Reference Dependence and Choice Overload

Joyee Deb and Jidong Zhou

School of Management, Yale University

2018

Reference Dependence and Choice Overload*

Joyee Deb†
Yale School of Management

Jidong Zhou‡
Yale School of Management

April 2018

Abstract

This paper offers an explanation for choice overload based on reference-dependent preferences. We assume that consumers construct an ideal object that combines the best attributes of all objects in their choice set, and use this as a reference point. They exhibit loss aversion in any attribute that is worse than the reference point. When a consumer's choice set expands, on the one hand, she is more likely to find a better object, but on the other hand, the reference point improves making all existing objects appear worse. We characterize when the latter reference-dependence effect dominates, thus making the probability of purchase decrease with the number of objects available. We also show that consumers' propensity to choose can decrease with object complexity, measured by the number of attributes.

Keywords: choice overload, reference dependence, loss aversion
JEL classification: D11, D91

---

*We are grateful to Jason Abaluck, Ravi Dhar, Laura Doval, Shane Frederick, Ryota Iijima, Alessandro Lizzeri, Barry Nalebuff, Larry Samuelson and Leeat Yariv for their helpful comments. The research assistance of Zixiong Wang is greatly appreciated.

†Email: joyee.deb@yale.edu
‡Email: jidong.zhou@yale.edu

1 Introduction

There is widespread evidence that an abundance of choice can make decision making harder and indeed may cause people to buy less often, or be less satisfied with what they purchase. This is often termed "choice overload" or "paradox of choice". Iyengar and Lepper (2000) were among the first to show, in their well-known jam experiment, that consumers when faced with a larger number of choices were less likely to purchase.[1] Similar evidence has been shown in a variety of contexts: A study by Iyengar, Huberman, and G.Jiang (2004) documented evidence of a negative impact on employee participation when a larger range of retirement plans were offered. Other things equal, every ten funds added was associated with 1.5-2% drop in participation rates. Bertrand, Karlan, Mullainathan, Shafir, and Zinman (2010) showed that offering a smaller range of loan products (1 vs 4 options) had the same effect on demand as a 25% decrease in interest rates. A survey by the Consumer Reports in 2014 also confirms that many consumers are overwhelmed by too many choices in supermarkets.[2] In fact some retailers have started to reduce the number of products they carry to make shopping easier.[3]

The evidence is, however, not unambiguous. There are also studies where choice overload does not arise. See, e.g., Scheibehenne, Greifeneder, and Todd (2010a) and Chernev, Böckenholt, and Goodman (2015) for surveys of lab and field studies of choice overload. Moreover, there is recent experimental evidence that suggests that choice overload is more likely to arise when the decision maker faces a more complex decision. For instance, Scheibehenne, Greifeneder, and Todd (2010b) conducted an experiment with pens and mp3 players, and showed that choice overload is more likely to arise when the objects are more complex in terms of having a greater number of attributes.

This body of evidence is the starting point for our paper. We present a plausible

[1] There is some earlier related research, but those papers usually consider a small number of objects and do not particularly focus on choice overload problem. For instance, Dhar (1997) studies whether the chance of taking the no-choice option is higher or lower when a new option is added to a singleton choice set, where the new option is similar to the existing one. Similarly Tversky and Shafir (1992) show experimentally, by expanding a singleton choice set, that it is possible to induce a decision maker to delay her purchase decision by adding an alternative.

[2] https://www.consumerreports.org/cro/magazine/2014/03/too-many-product-choices-in-supermarkets /index. htm

[3] See, e.g., https://www.theguardian.com/business/2015/jan/30/tesco-cuts-range-products

mechanism that can result in choice overload, and is consistent with the evidence. In particular, we propose a simple and natural model of reference dependence with loss aversion and present the conditions under which choice overload can or cannot arise. We also show how the phenomenon of choice overload is affected by the complexity of the objects.

Our main premise is that when a decision maker chooses a multi-attribute object, she needs to make trade-offs between various attributes of the object, and she faces some disutility if her chosen object is worse on some attribute than other objects that she has seen. It is as if when the decision maker observes multiple options, she imagines an ideal object that combines the best attributes of all the objects she has seen, and uses it as a reference point when she evaluates each available object. A comparison with such a reference point can leave her dissatisfied with every object in the sample. This type of reasoning has been suggested and studied by psychologists, though it has not been explored in a formal decision model. For instance, Schwartz (2004) writes:[4]

> *The existence of multiple alternatives makes it easy for us to imagine alternatives that don't exist—alternatives that combine the attractive features of the ones that do exist. And to the extent that we engage our imaginations in this way, we will be even less satisfied with the alternative we end up choosing.*

To fix ideas, consider the choice of a house. One might be interested in several attributes of a house: square footage, size of the backyard, quality of the view and style. Then, one goes house-hunting. The first open house is just the right size with a large backyard but no view, while the second is too large, with no backyard, an ocean view and constructed in mid-century style, while the third is small, has no backyard, has an ocean view and is in a modern minimalist style. As one samples more houses, one conjures up the image of an ideal house that combines the best features of each: a house that is the right size with a backyard, an ocean view and a modern minimalist design. Often, such an ideal house is not available, and purchase

[4] Sagi and Friedland (2007) provide some experimental evidence on the idea in Schwartz (2004). A related idea in psychology is that due to the contrast effect, adding an "attractive but unattainable alternative" tends to decrease the attractiveness of available alternatives, especially when the decision needs to be made soon. See, for example, the experimental study by Borovoi, Liberman, and Trope (2010).

is delayed, because every house that is available compares unfavorably with the ideal. Similar situations can arise in many examples that involve choosing multi-attribute objects such as furniture, job offers, and partners.

Formally, we consider a consumer who faces a choice set of n objects, each with m attributes. The consumer values all attributes, and her instrinsic valuations are drawn independently from some distribution. Beyond the intrinsic valuations, the consumer evaluates any object relative to a reference point which is some "ideal object" that is based on the set of choices she faces. She suffers a psychological loss that depends on how an object compares on each attribute with the reference point. A consumer will purchase her favorite object from the choice set if it gives her a positive overall utility. We say choice overload arises if the likelihood of purchase decreases with the number of options available.

We examine two natural possibilities for reference points in this context. We consider a Utopian reference point, where the consumer imagines an ideal object that combines the best attributes from all the existing objects. We also consider an Expectation-based reference point, where before inspecting the existing objects, the consumer imagines an object that has the expected best value on each attribute and regards it as the reference point. We ask two key questions. Can choice overload arise out of this reference dependence? Further, does the likelihood of purchase also vary with complexity of the object (measured by the number of attributes)?

We first show that choice overload can arise with such reference-dependent preferences, and provide sufficient conditions for this to happen. To see the intuition behind why choice overload can arise, note that increasing the size of the choice set has two effects. On the one hand, it increases the chance that better options are available, and so increases the consumer's propensity to purchase if there is no reference dependence. We call this the "sample-size effect". On the other hand, increasing the number of choices induces a better reference point, making every object in the choice set less desirable. We call this the "reference-dependence effect". When the second effect dominates, choice overload arises.

We also examine the relationship between choice overload and product complexity. Since our premise is that consumers dislike trade-offs, one may wonder whether making a decision about an object with many attributes is more likely to give rise to choice overload because there are potentially many more trade-offs.[5] As mentioned

[5]Bachi and Spiegler (2018) study a market competition model where consumers dislike trade-offs

earlier, Scheibehenne, Greifeneder, and Todd (2010b) provide experimental evidence of this. Indeed we find that a key prediction of our model is consistent with the evidence that choice overload is more likely in the case of objects with many attributes. We provide conditions under which the consumer in our model is less likely to buy when the number of attributes increases.

Other related literature. In a framework with standard consumers, it is usually hard to generate choice overload. This is because when the choice set expands, a rational consumer can always choose to ignore the newly added options. However, if the average quality of the options somehow decreases as the choice set expands, then even rational consumers can suffer from a larger choice set if they cannot easily identify each option's quality.[6] Kamenica (2008) offers a contextual inference explanation of choice overload in this vein. In his model, a firm knows that consumers have heterogeneous preferences and in equilibrium always provides the most popular varieties of a product. Then the average popularity of the available varieties decreases as the number of varieties increases. As a result, uninformed consumers who do not know their own preferences and so have to randomly choose are less likely to purchase. In a similar spirit, Kuksov and Villas-Boas (2010) offer a search model that features choice overload. They consider a Hotelling setup where there are n products located at $\frac{2i-1}{2n}$, $i = 1, \cdots, n$, respectively (which minimizes the expected consumer travelling distance when information is perfect). Consumers initially do not know which product is in which location, but can learn via a sequential search process. In such a setup, having more products implies more uncertainty of product match. When search is relatively costly, this can induce more consumers to leave the market without purchasing.[7]

Instead, we adopt a behavioral approach to explain choice overload. Our paper belongs to the large literature on choice-set dependent preferences (see, for example, Tversky and Simonson (1993) and Bordalo, Gennaioli, and Shleifer (2013)). It also relates to work in decision theory that studies preferences over menus with ex-post regret. Sarver (2008) provides axioms that deliver a unique regret representation, and can be seen as providing an axiomatic foundation for our preferences, since our

and they employ non-compensatory choice procedures in the absence of a dominant option. For example, consumers may then stick to the default option even if it is dominated.

[6] Abaluck and Gruber (2018) provide empirical evidence in this vein from the health insurance market in the state of Oregon. They show that a larger choice set makes consumers worse off mainly because it include worse choices on average.

[7] See Ke, Shen, and Villas-Boas (2016) for another search related explanation for choice overload.

utility function with a Utopian reference point can be interpreted as a particular case of the regret representation.[8] Choice overload can also naturally arise if we introduce cost of thinking and assume that the cost is higher when the choice set is larger (which requires that the decision maker cannot freely discard options). Ortoleva (2013) axiomatizes a utility function with such a feature.[9,10]

Different from the decision theory literature, our paper takes a more applied perspective. By assuming a utility function with a plausible behavioral component, we focus on studying how the number of options and the number of attributes affect the consumer choice behavior. In particular, the question of how object complexity affects choices has not been explored in the decision theory literature. We also examine how our framework can be extended to explain other commonly observed behavioral anomalies like the compromise effect and the attraction effect.

