UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays on the Expression and Measurement of Consumer Preferences

Permalink
https://escholarship.org/uc/item/1nz8m971

Author
O'Donnell, Michael

Publication Date
2019

Peer reviewed|Thesis/dissertation

Essays on the Expression and Measurement of Consumer Preferences

By

Michael O'Donnell, Jr.

A dissertation submitted in partial satisfaction of the requirements

for the degree of

Doctor of Philosophy in

Business Administration

in the

Graduate Division of

the

University of California, Berkeley

Committee in charge:

Professor Ellen R. K. Evers, Chair
Professor Leif D. Nelson Professor
Clayton R. Critcher
Professor Serena Chen

Spring 2019

Abstract

I present three papers that each investigate an aspect of consumer preferences. In the first paper, I document preference reversals between willingness-to-pay and choice, and find that consumers are more likely to pay more for utilitarian goods but choose hedonic goods. I show that this is due, in part, to the greater level of deliberation inherent in generating willingness-to-pay values. In the second paper, I investigate how different features of choice design can lead consumers to select greater variety. I show that consumers choose variety in relation to the number of pathways to variety available to them. Finally, in the third paper, I show that consumers allow anticipated feelings of guilt about spending money to affect their preferences. I show that consumers have standards about how they believe they ought to spend money, I develop a scale to measure anticipated spending guilt, and I validate this scale by using it to predict purchase intentions above and beyond product liking.

Acknowledgements

Completing graduate school and writing a dissertation has been a bit of a long and winding road. Fortunately, I've had a number of good people to help me along the way. First, I would like to acknowledge my parents, Mike and Joy O'Donnell. They have provided me with love, support, guidance, and advice for my whole life. Thank you.

I have been fortunate to have a number of incredible mentors during my time at Berkeley. Ellen Evers has been a selfless and supportive advisor, the best collaborator I've worked with, and a kind and generous friend during the time we've been at Berkeley together. Leif Nelson helped me to find the path to success when it was not clear that I would be able to succeed in graduate school and has been a great friend and mentor. Clayton Critcher and Don Moore have been excellent collaborators, wonderful friends, and always willing to let me stop by to talk to them about research or anything else that was on my mind.

Seven years at Berkeley has meant a lot of time to make friends and meet new people, and I am thankful for the people I have found. Fausto Gonzalez has been the best friend anyone can ask for, a great roommate, an ideal traveling companion, and someone who has always been there for me. Alice Moon has likewise been a wonderful friend, a source of great advice, and one of the best people to talk to about research and life. Minah Jung has always looked out for me in ways that few other peers would do and has been someone whom I relish spending time with in Berkeley, in New York, and at conferences. Hannah Perfecto has always been someone I can laugh about life and research with and has been a great friend and colleague. Craig Anderson and Isabelle Mousseau have been the best brunch partners and Muping Gan and Simon Cartoon have been amazing dinner companions, and all have been generous and caring friends. Chris Waldo has been someone I can look forward to sharing beer with and watching comedy together. My peers at IPSR in Psychology and the Marketing and MORS groups in Haas have been excellent friends. Finally, Valerie McGraw has made my final year at Berkeley so much brighter.

Table of Contents

Data and Supplementary Materials for Chapter 1 can be found at: https://osf.io/grxy8/

Data and Supplementary Materials for Chapter 2 can be found at: https://osf.io/nfgbm/

Data and Supplementary Materials for Chapter 3 can be found at: https://osf.io/325bx/

Chapter 1: Preference Reversals in Willingness-to-Pay and Choice

Suppose someone offers you a choice between two snacks: a chocolate bar or a granola bar. You choose the chocolate bar, because it tastes good and provides a rush of sugar and fat, which can be ideal in helping you to get through a midafternoon lull. Now, however, suppose that instead of offering you a choice between the two snacks, someone asks you to indicate the most you are willing to pay to receive either product right now. Because you chose the chocolate bar, your willingness to pay (WTP) for the chocolate bar *should* be higher than your WTP for the granola bar. You should both choose what you like and also pay more for the things you like, but we find that people often do not behave this way.

Researchers rely on measures of preferences such as choices and willingness to pay because they are both face-valid and should logically converge. For instance, if someone is willing to pay up to $5 for a chocolate bar and up to $2 for a granola bar, this implies that, if given a choice, the person would choose the chocolate bar over $3 in cash and would choose $3 in cash over the granola bar (see Grether and Plott 1979 for a mathematical proof). Despite this, we find that consumers quite frequently exhibit preferences that are highly dependent on the different decision-making processes activated by different preference elicitation procedures. Accordingly, we test and show evidence for two overarching hypotheses in the studies that follow:

H1: Consumers are more likely to indicate a preference for a relatively utilitarian good over a more hedonic one when their preference is expressed through WTP as compared to choices.

H2: WTP increases preference for the utilitarian good over the hedonic good by increasing deliberation.

