default Choice
S=11	0.134*
	(0.070)
S=16	0.154**
	(0.070)
S=21	0.177**
	(0.070)
Constant	0.255***
	(0.064)
R^2	0.006
N	954

OLS regression of default choice on set size dummies. Standard errors clustered at the subject level. * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

Table 9: Categorizing Choice Overload - No Learning Control

Size	**Choose Default**	**Search Never Starts**	N
2	26%	9%	*57*
11	39%	15%	*218*
16	41%	16%	*310*
21	43%	16%	*379*
s $\geq 2 - s = 2$	16%	7%	

For each choice set size, reports the fraction of sets that contain an above-default option in which the default was chosen (Chose Default), and the fraction of sets in which search never starts.

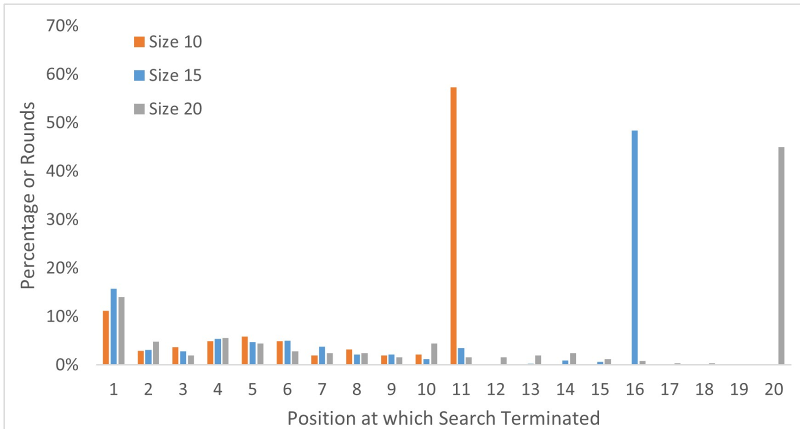


Figure 6: Histogram of search termination position for choice sets of size 10, 15 and 20 - No-learning control

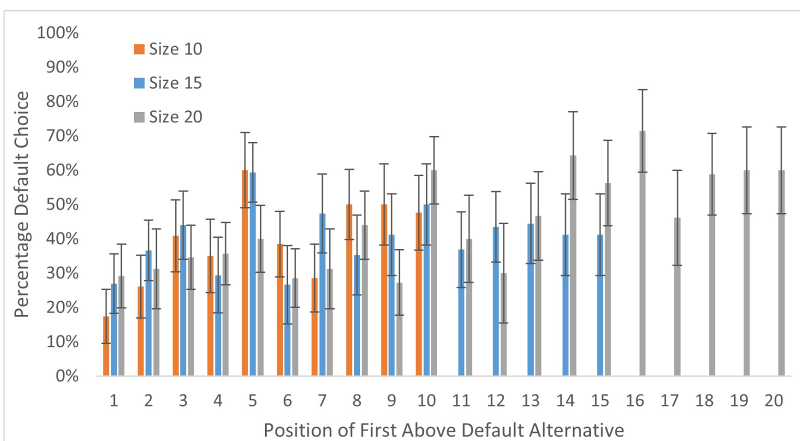


Figure 7: Effect of position of first above-default alternative on default choice - no learning control. Bars show the probability of default being chosen. Error bars represent standard errors.

Table 10: The Effect of Distribution Uncertainty - No Learning Control

	Min. Beliefs			**Min. Costs**		
k	**Coeff**	**s.e.**	**N**	**Coeff**	**s.e.**	**N**
1	-0.02	0.09	834	0.01	0.51	834
2	-0.03	0.18	765	-0.04	0.34	765
3	0.11	0.29	692	0.12	0.29	692
4	0.13	0.41	627	0.20	0.24	627
5	0.08	0.58	550	0.12	0.24	550
6	0.55	0.85	481	0.22	0.25	481
7	-1.60	1.07	425	0.10	0.25	425
8	0.28	1.43	359	0.22	0.27	359
9	0.73	1.84	302	0.34	0.28	302
10	-0.02	2.26	238	0.16	0.32	238
11	3.45	2.76	204	0.31	0.36	204
12	3.52	3.33	171	0.05	0.39	171
13	5.02	3.87	138	0.08	0.43	138
14	10.84***	3.48	107	0.36	0.50	107
15	11.86***	4.22	74	0.16	0.57	74

Each line reports the coefficient of a regression of default choice on minimum beliefs (left hand panel) and minimum costs (right hand panel) measured at period k, looking only at sets in which the first above-default alternative appears after position k. * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

Table 11: Determinants of Stopping Search - No Learning Control

	Search Stopped	
	(1)	**(2)**
Min. Cost	-0.403***	
	(0.081)	
Min. Beliefs		-0.061
		(0.038)
Searched	0.028***	0.031***
	(0.002)	*(0.002)*
Items Remaining		-0.001
		(0.002)
Above	0.156***	0.177***
	(0.016)	*(0.016)*
Const	0.199***	0.123***
	(0.023)	*(0.028)*
R^2	0.481	0.478
N	22,431	22,431
Subject f.e.	Yes	Yes

Dependent variable is whether or not search has stopped by given observation. OLS regression with standard errors clustered at the subject level * significant at 10% level, ** significant at 5% level, *** significant at 1% level.