The rest of the paper is organized as follows. Section 2 contains the model. In Section 3, we illustrate the main intuition by considering a limiting case of our model with extreme reference dependence. In Section 4, we first investigate the Expectation-based reference point, because this turns out to be an easier setting to analyze. In Section 5, we discuss the Utopian reference point, and show that the results are qualitatively similar. In Section 6, we conclude with a discussion of some applications of our framework.

2 The Model

Consider a consumer who faces a choice set consisting of n multi-attribute options. Let $\mathbf{x}^j = (x_1^j, \cdots, x_m^j)$ denote option $j \in \{1, \cdots, n\}$, where x_i^j is the consumer's

[8] Buturak and Evren (2017) extend Sarver (2008) by introducing a default option that is not subject to regret consideration and study the choice overload consequence of the modified setup. This is closer to our model as we also have a normalized outside option that is not subject to the reference-dependence effect.

[9] See also Frick (2016) and Gerasimou (Forthcoming) where a larger choice set is associated with a higher complexity related cost, even if a newly added option is a dominant option or dominated by the existing ones. This contrasts with our model where adding a dominant or dominated option never harms the decision maker.

[10] Other decision theory models that predict choice overload include Ravid (2015) and Fudenberg, Iijima, and Strzalecki (2016) that both involve stochastic choice by the decision maker. Ravid (2015) proposes a boundedly rational random choice procedure. The "Focus, Then Compare" procedure has agents picking an option from their choice set at random, and then making their choice through a sequence of pair-wise comparisons.

valuation for option j's attribute i and m is the number of attributes each option has. Suppose that the consumer has an additive intrinsic utility function:

$$u(\mathbf{x}^j) = \frac{1}{m}\sum_{i=1}^{m} x_i^j \ ,$$

where for simplicity all attributes are assumed to be equally important. The qualitative results will be largely unchanged if we allow different weights on different attributes. Suppose that x_i^j is drawn independently from a distribution $F(x)$ with support $[\underline{x}, \overline{x}]$ and a continuous density $f(x)$, where an infinite support is allowed. Suppose that the realization of each x_i^j is i.i.d. across both j and i. Therefore, all options are symmetric and all attributes of each option are symmetric. The symmetry assumptions are made mainly for tractability. Throughout the paper we assume

$$\mu_0 \equiv \mathbb{E}[x_i^j] > 0 \ ,$$

i.e., on average each attribute provides a positive utility.

This random utility model can have two interpretations: We can think of modeling a single consumer whose valuations for attributes are unknown to the analyst. In this case, we are interested in the ex-ante probability that the consumer will choose an option from her choice set. Alternatively, we can think of modeling a large number of ex-ante symmetric consumers whose valuations for attributes are drawn independently from the same distribution and can be known to the analyst. In this case, we are interested in the the fraction of consumers who will choose an option from the available choice set.

The consumer evaluates each option relative to an "ideal object" or a reference point $\mathbf{r} = (r_1, \cdots, r_m)$. Specifically, the consumer's valuation for option j is

$$u(\mathbf{x}^j) - \frac{\lambda}{m}\sum_{i=1}^{m} \max\{0, r_i - x_i^j\} \ , \tag{1}$$

where $\lambda \geq 0$ is the loss aversion parameter.[11] The second term reflects the weighted sum of the psychological losses from all the attributes which are worse than the

[11]This is the simplest possible loss-aversion setup. More generally we could consider a gain/loss function $l(\mathbf{r} - \mathbf{x}^j)$ and assume a loss looms larger than a gain of the same magnitude, but the analysis would be less tractable.

reference point. The standard case of no reference dependence corresponds to $\lambda = 0$.[12]

We consider two specifications of the reference point $\mathbf{r}$:

(1) Utopian reference point. The consumer, as suggested in Schwartz's quotation above, imagines an "ideal" option which has the best attributes from all available options, and this object acts as the reference point. That is, the reference point is

$$\mathbf{r} = \mathbf{x}^1 \vee \cdots \vee \mathbf{x}^n \qquad \text{with} \qquad r_i = \max\{x_i^1, \cdots, x_i^n\}$$

Notice that r_i is a random variable *ex ante*, and it increases in n stochastically.[13]

(2) Expectation-based reference point. Here, the consumer's reference point is

$$\mathbf{r} = \mathbb{E}[\mathbf{x}^1 \vee \cdots \vee \mathbf{x}^n] \qquad \text{with} \qquad r_i = \mathbb{E}[\max\{x_i^1, \cdots, x_i^n\}] = \int_{\underline{x}}^{\overline{x}} x dF(x)^n$$

where $F(x)^n$ is the CDF of $\max\{x_i^1, \cdots, x_i^n\}$. The interpretation is that before the consumer sees the available options, she imagines an ideal object that has the expected best possible value on each attribute, and this serves as the reference point. Notice that unlike the case of the Utopian reference point, here r_i is a constant and it increases in n deterministically. This is consistent with the idea that a larger choice set induces people to anticipate a better matched option.

We assume that if the consumer does not choose any of the options, she obtains an outside option that gives her utility normalized to zero. Hence, the consumer's problem is to choose the option with the highest positive utility if any. Let P_n denote the probability that the consumer will select one of the n options.[14]

---

[12] In our model reference dependence occurs at the attribute level. This is psychologically reasonable and has been extensively adopted in the literature of prospect theory (e.g., Kahneman and Tversky (2000)).

[13] Sarver (2008) provides an axiomatic foundation for such preferences: Our utility function can be interpreted as a special case of Sarver's regret representation:

$$z^j = \frac{1}{m}\sum_{i=1}^{m}[x_i^j - \lambda(\max_k\{x_i^k\} - x_i^j)] ,$$

where i is an index of state, $\frac{1}{m}$ is the probability of each possible state, and x_i^j is option j's valuation at state i. Then the λ term captures the regret of choosing option j when state i is realized *ex post*.

[14] Alternatively, if we consider our model to be one of a large population of consumers with i.i.d. preferences, then P_n is the fraction of consumers who will select one of the n options.

Denote by

$$z_i^j \equiv x_i^j - \lambda \max\{0, r_i - x_i^j\} \tag{2}$$

the valuation for option j's attribute i after taking into account loss aversion. Let

$$H_i(t_1, \cdots, t_n) = \Pr(z_i^1 \leq t_1, \cdots, z_i^n \leq t_n)$$

be the joint CDF of $\mathbf{z}_i = (z_i^1, \cdots, z_i^n)$. Note that z_i^j and z_i^k are correlated in the case of Utopian reference point where r_i is a random variable, but independent in the case of expectation-based reference point where r_i is a constant. Whenever there is no confusion, let $H_i(t_j)$ be the marginal CDF of z_i^j.

Define

$$z^j \equiv \frac{1}{m} \sum_{i=1}^{m} z_i^j \ .$$

Let

$$H(t_1, \cdots, t_n) = \Pr(z^1 \leq t_1, \cdots, z^n \leq t_n)$$

be the joint CDF of $\mathbf{z} = (z^1, \cdots, z^n)$, and whenever there is no confusion let $H(t_j)$ be the marginal CDF of z^j. (Here, as well, z^j and z^k are correlated in the case of Utopian reference point via the correlation in each attribute.). Then we have

$$P_n = \Pr(\max\{z^1, \cdots, z^n\} > 0) = 1 - H(0, \cdots, 0) \ . \tag{3}$$

We aim to investigate the following two questions: First, how does P_n change with n? Choice overload occurs if P_n decreases with n. Second, how does P_n change with the complexity of the options which is measured by m, the number of attributes?

For the first question, the basic trade-off is as follows: As n increases, if each z^j were independent of n, $\max\{z^1, \cdots, z^n\}$ would increase stochastically due to the standard *"sample-size effect"*. However, each z^j actually decreases with n stochastically due to the *"reference-dependence effect"* since the reference point $\mathbf{r}$ improves as n increases. These two effects work in opposite directions, and it is *ex ante* unclear which effect dominates.

3 Illustration: Strong Reference Dependence

We first consider the case with $\lambda \to \infty$ to illustrate the main intuition of the paper.

First, consider the Utopian reference point. When $\lambda \to \infty$, a consumer will choose an option if and only if it is the dominant option in her choice set. This is because if an option is not the dominant one, it is worse than some other product at least in one attribute, which will make its utility go to $-\infty$. In other words, the consumer cannot tolerate any trade-off among options.[15] For a given product j with utility realization $\mathbf{x}^j = (x_1^j, \ldots, x_m^j)$, the probability that it is the dominant option is $\prod_{i=1}^m F(x_i^j)^{n-1}$. Then the ex-ante probability that product j is the dominant option is[16]

$$\int \prod_{i=1}^{m} F(x_i^j)^{n-1} dF(\mathbf{x}^j) = \prod_{i=1}^{m} \int_{\underline{x}}^{\overline{x}} F(x_i^j)^{n-1} dF(x_i^j) = \left(\frac{1}{n}\right)^m .$$

Therefore, the probability of having one dominant option out of n is

$$P_n = n \times \left(\frac{1}{n}\right)^m = \left(\frac{1}{n}\right)^{m-1} .$$

It decreases in both n (except when $m = 1$) and m. The effect of m is intuitive: For a given n, when m increases, the chance of facing trade-offs increases so the consumer is less likely to buy. However, the effect of n on P_n is more surprising. For a given $m > 1$, when n increases, each option is less likely to be the dominant option, but there are also now more options. But it tuns out that the first effect always dominates the second.

Next, consider the expectation-based reference point. When $\lambda \to \infty$, an option is not acceptable if and only if at least one of its attributes is worse than the expected maximum of that attribute. This happens with probability $1 - (1 - F(r_i))^m$, where $r_i = \int_{\underline{x}}^{\overline{x}} x dF(x)^n$. Then

$$P_n = 1 - [1 - (1 - F(r_i))^m]^n .$$

It is immediate that P_n decreases in m, since with more attributes it is more likely

[15] This is similar to assuming incomplete preferences by which consumers cannot compare two options with trade-offs and so will remain indecisive when there is no dominant option.