Our aim in this article is greater than just making a mere, albeit important, methodological point. We believe the differences in expressed preferences between choices and WTP can also teach us about the psychological processes that underlie how consumers engage in decision-making. Specifically, we find that elicitation procedures wherein participants select what they like (such as choices, but also attractiveness ratings and choice likelihood measures) tend to rely more on affect and lead to consumers preferring products that are relatively more hedonic (goods that are fun, pleasurable, and exciting, such as ice cream, chocolate, and movie tickets; Dhar and Wertenbroch 2000). However, preference elicitation procedures that require participants to generate a valuation of products, such as indicating WTP (but also selecting which product they are willing to pay more for), lead consumers to express a relatively greater preference for utilitarian goods (goods that are considered instrumental and functional, such as trash bags, toilet paper, and toaster ovens; Dhar and Wertenbroch 2000).

In the studies that follow, we test the relative role of affect and deliberation in preference elicitation procedures as a mechanism that leads to these preference reversals, as well as a number of boundary conditions. In addition to hypothetical scenarios, we used incentive-compatible settings to show that these findings are robust. We ruled out the possibility that scale differences between choices and WTP cause these preference reversals. Finally, we found that these results are related to the greater level of deliberation that participants undertake in determining a value for a good relative to simply choosing a preferred product.

AFFECT AND DELIBERATION IN PREFERENCE ELICITATION PROCEDURES

We propose that different preference elicitation procedures engender different levels of reliance on affect and deliberation in the consumer's decision-making process. For instance, when a consumer chooses between two goods, an approach that requires minimal cognitive effort (i.e., a heuristic) is to just choose whatever feels best. This is especially likely to be the case when the choice options are common and not very valuable goods. Such an affect-based heuristic, which is consistent with an intuitive System 1 decision-making process (Kahneman 2003), is a fast and low-risk approach to making simple choices (Slovic et al. 2002, 2007). Given the relationship between the use of affective-based decision-making strategies and a preference for hedonic goods (Pham 1998), we argue that when consumers are expressing a preference between a hedonic and a utilitarian good, a context that is likely to engender reliance on affect— such as choosing the preferred product—will be associated with a greater preference for the hedonic good.

However, although affective responses may help guide consumers in making a choice between two options, these responses alone are insufficient to generate an appropriate monetary value for a good. Therefore, when a consumer is asked to indicate how much she is willing to pay for a particular good, the decision must logically be influenced by more than just her affective response to the options. This process, which is consistent with a reasoned System 2 decision-making approach (Kahneman 2003), requires consumers to consider a greater amount of information and could include taking into consideration quantity, perceived value of the product, and whether it is a good use of money, among other considerations. The process of generating a WTP value for a good likely involves the synthesis of different information than making a simple choice, and therefore results in consumers expressing a different preference than a simple choice.

The notion that choices are mostly influenced by immediate affective responses is consistent with the arguments brought forward by Slovic et al. (2002, 2007). Consumers often use their affective responses as a heuristic in decision-making, simply preferring the option they deem to be more "good." For instance, findings relating similarity to liking can be explained by an affect heuristic, as in Zajonc's (1968) studies demonstrating participants' preference for unknown characters that they had been repeatedly exposed to, which increased positive evaluations and liking. This is also consistent with findings from Chaiken (1980) and Petty, Cacioppo, and Goldman (1981) that source attractiveness is a source of persuasion when the target is evaluating in a low-effort manner. Furthermore, Evers, Inbar, and Zeelenberg (2014) showed that participants who evaluated bundles of goods quickly and intuitively selected a set of goods that was the relatively more affectively pleasing option yet took considerably longer (and were often unable) to articulate how they arrived at their decisions. Taken together, these findings support an argument that people make choices in a quick, affect-driven manner. People are able to articulate preferences through choices relatively swiftly and thoughtlessly, and the tendency to rely on these thoughtless processes is consistent with selecting more affectively pleasing, or hedonic, options.

We do not believe that a more effortful decision-making process causes consumers to change their underlying preferences. Instead, we suggest that increased deliberation creates a context where consumers are likely to integrate more information and focus on different product attributes. Within this enriched context, preferences are likely to be expressed differently.

Indeed, the notion that choice and willingness to pay lead to different levels of integration of multiple considerations into the decision-making process is consistent with the elaborative process of relying on background knowledge in decision-making (Wegener et al. 2010), and with the notion that different measurement processes are likely to call upon different motives for evaluating products (Frederick and Loewenstein 2008). This relates specifically to Hypothesis 2 in that the more deliberative context of WTP puts a "brake" on the fast, affective process of choosing preferences merely based on liking. In considering how to spend money, participants might be more concerned about what is a "good" use of money (a possibility that is bolstered in data reported in Chapter 3, which suggests consumers think it is better to spend money on necessities and useful goods), and therefore be inclined to prefer a more utilitarian product (this line of reasoning could also be consistent with the possibility of consumers experiencing spending guilt, which we advance in Chapter 3). We argue that processes that are structurally similar to choices, but lead to these kinds of tradeoffs (e.g., indicating which products a person is willing to pay more for) will also show a pattern of preferences similar to WTP. This idea is consistent with the finding that when consumers use more cognitive resources, they tend to prefer goods that are superior on cognitive (utilitarian) dimensions (Shiv and Fedorikhin 1999).