[16] The same argument will work here even if we allow the utility from each attribute to be drawn from different distributions.

that at least one of them will be worse than r_i. In other words, the propensity to purchase decreases with complexity of the object. What is less clear is whether choice overload can ever arise, i.e., whether P_n can decrease with n : if r_i were independent of n, P_n would increase in n, but the reference-dependence effect via r_i generates an opposite force. In fact we can show that, for sufficiently large m, the reference dependence effect dominates so that P_n decreases in n.[17] To see how, let us treat both n and m as continuous variables. Let $t(n) \equiv 1 - F(r_i)$, and note that $t'(n) < 0$. Then

$$\ln(1 - P_n) = n \ln(1 - t(n)^m) \ ,$$

and its derivative with respect to n is

$$\underbrace{\ln(1 - t(n)^m)}_{\text{sample-size effect}} + \underbrace{n(-t'(n))\frac{mt(n)^{m-1}}{1 - t(n)^m}}_{\text{reference-dependence effect}} \ . \tag{4}$$

Let $\tau = 1 - t(n)^m$, so $\tau \to 1$ as $m \to \infty$. Then

$$\lim_{m\to\infty} \frac{(1 - t(n)^m)\ln(1 - t(n)^m)}{mt(n)^m} = \lim_{\tau\to 1} \frac{\tau \ln \tau}{(1 - \tau)\log_{t(n)}(1 - \tau)} = 0 \ .$$

(The last step is from applying the L'Hospital's rule.) Therefore, for a fixed n, (4) is strictly positive (so P_n decreases in n) for sufficiently large m.

In the remainder of the paper, we consider general λ. The specification of the reference point does not affect our results qualitatively. But the Utopian reference point is analytically more challenging than the expectation-based one. So, we start with the simpler case.

4 Expectation-Based Reference Point

The consumer's expectation-based reference point in attribute i is

$$r_i = \mathbb{E}[\max\{x_i^1, \cdots, x_i^n\}] = \int_{\underline{x}}^{\overline{x}} x dF(x)^n,$$

[17]For small m, however, it is possible that P_n increases in n. For instance, we can verify that for both the uniform distribution and the exponential distribution, P_n increases in n if and only if $m = 1$.

and it is increasing in n. Using (2) we have

$$z_i^j = \begin{cases} x_i^j & \text{if } x_i^j \geq r_i \\ x_i^j - \lambda(r_i - x_i^j) & \text{if } x_i^j < r_i \end{cases} .$$

Since r_i is a constant, z_i^j is a kinked increasing function of x_i^j, and $\{z_i^1, \cdots, z_i^n\}$ are statistically independent of each other. Then the CDF of z_i^j is

$$H_i(z) = \begin{cases} F(\frac{z+\lambda r_i}{1+\lambda}) & \text{if } z \leq r_i \\ F(z) & \text{if } z > r_i \end{cases} ,$$

which has support $[(1+\lambda)\underline{x} - \lambda r_i, \overline{x}]$ and is (weakly) increasing in n. The mean of z_i^j is

$$\mu = \mu_0 - \lambda \int_{\underline{x}}^{r_i} F(x)dx , \tag{5}$$

which clearly decreases in n . Since $z^j = \frac{1}{m}\sum_{i=1}^m z_i^j$, $\{z^1, \cdots, z^n\}$ are also statistically independent of each other. The CDF of z^j, i.e. $H(\cdot)$, is an m-order convolution of $H_i(\cdot)$. Then

$$P_n = \Pr(\max\{z^1, \cdots, z^n\} > 0) = 1 - H(0)^n . \tag{6}$$

Notice that $H(0)$ increases in n, which makes it ex-ante unclear how P_n varies with n.

4.1 Choice overload

We first ask, for a fixed m, whether choice overload can arise at least for some range of n. We know that the lower bound of z^j is $(1+\lambda)\underline{x} - \lambda r_i$ with $r_i = \int_{\underline{x}}^{\overline{x}} x dF(x)^n$. We also know that r_i is increasing in n. It follows that if this lower bound is positive for $n = 1$ but negative for some $n > 1$, then we will have $P_1 = 1$ and $P_n < 1$ for some $n > 1$, which will imply that choice overload arises for some range of n. Therefore we have the following preliminary result.

Proposition 1. *For a fixed m, if $(1+\lambda)\underline{x} - \lambda\mu_0 > 0$ but $(1+\lambda)\underline{x} - \lambda r_i < 0$ for some $n > 1$ (where $r_i = \int_{\underline{x}}^{\overline{x}} x dF(x)^n$), then P_n decreases in n at least for a range of n.*

In particular, when each option has many attributes, i.e., when $m \to \infty$, it is easy to show that the probability of purchase P_n decreases weakly in n. As $m \to \infty$, each $z^j = \frac{1}{m}\sum_{i=1}^{m} z_i^j$ converges to the mean μ according to the *law of large numbers*. Then a consumer will choose one option if and only if $\mu > 0$. In this case the sample-size effect vanishes and only the reference-dependence effect remains. Given that μ decreases in n, the probability that the consumer will choose an option (weakly) decreases with n, i.e., $\lim_{m\to\infty} P_n$ decreases in n. If $\mu > 0$ for $n = 1$ and $\mu < 0$ for a sufficiently large n, then the consumer will not choose any option if the number of options exceeds a threshold.[18]

Now consider, a finite but large m such that we can approximate $H(\cdot)$ by using the *central limit theorem (CLT)*. If z_i^j has a mean μ and a variance σ^2, CLT implies that approximately $z^j - \mu \sim N(0, \frac{\sigma^2}{m})$. Then

$$H(0) = \Pr(z^j - \mu < -\mu) \approx \Phi(-\sqrt{m}\frac{\mu}{\sigma}) \ ,$$

where Φ is the CDF of the standard normal distribution. Therefore,

$$P_n = 1 - H(0)^n \approx 1 - \Phi(-\sqrt{m}\frac{\mu}{\sigma})^n \tag{7}$$

for a large m. (Notice that this formula justifies our previous discussion when $m \to \infty$.) We have the following proposition.

Proposition 2. *Let μ and σ^2 be the mean and variance of z_i^j, respectively. For a given n,*

1. *If $-\frac{\mu}{\sigma}$ decreases at n, then for m large enough, there is no choice overload.*
2. *If $\mu < 0$ and $-\frac{\mu}{\sigma}$ increases at n, there exists $\hat{m}_1$ such that $P_{n+1} < P_n$ for $m > \hat{m}_1$.*

Proof. For the first part, notice that if $-\frac{\mu}{\sigma}$ decreases at n, then P_n approximated in (7) must increase in n, such that there is no choice overload.

For the second part, let us treat n as a continuous variable. Let $\rho(n) \equiv -\frac{\mu}{\sigma}$. Then our assumption implies $\rho(n) > 0$ and $\rho'(n) > 0$. When m is sufficiently large,

[18] For instance, if x_j^i were drawn from the uniform distribution over $[0, 1]$, one can check that $\mu = \frac{1}{2}(1 - \lambda(\frac{n}{n+1})^2)$. If $\lambda \in (1, 4)$, $\mu > 0$ for $n = 1$ and $\mu < 0$ if and only if $n \geq \frac{1}{\sqrt{\lambda}-1}$. That is, P_n will drop from 1 to 0 when n exceeds $\frac{1}{\sqrt{\lambda}-1}$.

$P_n \approx 1 - \Phi(\sqrt{m}\rho(n))^n$. Consider

$$\ln \Phi(\sqrt{m}\rho(n))^n = n \ln \Phi(\sqrt{m}\rho(n)) \ .$$

Its derivative with respect to n is

$$\log \Phi(\sqrt{m}\rho(n)) + n \frac{\phi(\sqrt{m}\rho(n))}{\Phi(\sqrt{m}\rho(n))} \sqrt{m}\rho'(n) \ ,$$

where $\phi(\cdot)$ is the density of the standard normal distribution. The sign of this derivative is the same as that of

$$\frac{\log \Phi(\sqrt{m}\rho(n))}{\phi(\sqrt{m}\rho(n))} + n \frac{\sqrt{m}\rho'(n)}{\Phi(\sqrt{m}\rho(n))} \ . \tag{8}$$

For a given $\rho(n) > 0$, $\lim_{m\to\infty} \sqrt{m}\rho(n) = \infty$. By L'Hospital's rule, we have

$$\lim_{x\to\infty} \frac{\log \Phi(x)}{\phi(x)} = \lim_{x\to\infty} \frac{\phi(x)/\Phi(x)}{\phi'(x)} = \lim_{x\to\infty} \frac{\phi(x)}{\phi'(x)} = \lim_{x\to\infty} (-\frac{1}{x}) = 0 \ .$$

(The second last equality follows from the fact that $\phi'(x) = -x\phi(x)$ for the standard normal density function.) Therefore, for a fixed n, the first term in (8) tends to zero as $m \to \infty$, while the second one is positive and bounded away from zero. This implies that $\Phi(\sqrt{m}\rho(n))^n$ increases in n (and so P_n decreases in n) for a large m. $\square$

To illustrate, let us consider the uniform distribution example with $F(x) = x$ and $r_i = \frac{n}{n+1}$. The density function of z_i^j is then

$$h_i(z) = \begin{cases} \frac{1}{1+\lambda} & \text{if } z \leq r_i \\ 1 & \text{if } z > r_i \end{cases}$$

for $z \in [-\frac{n\lambda}{n+1}, 1]$. One can check that the mean of z_i^j is $\mu = \frac{1}{2}\left(1 - \lambda\left(\frac{n}{n+1}\right)^2\right)$, and its variance is $\sigma^2 = \frac{1}{3}\left(1 + \left(\frac{n}{1+n}\right)^3 (\lambda^2 - \lambda)\right) - \mu^2$. It is straightforward to verify that $-\frac{\mu}{\sigma}$ increases in n.

In Figure 1, the left panel depicts how the approximated P_n in (7) varies with n for several values of m when $\lambda = 2$. It is clear that choice overload can arise at least when n is not too large. The right panel shows how the true $P_n = 1 - H(0)^n$ varies with n (not restricting attention to large m). For $m = 5$, the two graphs are already almost identical, suggesting that our approximation works well even when m is not too large.

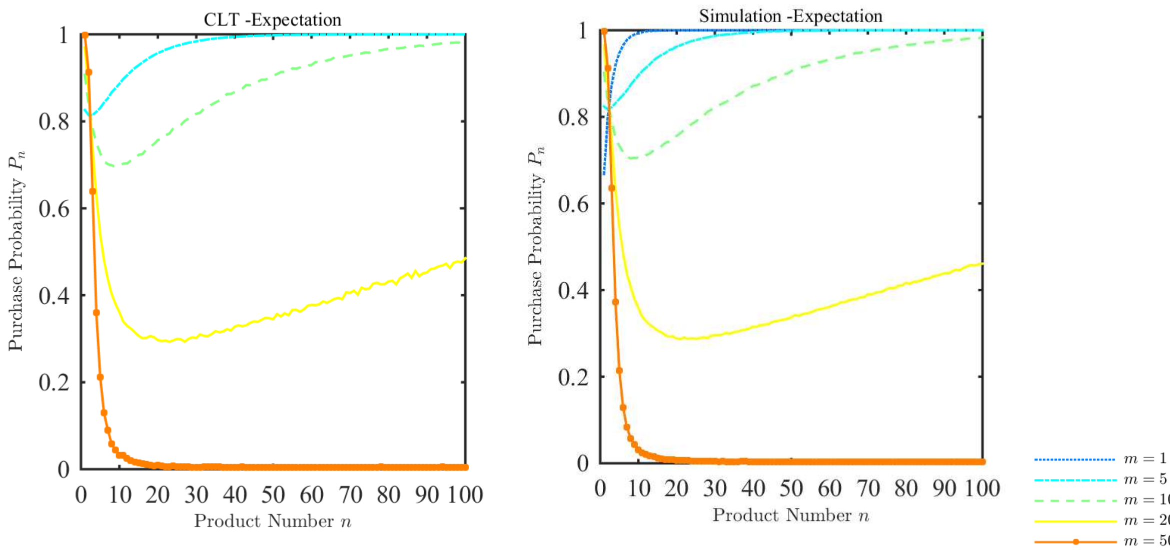


Figure 1: P_n **with expectation-based reference points and uniform distribution:** The left panel plots how P_n varies with n for large m, using the Central Limit Theorem approximation. The right panel plots the exact $P_n = 1 - H(0)^n$ for different values of m.

An observation that emerges from Figure 1 is that while P_n decreases with n in a certain range, eventually for sufficiently large n, P_n can increase again to 1. In fact, we can show that $\lim_{n\to\infty} P_n = 1$ for any fixed m if $f(x) > 0$ everywhere in $[\underline{x}, \overline{x}]$.[19] In other words, when there are enough options the sample-size effect dominates the reference-dependence effect such that there is no choice overload. However, this is not generally true if x_i^j has an unbounded support. Figure 2 plots the exact P_n for the exponential and normal distributions: In these cases, P_n always decreases in n when m is not too small.

[19]The proof is as follows: Given $f(x) > 0$ everywhere, so is $h_i(z)$. Then there exists a constant $\kappa > 0$ independent of n, such that $\Pr(z_i^j > 0) = \int_0^{\overline{x}} h_i(z)dz > \overline{x} \min_{z\in[0,\overline{x}]} h_i(z) > \kappa$. (The reason we introduce κ is that $\overline{x} \min_{z\in[0,\overline{x}]} h_i(z)$ usually depends on n.) Then $\Pr(\frac{1}{m}\sum_{i=1}^m z_i^j > 0) > \kappa^m > 0$. This implies $H(0) < 1 - \kappa^m$ for any n. Then $\lim_{n\to\infty} H(0)^n = 0$ and so $\lim_{n\to\infty} P_n = 1$.

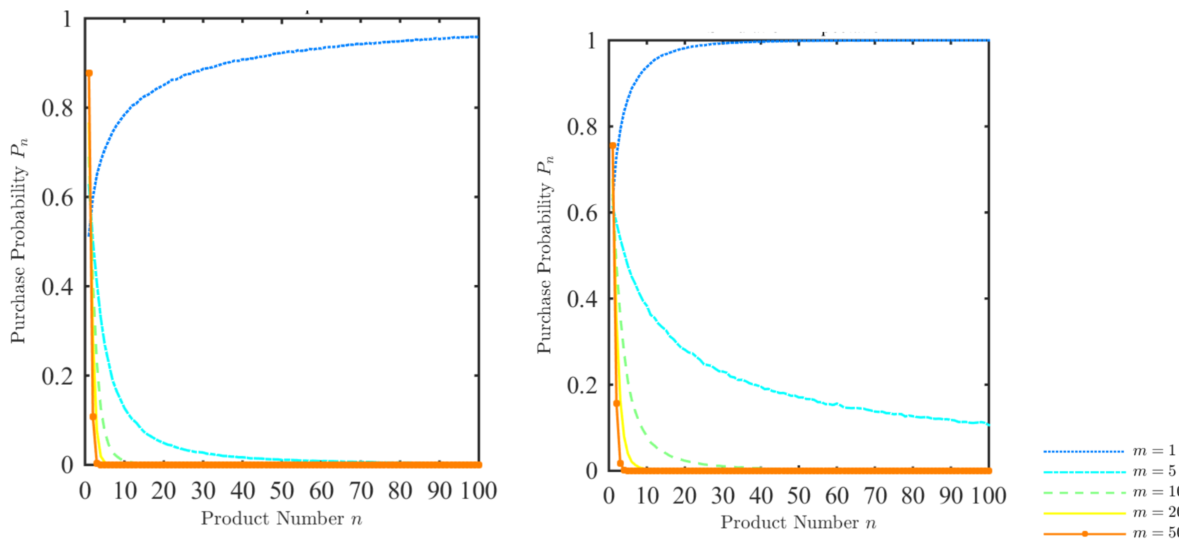


Figure 2: P_n **with expectation-based reference points under Exponential and Normal distributions:** The left panel shows how P_n varies with the number of options n for different values of m under the exponential distribution. The right panel shows the same for the normal distribution.

4.2 Impact of option complexity

Next, we turn to the question of how the purchase probability P_n changes with m. From Figures 1 and 2, we can see that P_n decreases in m when n is above a small threshold. To further understand how the number of attributes (or the complexity of the options) might play an important role in the choice overload problem, consider two cases: $m = \infty$ and $m = 1$. Recall, that in the limiting case of $m \to \infty$, choice overload arises if $\mu > 0$ for $n = 1$ and $\mu < 0$ for n large enough. Now consider the opposite case of choosing a single attribute object i.e., $m = 1$. In this case, $H = H_i$ and so we have

$$P_n = 1 - H(0)^n = 1 - F(\frac{\lambda r_i}{1+\lambda})^n \ .$$

Here both the sample-size effect and the reference-dependence effect are present. For the uniform distribution example with $F(x) = x$, one can check that $r_i = \frac{n}{n+1}$ and

$$P_n = 1 - \left(\frac{\lambda}{1+\lambda}\frac{n}{n+1}\right)^n \ .$$

Since both $(\frac{\lambda}{1+\lambda})^n$ and $(\frac{n}{n+1})^n$ decrease in n, P_n must increase with n. So, in this example with $m = 1$, the sample-size effect always dominates and there is no choice overload.

These two polar cases naturally gives rise to the question of how P_n varies with m generally. To that end, we need to know how $H(0) = \Pr(\frac{1}{m}\sum_{i=1}^{m} z_i^j < 0)$ changes in m. One may conjecture that if z_i^j has a negative mean, the sample mean $\frac{1}{m}\sum_{i=1}^{m} z_i^j$ should be negative more likely for a larger m. Conversely if z_i^j has a positive mean, the sample mean should be positive more likely for a larger m. This is, however, not true in general.[20] In the following we present sufficient conditions for this to be true: one is when m is large, and the other is when n is large.

Consider a setting with large m, so that (as we have shown before) $P_n \approx 1 - \Phi(-\sqrt{m}\frac{\mu}{\sigma})^n$. It is then clear that P_n decreases in m if $\mu < 0$ and increases in m if $\mu > 0$. We state this in the proposition below.

Proposition 3. *There exists $\hat{m}_2$ such that if $m > \hat{m}_2$, then P_n is decreasing in m if $\mu < 0$ and increasing in m if $\mu > 0$.*

In other words, if reference-dependence effect is strong enough such that $\mu < 0$ and the options are already complicated, then the probability that a consumer chooses an option from her choice set decreases as the decision becomes more complex (as m increases).

We next study the case when n is large. The following proposition characterizes when choice overload arises for symmetric and log-concave density functions.

Proposition 4. *Suppose f is log-concave and symmetric. Then if $\mu_0 < \frac{\lambda}{1+\lambda}\overline{x}$, there exists $\hat{n}_1$ such that for $n > \hat{n}_1$, P_n decreases in m. If $\mu_0 > \frac{\lambda}{1+\lambda}\overline{x}$, there exists $\hat{n}_2$ such that for $n > \hat{n}_2$, P_n increases in m.*

Proof. To prove this proposition we need a lemma.

Lemma 1. *Consider a sequence of i.i.d. random variables $\{x_1, \cdots, x_m\}$ with a common density function $f(x)$ and mean $\mathbb{E}[x]$. Suppose $f(x)$ is log-concave and symmetric about the mean. Then $\Pr(\frac{1}{m}\sum_{i=1}^{m} x_i < a)$ decreases in m if $\mathbb{E}[x] > a$ and increases in m if $\mathbb{E}[x] < a$.*

[20] Here is a simple counter example: Consider two i.i.d. binary random variables z_1 and z_2. Suppose $\Pr(z_i = -2) = \alpha$ and $\Pr(z_i = 1) = 1 - \alpha$. Then $\Pr(z_1 + z_2 = -4) = \alpha^2$, $\Pr(z_1 + z_2 = -1) = 2\alpha(1-\alpha)$, and $\Pr(z_1 + z_2 = 2) = (1-\alpha)^2$. It is easy to see that in this example $\Pr(\frac{z_1+z_2}{2} < 0) = 2\alpha - \alpha^2 > \Pr(z_i < 0) = \alpha$ for any $\alpha \in (0,1)$, regardless of whether the mean of z_i is positive or negative.

The interested reader can refer to the Appendix for a proof of the lemma. Note that in our model, the density of z_i^j is not symmetric due to the reference point, even if the density of x_i^j is. So we cannot apply this lemma directly. Nevertheless, notice that the density of z_i^j is

$$h_i(z) = \begin{cases} \frac{1}{1+\lambda} f(\frac{z+\lambda r_i}{1+\lambda}) & \text{if } z \leq r_i \\ f(z) & \text{if } z > r_i \end{cases} ,$$

and $f(\frac{z+\lambda r_i}{1+\lambda})$ is symmetric if $f(z)$ is. If n is sufficiently large such that r_i is close to $\overline{x}$, then the part of $z > r_i$ becomes negligible and $h_i(z)$ becomes almost symmetric. Then the lemma can be applied to obtain Proposition 4. See the details in the appendix. □

Notice that $\mu_0 < \frac{\lambda}{1+\lambda}\bar{x}$ is equivalent to $\lim_{n\to\infty} \mu < 0$. So this result is consistent with the above discussion that if z_i^j has a negative mean (e.g., because λ is sufficiently large), it is more likely that $\frac{1}{m}\sum_{i=1}^m z_i^j < 0$ when m increases. The conditions for this result are satisfied in both the uniform and the normal example. They are not satisfied in the exponential example where f is not symmetric, but the simulations in Figure 2 show that P_n still goes down with m when n is relatively large.