RELATIONSHIP TO OTHER TYPES OF PREFERENCE REVERSALS

We believe our findings are novel because we identify a reliable preference reversal showing a divergence in expressed preferences between supposedly equivalent measures that is dependent on the use of deliberative and affective decision-making strategies. However, as there are a number of other well-documented instances of preference reversals, we believe it worthwhile to highlight why our work is similar to and different from these other cases. Importantly, we use easy-to-understand and straightforward elicitation methods for goods that many consumers are likely to encounter in their day-to-day lives and find that these measures lead to unanticipated differences.

A canonical example of preference reversals between WTP and choices occurs when participants evaluate prospective lotteries (Lichtenstein and Slovic 1971, 1973). Participants have been shown to choose higher-probability (and lower-payout) lotteries but are willing to pay more for lower-probability (and higher-payout) lotteries. This difference is usually understood as a consequence of *scale compatibility*, a mechanical explanation positing that because WTP and payouts are on the same scale, WTP values are distorted by payoff values (Slovic, Griffin, and Tversky, 1990; Tversky, Slovic, and Kahneman 1990; Tversky and Thaler 1990).

There are also some similarities between the studies we present here and the work on joint versus separate judgments (Hsee 1996; Hsee and Leclerc 1998; Shaffer and Arkes 2009). Specifically, in work on joint versus separate evaluations, participants either price or evaluate one of two different products or both products at the same time. This work claims that preference reversals emerge between these two settings because, in separate evaluations, people lack a normative standard to judge certain product attributes and end up overvaluing attributes that are more easily assessed in isolation, while in joint evaluations those attributes that are difficult to judge in isolation can easily be judged in comparison to each other. While one could argue that choices are more similar to joint evaluations than WTP, the processes that underlie joint versus separate preference reversals cannot explain the effects documented in this article. More specifically, joint-separate preference reversals are believed to occur because some dimensions of a product might be more difficult to evaluate in isolation than when the consumer has a direct

comparison. If this is the case, the manufacturer's suggested retail price (MSRP), the easiest-to-compare attribute, should weigh more heavily in choices as compared to WTP. We test for this possibility and find a preference for the hedonic good in choices regardless of whether its MSRP is displayed or not.

Finally, Nowlis and Simonson (1997) have documented a different type of preference reversal that at first glance seems to conflict with the findings from the set of studies that we present here. They find that more easily comparable attributes, such as price, are more likely to be weighted in comparative situations (such as choice), whereas "enriched" attributes weigh more heavily when participants make separate evaluations of their individual options. Notably, this would hold that MSRP should weigh heavily in choice, while enriched affective responses should weigh more heavily in separate evaluations like willingness to pay. We consistently find that this is not the case in a number of settings. We include a more detailed discussion of these differences in the General Discussion.

OVERVIEW OF STUDIES

We tested hypothesis 1 in studies 1a–1j and in studies 2a and 2b. In studies 1a–1j we present 10 scenario studies, for which we collected data online, in which participants either chose between or indicated their WTP for affective (or hedonic) and functional (utilitarian) products. We varied features of the products under consideration in these scenario studies to test for possible boundary conditions. In studies 2a and 2b, we further tested hypothesis 1 in two fully incentivized experiments. These 12 experiments provide strong and consistent evidence that WTP is associated with a relatively greater preference for a utilitarian good as opposed to a hedonic good. This can be restated such that choices are associated with a relatively greater preference for a hedonic good than a utilitarian one. Thus, the implication of these findings is that varying elicitation methods can lead consumers to maximize on distinct dimensions, leading to these preference reversals.

In studies 3a and 3b and study 4 we probed for evidence of the underlying mechanism and tested hypothesis 2. In studies 3a and 3b we tested an important boundary condition and ruled out the possibility that the differences in preferences between choices and WTP occurred merely due to mechanical, rather than psychological, determinants. Specifically, we tested for the possibility that binary or continuous response scales cause the differences between choices and WTP and found that they do not cause the preference reversals that we observe. Instead, we found—consistent with hypothesis 2—that continuous, but affective, response measures are associated with a relatively greater preference for hedonic goods, while categorical, but more deliberative, choice measures are associated with a relatively greater preference for utilitarian goods. In study 4 we tested for the role of deliberation in causing these preference reversals and found that instructing participants to deliberate causes preferences in choice to be more similar to those in WTP.

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures for each study. For all studies, we predetermined a minimum sample size of 100 participants per condition; specifically, for most of the studies we collected 100 per cell, but for those in which we expected a smaller effect size, we preregistered a larger sample size and did not analyze any data until we met our predetermined sample size. Most of the experiments that follow were preregistered. Additional robustness checks (such as testing for

order effects) and descriptive statistics (such as the average willingness to pay) for each study can be found in the web appendix.