While we show above that P_n can decrease with m for a fixed n, one may ask whether increased product complexity also increases the range of n over which P_n decreases in n. Figure 1 suggests that this is indeed the case. Increased product complexity can amplify choice overload in the sense that choice overload happens for a larger range of n.

5 Utopian Reference Point

Recall that the Utopian reference point has the highest values on each attribute from the available objects in the choice set, i.e., $r_i = \max_j\{x_i^j\}$ in attribute i is a random variable, and

$$z_i^j = (1+\lambda)x_i^j - \lambda r_i .$$

The support of z_i^j is $[\underline{x}, \overline{x}]$ if $n = 1$ and $[(1+\lambda)\underline{x} - \lambda\overline{x}, \overline{x}]$ if $n \geq 2$. The mean of z_i^j is

$$\mu = (1+\lambda)\mu_0 - \lambda \int_{\underline{x}}^{\overline{x}} x dF(x)^n . \tag{9}$$

When n increases from 1 to infinity, it decreases from μ_0 to $(1+\lambda)\mu_0 - \lambda\overline{x}$. Notice that $\{z_i^1, \cdots, z_i^n\}$ are now correlated due to the random reference point, and so are $\{z^1, \cdots, z^n\}$. This is the main difference compared to the case of the expectation-based reference point. It turns out that, though this difference makes the analysis more difficult, it does not qualitatively affect the main results.

5.1 Choice overload

When $m = 1$, there is no choice overload in this setting, as we will discuss later. For $m > 1$, we have a result analogous to Proposition 1:

Proposition 5. *For a fixed $m > 1$, if $0 < \underline{x} < \frac{1}{1+\lambda}\left(\lambda - \frac{1}{m-1}\right)\overline{x}$, then P_n decreases in n at least for a range of n.*

Proof. Given $\underline{x} > 0$, we must have $P_1 = 1$ in the current setup. It then suffices to show $P_n < 1$ for some $n > 1$.[21] Given the continuity of our distribution, we only need to find that for some n there exists one realization of $\{x_1^j, \cdots, x_m^j\}_{j=1}^n$ such that every option has a negative valuation. Consider $n > m$. Suppose option j's attribute j has value $x_j^j = \overline{x}$ for $j = 1, \cdots, m$, and $x_i^j = \underline{x}$ for all other j and i. Then each of the first m options has valuation

$$\frac{1}{m}\left[\overline{x} + (m-1)((1+\lambda)\underline{x} - \lambda\overline{x})\right] < 0$$

given the stated condition, and the other options have an even lower valuation. □

As before, we next consider the case with large m such that H can be approximated by CLT. Let $\mu = (\mu, \cdots, \mu)$ be the mean of $\mathbf{z}_i = (z_i^1, \cdots, z_i^n)$, and let

$$\Sigma = \begin{pmatrix} \sigma^2 & \cdots & \sigma_{12} \\ \vdots & \ddots & \vdots \\ \sigma_{12} & \cdots & \sigma^2 \end{pmatrix} \tag{10}$$

be its covariance matrix, where σ^2 is the variance of z_i^j and σ_{12} is the covariance of (z_i^1, z_i^2). Then when m is large, $\mathbf{z} = \frac{1}{m}\mathbf{z}_i$ has approximately a multivariate normal

[21] Unlike the case of expectation-based reference point, it is now not enough to just show that the lower bound of z_i^j is negative. This is because in the current setup it is impossible that z_i^j is equal to the lower bound for all i and j.

distribution $\mathcal{N}(\mu, \frac{1}{m}\Sigma)$, and P_n can be approximated as follows.

Lemma 2. *When m is large,*

$$P_n \approx 1 - \int_{-\infty}^{\infty} \Phi\left(\frac{-\sqrt{m}\frac{\mu}{\sigma} - \sqrt{\rho}x}{\sqrt{1-\rho}}\right)^n \phi(x)dx \ , \tag{11}$$

where $\rho = \sigma_{12}/\sigma^2$ and Φ and ϕ are the CDF and density function of the standard normal distribution.[22]

Proof. Consider an n-dimensional random variable $\mathbf{x}$ which has a multivariate normal distribution $\mathcal{N}(\mu, \frac{1}{m}\Sigma)$. Define $\hat{\mathbf{x}} \equiv \frac{\sqrt{m}}{\sigma}(\mathbf{x} - \mu)$. Then $\hat{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \hat{\Sigma})$, where

$$\hat{\Sigma} = \begin{pmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}$$

with $\rho = \sigma_{12}/\sigma^2$. That is, $\hat{\mathbf{x}}$ has an *equicorrelated* multivariate normal distribution. Then

$$\Pr(\mathbf{x} < \mathbf{0}) = \Pr(\hat{\mathbf{x}} < -\frac{\sqrt{m}}{\sigma}\mu) = \int_{-\infty}^{\infty} \Phi\left(\frac{-\sqrt{m}\frac{\mu}{\sigma} - \sqrt{\rho}x}{\sqrt{1-\rho}}\right)^n \phi(x)dx \ . \tag{12}$$

The last step is from the formula of calculating orthant probability for an equicorrelated multivariate normal distribution (see, e.g., Steck and Owen, 1962).[23] When m is large, $P_n \approx 1 - \Pr(\mathbf{x} < \mathbf{0})$. □

The term inside $\Phi(\cdot)$ in (11) is complicated, so it is still hard to analytically investigate how P_n varies with n. This complication is caused by the correlation among $(z_i^1, \cdots, z_i^n)$. Intuitively z_i^1 and z_i^2 are correlated only if the realization of r_i is x_i^1 or x_i^2 (in which case they must be negatively correlated). The probability of this event decreases when n increases, and so when n is large the correlation between z_i^1

[22]If $\rho < 0$, $\sqrt{\rho}$ is a complex number and $\Phi(\cdot)$ is defined as follows:

$$\Phi(a + ib) = e^{\frac{1}{2}b^2} \int_{-\infty}^{a} e^{-itb}\phi(t)dt \ .$$

It is an integration along a path in the complex plane parallel to the a-axis from $-\infty + ib$ to $a + ib$.

[23]Notice that for a *non-equicorrelated* multivariate normal distribution, the orthant probabity $\Pr(\hat{\mathbf{x}} < \mathbf{0})$ usually does not have an analytical expression. Our result makes use of the fact that $\{z_i^1, z_i^2, \cdots, z_i^n\}$ are symmetric.

and z_i^2 should be negligible (i.e. $\rho \approx 0$). Therefore, when both m and n are large, an approximation of P_n is

$$1 - \Phi\left(-\sqrt{m}\frac{\mu}{\sigma}\right)^n ,$$

and so the result in Proposition 2 applies.

As shown in the appendix, in the uniform distribution example we have

$$\mu = \frac{1}{2} - \left(\frac{n}{n+1} - \frac{1}{2}\right)\lambda, \qquad \sigma^2 = \frac{(1+\lambda)^2}{12} - \frac{\lambda(n+1+\lambda)}{(n+2)(n+1)^2}, \qquad \sigma_{12} = -\frac{\lambda\,(n+1+\lambda)}{(n+2)\,(n+1)^2}.$$

(As discussed above, here σ_{12} is indeed negative and goes to zero as $n \to \infty$.) Figure 3 below depicts how the approximated P_n in (11) and the simulated true P_n in this example change with n when $\lambda = 2$. They are qualitatively similar to those in the case of expectation-based reference point.

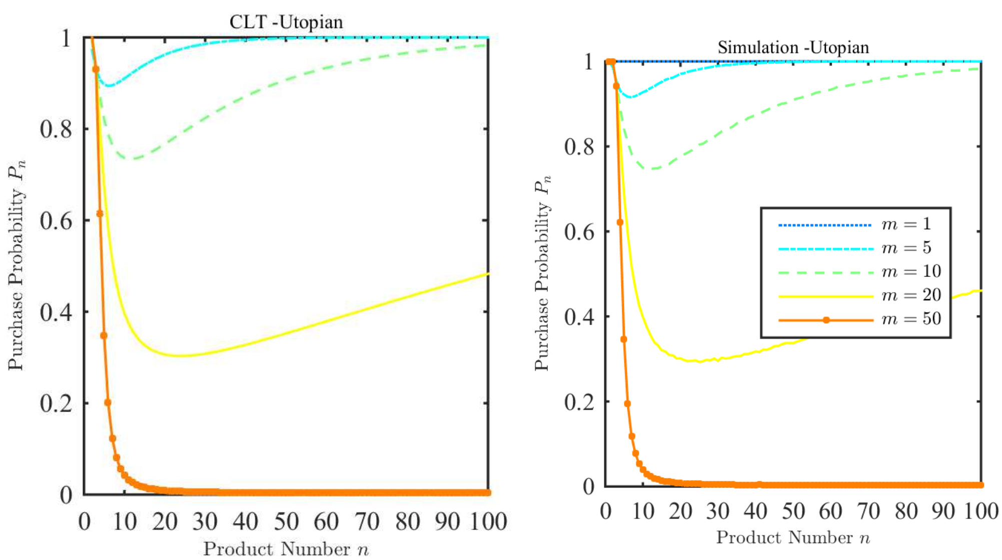


Figure 3: P_n **with Utopian reference points with the uniform distribution:** The left panel shows how P_n varies with the number of options n for large m, using the Central Limit Theorem approximation. The right panel plots the exact P_n.

The normal and exponential examples are also similar to the case of expectation-based reference point, as shown in Figure 4 below.