STUDY 1: DEMONSTRATION OF WTP AND CHOICE PREFERENCE REVERSALS

Study 1 consists of 10 experiments in which we tested hypothesis 1—that choices and WTP are associated with predictable differences in how consumers express preferences between hedonic and utilitarian goods. Specifically, we expected that preferences for the utilitarian products would be higher when measured by WTP as compared to choices. While the studies share a common design, we break down the discussion of these studies into three groups to highlight what we believe are the relevant points to be gleaned from this set of experiments.

Participants and Procedure

All experiments in study 1 shared the same basic design; participants either indicated their WTP for or made a choice between a pair of goods. In each scenario, one good was more hedonic (e.g., a chocolate bar) while the other was more utilitarian (e.g., trash bags). We conducted a post-test in which a separate set of participants rated each of the items (in the pairs used in the studies) on a 10-point scale ranging from –5 = "totally utilitarian; not at all hedonic" to 5 = "totally hedonic; not at all utilitarian." For the pair of items used in each study, the participants rated the hedonic items as significantly more hedonic than the utilitarian items, except for the toasters and ice cream, which did not differ significantly from each other. For these pairs of items, we leaned on both our preregistered intuition and the existing literature suggesting that name brands are more hedonic than store brands (Ailawadi, Neslin, and Gedenk 2001). We report the full post-test in the web appendix.

Because these 10 experiments share a common design, we will explain the procedure in detail once and summarize the participants and procedure details in table 1. Participants in each experiment were workers recruited through the Amazon Mechanical Turk (MTurk) platform. All participants were paid about $.12 to take part in the experiment and were randomly assigned to either a willingness-to-pay or choice condition. Participants in each condition were shown a picture of one relatively hedonic and one relatively utilitarian good. In the WTP condition, participants were asked to indicate their willingness to pay for each item by typing their answer into a text box. We varied the wording of the WTP item across the studies, e.g. "How much would you maximally pay for [this good]?" or "If you could buy these items now, what is the most you would be willing to pay to receive each of them?" The exact wording for each study can be found in the materials posted on this paper's OSF project page. The order in which the products appeared was counterbalanced. However, as we do not find order effects in most of our studies (and when we do, they do not affect the interpretation of the results), we report these tests and their results in the web appendix.

The participants who were assigned to the choice condition were presented with the same pictures as those in the WTP condition, and the prompt, "Imagine that you were offered a choice between the following two options." They were then asked to choose the good they would prefer by answering the question, "I would prefer the:" from a list of options. These participants could indicate a preference for either item or select an option indicating no preference between the two.

Studies 1a–1d: Basic Demonstrations of Effect and Accounting for Study-Design Moderators

The first four experiments were conducted to demonstrate the basic effect. Across experiments, we varied features of the study design that could potentially moderate the effect, such as whether we included MSRP or not, and whether the products were presented vertically or horizontally.

Studies 1a–1d: Results and Discussion

We calculated preferences in the WTP conditions by comparing the indicated WTP for the two goods. We classified participants as preferring the good for which they indicated the higher WTP. If participants indicated the same WTP for both goods, they were classified as indifferent. For parsimony and because indifferent participants appear to be equally present in both conditions in nearly all of our studies, we excluded them from the analyses. Results including indifferent participants are reported in the web appendix for each study, but these analyses do not meaningfully change the results in any of these studies.

The results in table 1 show that in each of the scenarios, participants preferred the utilitarian good at a substantially greater rate in the WTP condition as compared to the choice condition. In the most extreme case, study 1d, the preference for toilet paper increased by 55.4 percentage points from 36% to 91.4%. In studies 1a–1c, the stimuli were presented in a side-by-side fashion in the choice condition, but not in WTP; this left open the possibility that when participants chose between the two options, the items were more directly comparable—that is, more like a joint evaluation—while the vertical presentation of items in WTP could have been construed as more of a separate evaluation (Hsee 1996). We tested whether the presentation format mattered in study 1d by presenting the options side by side in both conditions. Although we did not expect this to change our findings, it is reassuring that the differences between choices and WTP remain robust when this design feature is modified. We further tested for the generalizability of our findings in studies 1e–1g by varying features of the products participants are evaluating, allowing us to further test the degree to which these effects generalize.

Studies 1e–1g: Testing Product Features

Studies 1e–1g consist of three additional scenario studies. In the experiments in studies 1a–1d, participants chose between two relatively cheap and well-known common products. While we believe this to be a strength of these scenarios, because most consumer decisions involve choices between and purchasing decisions for lower-cost well-known goods, one can wonder to which degree these effects hold for different types of products. Accordingly, we tested whether the effect generalizes to products with which consumers are less familiar. Furthermore, one might wonder whether the effect is limited to food products or whether it holds for nonedible items as well. In the next three experiments, we tested whether the same effect would emerge when we used products whose market value was less known to consumers (a foreign bar of chocolate and a can opener in study 1e), products that are substantially more expensive (a toaster oven and a fancy dinner for two in study 1f), and nonfood products (a toaster and movie tickets for two in study 1g).

Studies 1e–1g: Results and Discussion

Data for studies 1e–1g were analyzed in the same way as for studies 1a–1d. Despite considerably changing key features of the products under evaluation, we found the same effect in these scenarios as in studies 1a–1d. When asked to indicate their WTP, participants expressed a much stronger preference for the more utilitarian good as compared to when they expressed their preference through choices. The results of studies 1e–1g suggest that this type of preference reversal is not limited to cheap, well-known products, but emerges across a wide variety of products. Studies 1e and 1f show that the effect occurs for moderately expensive products and that the effect is not dependent on the hedonic good being physically consumable (i.e., eaten or drunk). Finally, study 1g shows that the preference for utilitarian goods in WTP remains even when consumers evaluate products about which they are less knowledgeable.

Study 1h–1j: Testing Comparison Features

In studies 1e–1g we tested whether certain features of the products could moderate or eliminate the effect. More specifically, in previous experiments, the products involved were always extremely different and difficult to compare to each other. One product was highly utilitarian with few (if any) hedonic functions, whereas the other was highly hedonic with very little functional benefits. Because Johnson (1989) has demonstrated that evaluating noncomparable products leads to different modes of decision processing, we wanted to test that our finding generalizes to closely related products and is not merely limited to noncomparable items. Specifically, it could be possible that there is no difference in how consumers express their preference between WTP and choices, and instead there is an underlying difference in how consumers think about household goods and snack foods. Therefore, we selected two products that are substitutes of each other in study 1h to demonstrate that even when the possibility of different processing styles associated with different product types is accounted for, there is a marked distinction between preferences expressed through WTP and preferences expressed through choices. In studies 1i and 1j we also used substitutes, but in these scenarios, we used products that vary in the degree to which they are hedonic and utilitarian within the same product category.

Studies 1h–1j: Results and Discussion

The results of studies 1h–1j are presented in table 1 and are similar to those of studies 1a–1g; participants reliably exhibited a stronger preference for the utilitarian goods in the WTP conditions as compared to the choice conditions. Crucially, in each of these scenarios, the items were close substitutes for each other.

Study 1h shows that even when products are direct substitutes for each other, thereby removing the possibility that participants used different modes of processing for the products under consideration, participants still tended to prefer the hedonic option to a greater extent in choices than in WTP. Studies 1i and 1j tested whether the relative difference in how hedonic or utilitarian the two products are would lead to similar differences in choices and WTP, even if the product category under evaluation is fundamentally hedonic or utilitarian in nature. Specifically, we found that even in comparing between two utilitarian or two hedonic products, the degree to which they could be perceived as relatively more hedonic (or utilitarian) can be associated with elicitation-dependent preference reversals. These studies suggest that, if there is a difference in

the degree of how hedonic or utilitarian two products are, even if the products are fundamentally hedonic or utilitarian, the relatively more utilitarian product will still be favored at a higher rate in WTP, while the relatively more hedonic product will be favored at a higher rate in choices.

We believe that the results of these 10 experiments occur because the preference elicitation procedures themselves are leading participants to engage with the products differently, which leads to the predictable and consistent divergence in expressed preferences. Before testing the process underlying these differences in expressed preferences, we first investigate whether the same effects hold for fully incentivized decision.

Table 1. "Studies 1a-1j preference for utilitarian good, participant information, and study features."

Study	N	Preference (WTP)*	Preference (Choice)*	Significance Test*	Study Features	Key Point
1a†	300 (11 indifferent)	86.2% (81/94)	27.7% (54/195)	X^2 (1, N = 289) = 87.13, p < .001.	Trash bags (MSRP: $11.99) and two pints of ice cream ($7.98).	WTP is associated with a relatively greater preference for a utilitarian good (relative to a hedonic good) than choice.
1b†	200 (14 indifferent)	25.6% (23/90)	12.5% (12/96)	X^2 (1, N = 186) = 5.18, p = .02	Pen (MSRP: $2.33) and chocolate bar ($1.82).	Effect generalizes to stimuli in which there is a relative difference in preferences between WTP and choices. Even in the cases in which there is not a "full flip" between WTP and choices, there is a difference in preferences.
1c†	202 (9 indifferent)	89.6% (86/96)	49.5% (48/97)	X^2 (1, N = 193) = 36.55, p < .001	12-pack of Cottonelle toilet paper and pint of ben and Jerry's ice cream (No MSRP info).	Effect persists without mention of MSRP.
1d	203 (9 indifferent)	90.4% (85/94)	36.0% (36/100)	X^2 (1, N = 194) = 61.15, p < .001	12-pack of Cottonelle toilet paper and pint of ben and Jerry's ice cream. Direct replication of Scenario 3 (no MSRP info).	Effect is robust to different presentation orientations of the two goods.
1e	202 (18 indifferent)	45.9% (39/85)	27.3% (27/99)	X^2 (1, N = 184) = 6.89, p = .009	Toaster oven (MSRP: $64.99) and dinner for two (MSRP: $55) (more expensive items).	Effect emerges for moderately expensive goods (not just cheap household or grocery items).
1f	205 (13 indifferent)	79.3% (73/92)	43.0% (43/100)	X^2 (1, N = 192) = 26.47, p < .001	Toaster (MSRP: $39.99) and pair of movie tickets (MSRP: $35) (Non-consumable items).	Effect generalizes to non-food domain.