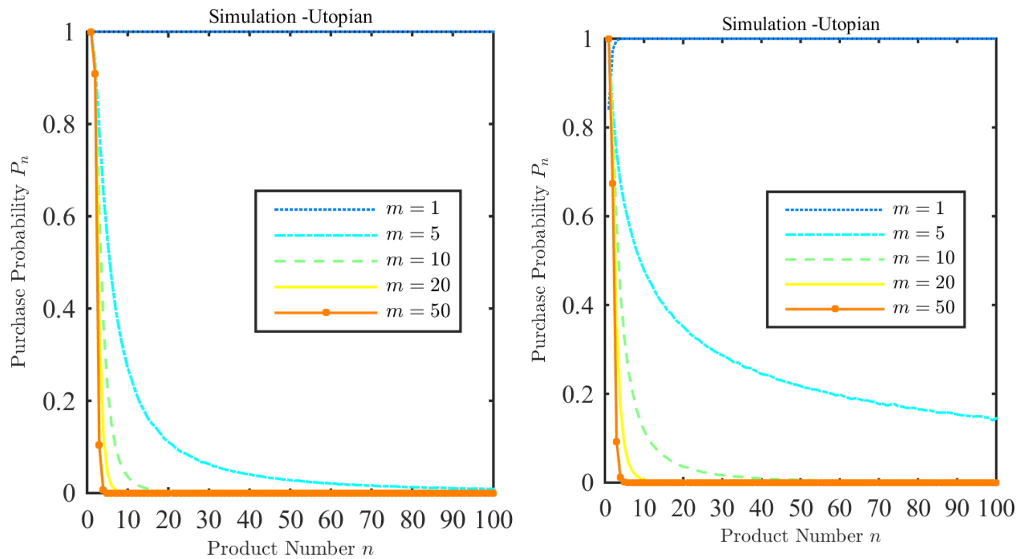


Figure 4: P_n **with Utopian reference points under exponential and normal distributions:** The left panel shows how the true P_n varies with n under the exponential distribution. The right panel does the same for the normal distribution.

5.2 Impact of option complexity

The results on impact of the complexity of the object on the probability of purchase are also analogous to the case of expectation-based reference point. When $m = 1$, there is no choice overload. To see why, note that

$$P_n = \Pr(\max\{x_1^1, \cdots, x_1^n\} \geq 0) = 1 - F(0)^n$$

is clearly increasing in n. The intuition is straightforward: With $m = 1$ there is always a dominant option, and the dominant option is not subject to the reference-dependence effect.

On the other hand, when m is large, from (11) we see that P_n decreases in m if $\mu < 0$ and increases in m if $\mu > 0$. This is also the same as in the previous case with expectation-based reference point. Finally, consider the case when n is large. A result similar to Proposition 4 holds here. The proof is in the Appendix.

Proposition 6. *Suppose that f is log-concave and symmetric. Then, if $\mu_0 < \frac{\lambda}{1+\lambda}\overline{x}$, there exists $\hat{n}_3$ such that for $n > \hat{n}_3$, P_n decreases in m. If $\mu_0 > \frac{\lambda}{1+\lambda}\overline{x}$, there exists $\hat{n}_4$ such that for $n > \hat{n}_4$, P_n increases in m.*

6 Discussion

In this paper, we propose an explanation for choice overload that stems from reference dependent preferences. When consumers use a reference point that combines the best attributes of existing choices, then increasing the number of choices improves the reference point and can make each available choice look less appealing, thus giving rise to choice overload. Below we argue that our reference dependence framework can be extended to explain other behavioral anomalies beyond choice overload.

6.1 Reference dependence and other behavioral biases

Two of the most robust departures from standard rational choice are the *compromise effect* and the *attraction effect*. Both these effects are well documented both empirically and in experimental settings (e.g. Tversky and Simonson (1993)). The compromise effect refers to the phenomenon that the introduction of an "extreme" but not inferior option into the choice set increases the probability with which the decision maker chooses an "intermediate alternative." It captures the decision maker's inclination to choose a "compromise option." The attraction effect refers to the phenomenon that the introduction of a relatively inferior (or dominated) option into the choice set increases the probability that the decision maker chooses the dominating alternative in the choice set. It captures the idea that the decision maker is attracted to options that dominate some other option in the choice set. The attraction effect was first documented by Huber, Payne, and Puto (1982) and the compromise effect by Simonson (1989). In both cases, decision makers violate the regularity in a standard choice model that the chance of an option being chosen cannot increase when the choice set is expanded. We show below that a modified version of our reference dependent preferences can also yield these two effects.

Reference dependence and the compromise effect. Consider the three two-attribute options $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ in Figure 5 below.

The compromise effect arises if adding option $\mathbf{z}$ to a decision maker's choice set increases the chance that option $\mathbf{y}$ (which is now the compromised option) is chosen. Let us now extend our reference-dependence framework by allowing the loss function to be convex (i.e., the pain from two small losses is less than the pain from a big loss which equals the sum of the two small losses). Formally, suppose the utility function

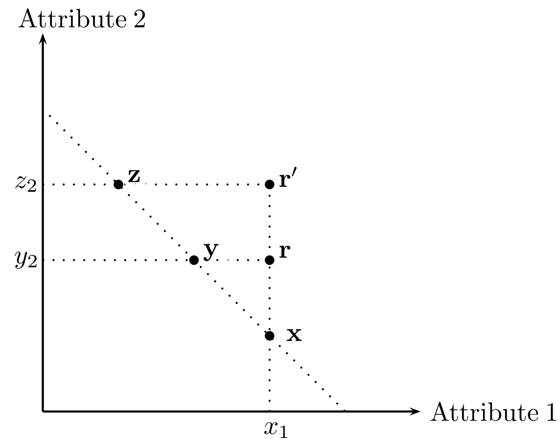


Figure 5: Reference dependence and the compromise effect

with a reference point $\mathbf{r}$ is

$$\frac{1}{m}\sum_{i=1}^{m}[x_i - l(\max\{0, r_i - x_i\})] \ ,$$

where the loss function $l(\cdot) \geq 0$ is convex and $l(0) = 0$. The previous basic model is the special case with a linear $l(\cdot)$. For a transparent illustration, we focus on the two-attribute case with $m = 2$, and assume $x_1 + x_2 = y_1 + y_2 = z_1 + z_2$ in Figure 5. Before option $\mathbf{z}$ is introduced, the reference point is $\mathbf{r} = (x_1, y_2)$ and the decision maker is indifferent between $\mathbf{x}$ and $\mathbf{y}$. After option $\mathbf{z}$ is introduced, the reference point becomes $\mathbf{r}' = (x_1, z_2)$. Then the decision maker is indifferent between $\mathbf{x}$ and $\mathbf{z}$, and the utility from each of them equals half of

$$x_1 + x_2 - l(z_2 - x_2) = z_1 + z_2 - l(x_1 - z_1) \ .$$

While the utility of $\mathbf{y}$ becomes half of

$$y_1 + y_2 - l(x_1 - y_1) - l(z_2 - y_2) \ .$$

Since $(x_1 - y_1) + (z_2 - y_2) = z_2 - x_2$ (where we have used $x_1 + x_2 = y_1 + y_2$), option $\mathbf{y}$ is perferred over $\mathbf{x}$ and $\mathbf{z}$ whenever $l(\cdot)$ is strictly convex. Intuitively, $\mathbf{x}$ now has a large disadvantage relative to $\mathbf{r}'$ on attribute 2, while $\mathbf{y}$ has two relatively small disadvantages on both attributes. The convexity of the loss function then implies that **x** is worse than **y**.

Reference dependence and the attraction effect. Consider the three two-attribute options **x**, **y**, and **z** in Figure **??** below.

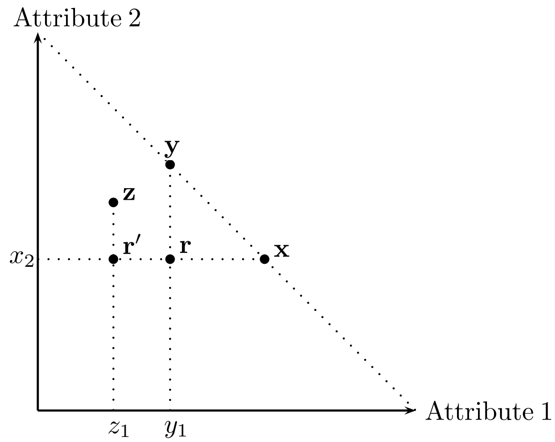


Figure 6: Reference dependence and the attraction effect

The attraction effect arises if adding option **z** (which is dominated by **y** but not **x**) to a decision maker's choice set increases the chance that option **y** is chosen. To explain this effect, we now suppose that the decision maker regards the combination of the *worst* attributes from all the existing options as the reference point **r**, and the utility function becomes

$$\frac{1}{m}\sum_{i=1}^{m}[x_i + g(\max\{0, x_i - r_i\})] \ ,$$

where the gain function $g(\cdot) \geq 0$ is concave and $g(0) = 0$. A concave gain function implies that the decision maker prefers two small gains over a big gain which equals the sum of the two small gains. Again, for expositional simplicity we focus on the two-attribute case with $m = 2$, and assume $x_1 + x_2 = y_1 + y_2$, $z_i < y_i$, $i = 1, 2$, but $z_2 > x_2$. Before option **z** is introduced, the reference point is $\mathbf{r} = (y_1, x_2)$ and the decision maker is indifferent between **x** and **y**. After option **z** is introduced, the reference point becomes $\mathbf{r}' = (z_1, x_2)$. Then the utility of **x** is half of

$$x_1 + x_2 + g(x_1 - z_1) \ ,$$

and the utility of $\mathbf{y}$ is half of

$$y_1 + y_2 + g(y_1 - z_1) + g(y_2 - x_2) \ .$$

Since $(y_1 - z_1) + (y_2 - x_2) = x_1 - z_1$ (where we have used $x_1 + x_2 = y_1 + y_2$), $\mathbf{y}$ is preferred over $\mathbf{x}$ whenever $g(\cdot)$ is strictly concave. Intuitively, $\mathbf{x}$ now has a large advantage relative to $\mathbf{r}'$ on attribute 1, while $\mathbf{y}$ has two relatively small advantages on both attributes. The concavity of the gain function then implies that $\mathbf{x}$ is worse than $\mathbf{y}$.

Notice that this second alternative framework with the worst combination as reference point can also explain the compromise effect. More generally, we can consider a utility function with two reference points $\overline{\mathbf{r}}$ (the best combination) and $\underline{\mathbf{r}}$ (the worst combination):

$$\frac{1}{m}\sum_{i=1}^{m}[x_i - \mu_l \times l(\overline{r}_i - x_i\} + \mu_g \times g(x_i - \underline{r}_i))] \ ,$$

where $l(\cdot)$ and $g(\cdot)$ have the same properties as above, and μ_l and μ_g indicate the importance of loss and gain, respectively. In particular, μ_l can be interpreted as how likely a consumer is "greedy" in the sense that she compares what is available to the best possibility, and μ_g can be interpreted as how likely a consumer is "contented" in the sense that she compares what is available to the worst possibility. This framework can account for both attraction and compromise effect, and when the loss is more important than the gain it can also account for choice overload.