1g	202 (13 indifferent)	85.87% (79/92)	24.74% (24/97)	X^2 (1, N =189) = 71.15, $p < .001$	Can opener and Dutch chocolate bar (items that participants were unlikely to have lay beliefs about cost; no MSRP info).	Effect generalizes to domains in which consumers have less knowledge about the value of the products.
1h	203 (24 indifferent)	81.3% (65/80)	51.5% (51/99)	X^2 (1, $N = 179$) = 17.15, $p < .001$	A Kind Healthy Grains bar (relatively more utilitarian) and a Hershey's Milk Chocolate Bar (relatively more hedonic). Items are substitutes for one another (no MSRP info).	Effect persists when the products are substitutes (i.e., the differences are not due to non-comparable products engendering different processing styles).
1i	200 (14 indifferent)	47.8% (43/90)	32.3% (31/96)	X^2 (1, $N = 186$) = 4.65, $p = .03$	Three pints of Great Value chocolate ice cream (relatively more utilitarian) and a pint of Ben and Jerry's Half-Baked ice cream (relatively more hedonic; no MSRP info).	Effect persists within a hedonic product category and is dependent on the relative difference in how hedonic or utilitarian the two goods are.
1j	403 (35 indifferent)	77.3% (136/176)	51.0% (98/192)	X^2 (1, $N = 368$) = 27.29, $p < .001$	A JC Penney store brand 4-slice toaster (relatively more utilitarian; MSRP: $85) and a KitchenAid 2-slice toaster (relatively more hedonic; MSRP: $79.99)	Effect persists within a utilitarian product category and is dependent on the relative difference in how hedonic or utilitarian the two goods are.

† Indicates items were presented in a vertical orientation; all other studies used a horizontal side-by-side orientation
* The denominator and *df* for these columns are based on the sample size excluding those who were indifferent. See the Web Appendix for details.

**STUDY 2: PREFERENCE REVERSALS IN AN INCENTIVE-COMPATIBLE
CONTEXT**

The effects documented in study 1 all occurred in hypothetical situations. While this is common in many domains of consumer research, one might question whether the documented effects are a result of the hypothetical nature of the previous experiments, or whether the same effects would emerge when people make choices involving real products and bid on these products using their own money (Ding, Grewal, and Liechty, 2005). Thus, we conducted two fully incentive-compatible studies. A further advantage of these incentive-compatible studies is that they rule out the possibility, which we regard as unlikely, that participants in the hypothetical scenarios were interpreting the choice measure as being related to what they wanted in the moment, whereas those in the WTP condition were interpreting the question as asking the amount they would pay for an item when they needed it. In studies 2a and 2b, all participants received the item at the time of the study and were aware of this before indicating their preferences.

Study 2a: Participants and Procedure

We recruited 300 participants (52% female, $M_{age} = 22.4$) from public locations on the campus of a large American university and from an undergraduate subject pool in the marketing department. The sample size was set a priori to ensure that 150 participants were assigned to each of the two cells in the study.

Participants were randomly assigned to either indicate their WTP or choose between a 1-ounce bottle of Purell hand sanitizer and a single 15-gram Ghirardelli milk chocolate caramel square. All participants completed the task with a pen and paper. Participants who were assigned to the WTP condition were endowed with $1.00 in cash and given a brief primer on the Becker-DeGroot-Marschak procedure (Becker, DeGroot, and Marschak 1964) for indicating their incentive-compatible WTP. After completing a comprehension test, participants indicated their WTP independently for each product (they could bid up to $1.00 for each product). To guarantee independent bids, we informed participants that only one of the two auctions actually would be played out, and it would be determined with a coin flip. After the coin flip, we rolled a 100-sided die (two 10-sided die) to determine a price, in cents, from $.01 to $1.00. If the participant's WTP was greater than or equal to the rolled price, we sold the item to the participant at the generated price and returned the change from the initial $1.00 endowment. If the participant's WTP was less than the generated price, the participant kept the dollar. Participants in the choice condition simply chose which item they preferred to have and were given that item.