6.2 Other Applications

In this paper, we restrict attention to understanding the individual decision maker's choice behavior, given that she has reference-dependent preferences. An important question is what implications such reference-dependent preferences have on firm behavior. For example, if consumers use Utopian reference points, competing firms may be led to producing similar multi-attribute products even though that intensifies price competition. Consider a situation where two firms compete in producing a multi-attribute product. Each firm faces a budget which is not enough to invest in all the potential attributes. With standard consumer preferences, firms will maximize product differentiation by investing in as many different attributes as possible, thereby weakening price competition. However, with our reference-dependent consumers the incentives for product differentiation are different. When firms invest in different attributes, consumers will imagine an ideal product which has all the potential attributes being invested as a reference point. This will make each firm's product less attractive and more consumers may thus leave the market. Taking into account this new effect, firms will have less incentive to differentiate and instead will invest in similar attributes to ensure that no firm is left behind significantly in any attribute.

Another application is product line design. Consider the classic Mussa-Rosen model where a firm can design products with different qualities to screen consumers with different willingness-to-pay for quality. If consumers have reference-dependent preferences as in our model, they will regard an ideal product which has the highest quality and the lowest price in the product line as the reference point. This will make each version of the product less attractive. Taking into account this effect, the firm will have an incentive to compress the product line. Exploring such applications is left for future research.

References

Abaluck, J., and J. Gruber (2018): "Less is More: Improving Choices by Limiting Choices in Health Insurance Markets," mimeo.

Bachi, B., and R. Spiegler (2018): "Buridanic Competition," *Games and Economic Behavior*, 107, 298–315.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010): "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *The Quarterly Journal of Economics*, 125(1), 263–306.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): "Salience and Consumer Choice," *Journal of Political Economy*, 121(5), 803–843.

Borovoi, L., N. Liberman, and Y. Trope (2010): "The effects of attractive but unattainable alternatives on the attractiveness of near and distant future menus," *Judgment and Decision Making*, 5(2), 102–109.

Buturak, G., and O. Evren (2017): "Chocie Overload and Asymmetric Regret," *Theoretical Economics*, 12(3), 1029–1056.

Chernev, A., U. Böckenholt, and J. Goodman (2015): "Choice Overload: A Conceptual Review and Meta-Analysis," *Journal of Consumer Psychology*, 25(2), 333–358.

Dhar, R. (1997): "Consumer Preference for a No-Choice Option," *Journal of Consumer Research*, 24(2), 215–231.

Frick, M. (2016): "Monotone Threshold Representations," *Theoretical Economics*, 11(3), 757–772.

Fudenberg, D., R. Iijima, and T. Strzalecki (2016): "Stochastic Choice and Revealed Perturbed Utility," Working paper version.

Gerasimou, G. (Forthcoming): "Indecisiveness, Undesirability and Overload Revealed Through Rational Choice Deferral," *Economic Journal*.

Huber, J., J. W. Payne, and C. Puto (1982): "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis," *Journal of Consumer Research*, 9(1), 90–98.

Iyengar, S. S., G. Huberman, and G.Jiang (2004): "How Much Choice is Too Much? Contributions to 401(k) Retirement Plans," in *Pension Design and Structure: New Lessons from Behavioral Finance*, ed. by O. S. Mitchell, and S. P. Utkus. Oxford University Press.

Iyengar, S. S., and M. R. Lepper (2000): "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?," *Journal of Personality and Social Psychology*, 79(6), 995–1006.

Kahneman, D., and A. Tversky (eds.) (2000): *Choices, values, and frames.* Cambridge University Press.

Kamenica, E. (2008): "Contextual Inference in Markets: On the Informational Content of Product Lines," *American Economic Review*, 98(5), 2127–2149.

Ke, T., Z. Shen, and J. Villas-Boas (2016): "Search for information on multiple products," *Management Science*, 62(12), 3576–3603.

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

Ortoleva, P. (2013): "The Price of Flexibility: Towards A Theory of Thinking Aversion," *Journal of Economic Theory*, 148, 903–934.

Proschan, F. (1965): "Peakedness of Distributions of Convex Combinations," *Annals of Mathematical Statistics*, 36(6), 1703–1706.

Ravid, D. (2015): "Focus, Then Compare," mimeo, University of Chicago.

Sagi, A., and N. Friedland (2007): "The Cost of Richness: The Effect of the Size and Diversity of Decision Sets on Post-Decision Regret," *Journal of Personality and Social Psychology*, 93(4), 515–524.

Sarver, T. (2008): "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica*, pp. 263–305.

Scheibehenne, B., P. Greifeneder, and P. Todd (2010a): "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload," *Journal of Consumer Research*, 37(3), 409–425.

——— (2010b): "Less may be more when choosing is difficult: Choice complexity and too much choice," *Acta Psychologica*, 133(1), 45 – 50.

Schwartz, B. (2004): *The Paradix of Choice: Why More is Less*. HarperCollins.

Simonson, I. (1989): "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16(2), 158–174.

Tversky, A., and E. Shafir (1992): "Choice Under Conflict: The Dyamics of Deferred Decision," *Psychological Science*, 4(6), 358–361.

Tversky, A., and I. Simonson (1993): "Context-Dependent Preferences," *Management Science*, 39(10), 1179–1189.

Appendix A

A.1 Proof of Lemma 1

Proof. Denote $y_m \equiv \sum_{i=1}^{m} x_i/m$. Notice that y_m has the same mean $\mathbb{E}[x]$ and its density is also symmetric. Given f is strictly log-concave and symmetric, Theorem 2.3 in Proschan (1965) implies that y_{m+1} is more peaked than y_m in the sense that

$$\Pr(|y_m - \mathbb{E}[x]| \leq t) < \Pr(|y_{m+1} - \mathbb{E}[x]| \leq t) \tag{13}$$

for any $0 < t < \overline{x} - \mathbb{E}[x]$.

Suppose first $\mathbb{E}[x] > a$. Substituting $t = \mathbb{E}[x] - a$ in (13) yields

$$\Pr(a \leq y_m \leq 2\mathbb{E}[x] - a) < \Pr(a \leq y_{m+1} \leq 2\mathbb{E}[x] - a) \ .$$

Then the symmetry of y_m and y_{m+1} implies that

$$\Pr(y_m < a) > \Pr(y_{m+1} < a) \ .$$

When $\mathbb{E}[x] < a$, substituting $t = a - \mathbb{E}[x]$ in (13) yields the opposite result. $\square$

A.2 Proof of Proposition 4

Proof. We suppress the superscript j for convenience. Notice that

$$H(0) = \Pr(\sum_{i=1}^{m} z_i < 0) = \Pr(\sum_{i=1}^{m} x_i < \lambda \sum_{i=1}^{m} \max\{0, r_i - x_i\}).$$

When n is sufficiently large, $r_i \approx \overline{x}$ and so

$$\Pr(\sum_{i=1}^{m} z_i < 0) \approx \Pr(\frac{1}{m} \sum_{i=1}^{m} x_i < \frac{\lambda}{1+\lambda} \overline{x}) \ .$$

According to Lemma 1, the right-hand side decreases in m (so P_n increases in m) if $\mathbb{E}[x] = \mu_0 > \frac{\lambda}{1+\lambda}\overline{x}$, and increases in m (so P_n decreases in m) if $\mathbb{E}[x] = \mu_0 \mathbb{E}[x] < \frac{\lambda}{1+\lambda}\overline{x}$. $\square$

A.3 Joint distribution with Utopian reference point

The following lemma characterizes the joint distribution of $\{z_i^1, \cdots, z_i^n\}$ in the case of the Utopian reference point.

Lemma 3. *(i) The joint CDF of* $\mathbf{z}_i = (z_i^1, \cdots, z_i^n)$ *is*

$$H_i(t_1, \cdots, t_n) = \sum_{j=1}^{n} \int_{\underline{x}}^{t_j} \prod_{k \neq j} F\left(\min\{r, \frac{t_k + \lambda r}{1+\lambda}\}\right) dF(r) \ , \tag{14}$$

where $t_j \in [(1+\lambda)\underline{x} - \lambda\overline{x}, \overline{x}]$ *if* $n \geq 2$ *and* $t_j \in [\underline{x}, \overline{x}]$ *if* $n = 1$*, and* $t \in [\underline{x}, \overline{x}]$*.*
(ii) The marginal CDF of z_i^j *is*

$$H_i(t_j) = F(t_j)^n + \int_{t_j}^{\overline{x}} F\left(\frac{t_j + \lambda r}{1+\lambda}\right) dF(r)^{n-1} \ . \tag{15}$$

(iii) The marginal CDF of (z_i^1, z_i^2) *is*

$$H_i(t_1, t_2) = F(t_1)^n + \int_{t_1}^{t_2} F\left(\frac{t_1 + \lambda r}{1+\lambda}\right) dF(r)^{n-1} + \int_{t_2}^{\overline{x}} F\left(\frac{t_1 + \lambda r}{1+\lambda}\right) F\left(\frac{t_2 + \lambda r}{1+\lambda}\right) dF(r)^{n-2} \tag{16}$$

for $t_1 \leq t_2$*, and the expression for* $t_1 > t_2$ *is analogous.*

Proof. (i) Notice that

$$H_i(t_1, \cdots, t_n) = \int_r \Pr(z_i^1 \leq t_1, \cdots, z_i^n \leq t_n | r_i = r) dF(r)^n \ ,$$

where $F(r)^n$ is the CDF of r_i. We claim that

$$\Pr(z_i^1 \leq t_1, \cdots, z_i^n \leq t_n | r_i = r) = \frac{1}{n} \sum_{j=1}^{n} \frac{\mathbb{I}_{\{t_j \geq r\}}}{F(r)^{n-1}} \prod_{k \neq j} F\left(\min\{r, \frac{t_k + \lambda r}{1+\lambda}\}\right) \ ,$$

where $\mathbb{I}_{\{\cdot\}}$ is the standard indicator function. Conditional on $r_i = r$, $x_i^j = r$ with probability $\frac{1}{n}$. In that case, $z_i^j = r$, and so $z_i^j \leq t_j$ holds iff $t_j \geq r$. This explains the indicator function term. For $k \neq j$, x_i^k must be less than r (conditional on $x_i^j = r$). Notice that

$$z_i^k = (1+\lambda)x_i^k - \lambda r \leq t_k \Leftrightarrow x_i^k \leq \frac{t_k + \lambda r}{1+\lambda} \ .$$

So the conditional probability that $z_i^k \le t_k$ is

$$\frac{F\left(\min\{r, \frac{t_k+\lambda r}{1+\lambda}\}\right)}{F(r)} \ . \tag{17}$$

Also notice that conditional $x_i^j = r$, all z_i^k, $k \neq j$, are independent of each other. Then multiplying (17) over $k \neq j$ and summing over j yeild the above expression.

Therefore,

$$H_i(t_1, \cdots, t_n) = \frac{1}{n} \sum_{j=1}^{n} \int_r \frac{\mathbb{I}_{\{t_j \ge r\}}}{F(r)^{n-1}} \prod_{k \neq j} F\left(\min\{r, \frac{t_k + \lambda r}{1+\lambda}\}\right) dF(r)^n \ .$$

This can be easily simplified to (14).

(ii) Notice that

$$H_i(t_j) = \int_r \Pr(z_i^j \le t_j | r_i = r) dF(r)^n \ .$$

Conditional on $r_i = r$, $x_i^j = r$ with probability $\frac{1}{n}$, in which case $z_i^j = r$ and so $z_i^j \le t_j$ iff $t_j \ge r$. With probability $\frac{n-1}{n}$, $x_i^j < r$ in which case the conditional probability of $z_i^j \le t_j$ is $F\left(\min\{r, \frac{t_j+\lambda r}{1+\lambda}\}\right)/F(r)$. Hence,

$$H_i(t_j) = \int_r [\frac{1}{n}\mathbb{I}_{\{t_j \ge r\}} + \frac{n-1}{n} \frac{F\left(\min\{r, \frac{t_j+\lambda r}{1+\lambda}\}\right)}{F(r)}] dF(r)^n \ .$$

This simplifies to (15).

(iii) Notice that

$$H_i(t_1, t_2) = \int_r \Pr(z_i^1 \le t_1, z_i^2 \le t_2 | r_i = r) dF(r)^n \ .$$

By a similar logic as before, the integrand is equal to

$$\frac{1}{n}\mathbb{I}_{\{r \le t_1\}} \frac{F\left(\min\{r, \frac{t_2+\lambda r}{1+\lambda}\}\right)}{F(r)} + \frac{1}{n}\mathbb{I}_{\{r \le t_2\}} \frac{F\left(\min\{r, \frac{t_1+\lambda r}{1+\lambda}\}\right)}{F(r)}$$
$$+ \frac{n-2}{n} \frac{F\left(\min\{r, \frac{t_1+\lambda r}{1+\lambda}\}\right) F\left(\min\{r, \frac{t_2+\lambda r}{1+\lambda}\}\right)}{F(r)^2} \ .$$

Due to the symmetry, we can focus on the case with $t_1 \le t_2$. Then one can readily verify (16). $\square$

It is useful to have the densify functions of z_i^j and (z_i^1, z_i^2):

$$h_i(t_j) = F(t_j)^{n-1} f(t_j) + \frac{1}{1+\lambda} \int_{t_j}^{\overline{x}} f\left(\frac{t_j + \lambda r}{1+\lambda}\right) dF(r)^{n-1} \ ,$$

and for $t_1 \leq t_2$

$$h_i(t_1, t_2) = \frac{1}{1+\lambda} f(\frac{t_1 + \lambda t_2}{1+\lambda}) f(t_2) F(t_2)^{n-2} + \frac{1}{(1+\lambda)^2} \int_{t_2}^{\overline{x}} f\left(\frac{t_1 + \lambda r}{1+\lambda}\right) f\left(\frac{t_2 + \lambda r}{1+\lambda}\right) dF(r)^{n-2} \, .$$

Now consider the uniform example with $F(x) = x$. When $n \geq 2$ the support of t is $[-\lambda, 1]$, and the support of r is $[0, 1]$. Then one can check that

$$h_i(t) = \begin{cases} \frac{1}{1+\lambda}(1 + \lambda t^{n-1}) & \text{if } t \in [0,1] \\ \\ \frac{1}{1+\lambda}(1 - (-\frac{t}{\lambda})^{n-1}) & \text{if } t \in [-\lambda, 0) \end{cases} \, .$$

Then

$$\mu = \frac{1}{2} - \left(\frac{n}{n+1} - \frac{1}{2}\right) \lambda \qquad \mathit{and} \qquad \sigma^2 = \mathbb{E}[(z_i^j)^2] - \mu^2 = \frac{(1+\lambda)^2}{12} - \frac{\lambda(n+1+\lambda)}{(n+2)(n+1)^2} \, .$$

On the other hand, one can check that when $t_1 \leq t_2$,

$$h_i(t_1, t_2) = \begin{cases} \frac{1 - \left(\frac{-t_1}{\lambda}\right)^{n-2}}{(1+\lambda)^2} & \text{if } -\lambda \leq t_1 < t_2 \leq 0 \text{ or if } -\lambda \leq t_1 < 0 < t_2 < \frac{-t_1}{\lambda} \leq 1 \\ \\ \frac{1 + \lambda t_2^{n-2}}{(1+\lambda)^2} & \text{if } -\lambda \leq t_1 < 0 < \frac{-t_1}{\lambda} < t_2 \leq 1 \text{ or if } 0 \leq t_1 < t_2 \leq 1 \end{cases} \, .$$

Then

$$\mathbb{E}[z_i^1 z_i^2] = 2 \int_{t_1 \leq t_2} t_1 t_2 h_i(t_1, t_2) dt_1 dt_2 = \frac{1}{4} \frac{1-\lambda}{n+2} \left(n(1-\lambda) + 2(1+\lambda)\right) \ ,$$

and

$$\sigma_{12} = \mathbb{E}[z_i^1 z_i^2] - \mu^2 = -\frac{\lambda \left(n+1+\lambda\right)}{\left(n+2\right) \left(n+1\right)^2} \ .$$

A.4 Proof of Proposition 6

The proof of Proposition 6 follows by deriving a multidimensional version of Lemma 1, which we present below.

Lemma 4. *Consider a sequence of i.i.d. n-dimensional random vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$, where $\mathbf{x}_i = (x_i^1, \cdots, x_i^n)$. Let $f(\mathbf{x})$ be the joint density function of $\mathbf{x}_i$ and let μ be the mean. Suppose x_i^j is symmetric across $j = 1, \cdots, n$, so that $\mu = (\mu, \cdots, \mu)$ and $f(\cdots, x_i^j, \cdots, x_i^k, \cdots) = f(\cdots, x_i^k, \cdots, x_i^j, \cdots)$ for any $j \neq k$. Suppose $f(\mathbf{x})$ is log-concave and symmetric about the mean (i.e., $f(\mathbf{x} - \mu) = f(\mu - \mathbf{x})$). Then for any constant vector $\mathbf{a} = (a, \cdots, a)$, $\Pr(\frac{1}{m}\sum_{i=1}^m \mathbf{x}_i < \mathbf{a})$ increases in m if $\mu < \mathbf{a}$ and decreases in m if $\mu > \mathbf{a}$.*

Proof. Denote

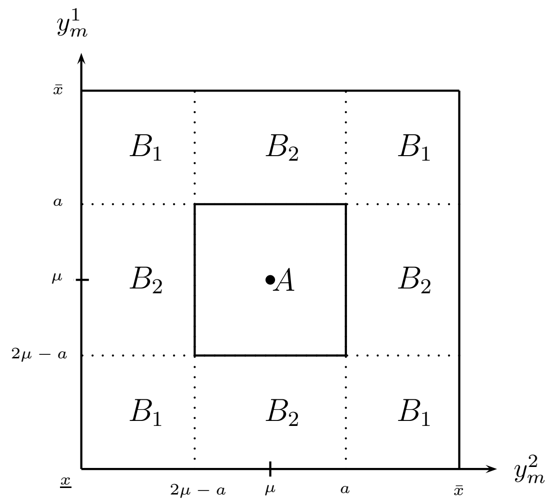
$$\mathbf{y}_m \equiv \frac{1}{m}\sum_{i=1}^{m} \mathbf{x}_i \ .$$

It is clear that $\mathbf{y}_m$ has the same mean μ and is also symmetric about μ. Given f is log-concave and symmetric, a multivariate version of Theorem 2.3 in Proschan (1965) (which is proved in Olkin and Tong, 1988) implies that $\mathbf{y}_{m+1}$ is more peaked than $\mathbf{y}_m$ in the following sense: for any compact, convex, and symmetric (about μ) $A \subset \mathbb{R}^n$ which is non-empty and is a subset of the domain of $\mathbf{x}_i$, we have

$$\Pr(\mathbf{y}_{m+1} \in A) > \Pr(\mathbf{y}_m \in A) \ . \tag{18}$$

Suppose first $\mu < \mathbf{a}$, and let A in (18) be the hypercube $[2\mu - a, a]^n$ which is centered at μ and with a corner at $\mathbf{a}$. Since A is compact, convex and symmetric about μ, $\Pr(\mathbf{y}_m \in A)$ increases in m. In the following, for convenience we consider the two-dimensional case with $n = 2$. (The argument works for the general case.) Let us divide the domain of (y_m^1, y_m^2) into multiple regions along the boundaries of A as in the figure below.

Let B_i, $i = 1, 2$, denote the probability that $\Pr(\mathbf{y}_m \in B_i)$. We want to show $3B_1 + 2B_2$ decreases in m because $\Pr(\mathbf{y}_m < \mathbf{a}) = 1 - (3B_1 + 2B_2)$. From $\Pr(\mathbf{y}_m \in A) = 1 - 4(B_1 + B_2)$ being increasing in m, we deduce that $4(B_1 + B_2)$ decreases in m, so does $B_1 + B_2$. In the same time, if we apply the multivariate version of the result by Proschan to the stripe which consists of A and two B_2's, we deduce $2B_1 + B_2$ decreases in m. Then we claim $(B_1 + B_2) + (2B_1 + B_2) = 3B_1 + 2B_2$ decreases in m.



The case with $\mu > \mathbf{a}$ can be dealt with similarly. $\square$