Study 2a: Results

Upon excluding participants who were indifferent in the WTP condition (those who indicated the same price for either item, $n = 17$; indifference was not an option in the choice condition), we were left with data from 283 participants. We determined preferences in the same way as in study 1, such that the chosen item or the item with the higher WTP was interpreted as the preferred product. A chi-square analysis of the preferences for these participants revealed that while 71.4% (95/133) of participants preferred the hand sanitizer in the WTP condition, only

56.0% (84/150) of participants preferred the hand sanitizer in the choice condition. This difference is statistically significant, X^2 (1, N = 283) = 7.22, p = .007.

Study 2b

While study 2a shows that the difference in relative preferences for hedonic and utilitarian goods between choices and WTP persists in an incentive-compatible setting, we designed an additional study to more conclusively demonstrate the effect in a setting where participants' choices were consequential. Because consumers are more likely to spend money on hedonic items when this money is part of a windfall, as compared to when it is not (O'Curry and Strahilevitz 2001), we designed study 2b in such a way that we could control the size of the experienced windfall in both conditions. Because controlling for the windfall results in a fairly complicated design, we describe only the crucial characteristics of the study here. For a thorough explanation of the logic underlying this reasoning and a more detailed accounting of the methods, please see the web appendix.

Study 2b was conducted in person on a central plaza of a large American university on the West Coast. To be able to control the size of the experienced windfall, we conducted this study in two waves. In wave 1, we assessed preferences for each good measuring WTP. Doing this first allowed us to establish the average subjective value for each good. Then, in wave 2, participants chose which of the two goods they would like to receive, while also receiving a small amount of extra cash to make the total value received as similar to that of wave 1 as possible. While this results in a more complicated design, it allows us to more directly equate the utility that participants receive in the two conditions, enabling a more conservative test of the effect. We again expected participants in the choice condition to be more likely to prefer the hedonic good as compared to participants in the WTP condition.

Study 2b, Wave 1: Participants and Procedure

Participants were 306 individuals (51.2% female M_{age} = 22.2, SD = 8.6) approached by research assistants on the campus of a large American university in April 2016. Participants were immediately given $2.00 in cash upon agreeing to complete the study and were informed that the $2.00 was theirs to keep, but that they had the opportunity to use the money to purchase a product from the researchers. They were then directed to a laptop on which they were shown further instructions. First, participants were taught about the BDM[1] method for eliciting WTP. Next, participants used this method in a practice round. At the end of the practice round, the participants were asked whether they understood the procedure. None of the respondents responded "no" to this question. Upon completing the practice round, they were then told that they would be bidding on both a Toblerone Swiss milk chocolate bar and a tube of Crest Complete Whitening + Scope toothpaste. The participants were further told that one of the items would be randomly selected, and we would compare their bid for that item to a randomly

---

[1] Bids were censored to a maximum of $2; 35 participants bid $2 for chocolate, 66 for the toothpaste. If censoring had any effect it would have worked against our hypothesis since it restricted bids for toothpaste more than for chocolate.

generated price and enforce the outcomes of the BDM auction.[2] Participants then indicated their WTP using the BDM procedure for both goods (counterbalanced). Once participants completed both rounds, one of the two products was selected randomly and the outcome for that product was enforced.

Study 2b, Wave 2: Participants and Procedure

Participants were 306 individuals recruited from the campus of a large American university (49.35% female, M_{age} = 22.1, SD = 7.3) in April 2016. Participants were approached on campus by the researchers and research assistants. Upon agreeing to be in the study, participants were offered a choice between two bundles; one bundle contained Toblerone chocolate plus some money, while the other contained Crest Complete toothpaste plus some money. The small amount of money was included in the bundles in an attempt to equate the experienced windfall in the WTP condition with that in the choice condition; moreover, including money in the bundles made the tradeoff between money and products more salient in the choice condition, but there was no *exchange* of money for products as there is in WTP.

For half of the participants in wave 2—wave 2a—we aimed to keep the average value of each bundle at $2.00 (as in wave 1). To achieve this, we had these participants choose either the chocolate bar + $.82, or the toothpaste + $.66. The cash amounts of $.82 and $.66 were chosen because they were the average amounts that participants in wave 1 received back in change after using their $2.00 to purchase the chocolate and toothpaste through the BDM auctions. For the other half of the participants in wave 2—wave 2b—we kept the value of each bundle below $2.00 (a lower windfall than that experienced in wave 1). To achieve this, we had these participants choose either the chocolate bar + $.27, or the toothpaste + $.11.

If the tendency for participants to choose relatively more hedonic goods in the choice condition is purely the result of windfall effects, we would expect preferences for chocolate to be lowest in the lower windfall version of the choice condition and greatest in the higher windfall version of the choice condition, with preferences in the WTP condition being between these two. On the other hand, if the differences in preferences are caused by the method by which preferences are elicited (i.e., choice vs. WTP), we should find higher preferences for the chocolate bar in *both* choice conditions as compared to the WTP condition.

Study 2b: Results

As planned, we excluded data from 19 participants in wave 1 and 13 participants in wave 2 who indicated a nut allergy. This left 287 participants in wave 1, 146 participants in wave 2a, and 147 participants in wave 2b, for a total of 293 participants in wave 2. We compared choices between the bundles in wave 2 with extrapolated choices based on the WTP values in wave 1.[3] Our method for determining extrapolated choices left us with 16 participants for whom we could not determine what they would have chosen. We excluded these participants from our analyses, leaving a final sample of 271 participants in wave 1. Overall, 59.4% of participants in wave 2 preferred a bundle containing chocolate compared to 47.6% in wave 1 (174/293 vs.

---

[2] BDM: Participants indicated their WTP for each good. Qualtrics randomly determined a price. If the price was less than or equal to the participant's WTP, the participant purchased the product and received any leftover money; otherwise the participant kept the $2.

[3] For a detailed description of how we calculated extrapolated choices in Wave 1, see the Web Appendix.

129/271). Likewise, 40.6% of participants in wave 2 preferred a bundle containing toothpaste compared to 52.4% in wave 1 (119/293 vs. 142/271). These distributions of preferences are significantly different from each other X^2 (1, $N = 564$) = 7.86, $p = .005$. Thus, participants in the choice condition preferred chocolate at a significantly greater rate than participants in the WTP condition. There was not a significant difference between wave 2a and wave 2b, X^2 (1, $N = 293$) = .299, $p = .585$. Considering wave 2a separately, we found that 61% of participants preferred chocolate ($n = 89$), whereas 39% ($n = 57$) preferred toothpaste. This is significantly different from the distribution of preferences in wave 1, X^2 (1, $N = 417$) = 6.79, $p = .01$. Restricting our analysis to wave 2b, the most conservative test, yields 57.8% of participants preferring chocolate ($n = 85$) and 42.2% preferring toothpaste ($n = 62$). This is significantly different from the distribution of preferences in wave 1, X^2 (1, $N = 418$) = 3.99, $p = .046$. Therefore, the tendency to prefer the chocolate to a greater extent in choice than in WTP cannot be explained by a windfall effect.

Studies 2a and 2b: Discussion

Studies 1a–1c and studies 2a and 2b clearly reveal that using choice as a preference elicitation procedure leads to a greater preference for hedonic goods than using WTP. Taken together, these studies provide evidence that, even in an incentive-compatible setting in which participants trade real money for real products, there is a predictable pattern of preference reversals that is dependent on how preferences are elicited. In studies 3a and 3b, we investigated whether the different expressions of preferences between choices and WTP are due to mechanical, rather than psychological, causes. We tested whether the difference is merely caused by categorical and continuous response scales, or if there are different underlying psychological contexts associated with each elicitation procedure. We believe the tendency to prefer hedonic goods in choices is due to a stronger reliance on one's affective responses to the choice options (akin to an affect heuristic) and explore this possibility in study 4.

STUDY 3: RULING OUT SCALE EFFECTS

In studies 3a and 3b we test whether the relative differences in preferences for hedonic and utilitarian goods between choices and WTP are due to the categorical or continuous natures of the response scales, or if they are caused by different modes of cognition associated with each preference elicitation procedure. For this study we test two new methods of preference elicitation, in addition to the usual choice and WTP measures. Each of these measures allows us to test whether or not the response scale itself is associated with different preferences for hedonic and utilitarian goods, or if the cognitive context created by the measure is associated with these differences.

In one of these new conditions, participants were asked to choose which product they would pay more for. If the categorical nature of the response scale is associated with the differences between choices and WTP, then this measure should look like the choice condition, and show a relatively stronger preference for the hedonic good. However, if the psychological context created by this measure requires more deliberation and a greater use of cognitive resources by inducing participants to determine a valuation rather than merely choosing the more affectively pleasing option, then this measure should look more like WTP and show a stronger preference for the utilitarian good.

In the other new condition, participants were asked to indicate how likely they are to choose each product on continuous response scales. Here, if elicitation-dependent preference reversals are caused by the use of continuous response scales, we would expect this measure to look more like WTP and show a stronger preference for the utilitarian good. However, if indicating choice likelihood taps into a similar affective process as simple choices and is associated with a preference for the more pleasurable, exciting, or fun item, then we would expect this condition to look more like choices and show a stronger preference for the hedonic good. This logic draws on Slovic et al.'s (2002, 2007) findings attractiveness ratings rely on a similar, thoughtless process, even if preferences are measured on a continuous scale. Specifically, because the measure for study 3a is merely a re-expression of a simple choice, which already taps into a more thoughtless System 1 process, then changing the scale from a categorical to a continuous measure should not affect our findings.

The designs of studies 3a and 3b allow us to begin testing for an underlying mechanism for the differences between choices and WTP. We believe that the psychological explanation, rather than the mechanical one, is likely to be supported by these studies. Specifically, we predicted that the categorical valuation measure would show results similar to WTP and the continuous choice measure would look more like the typical choice procedure. We used an identical method and analysis plan for studies 3a and 3b but used different stimuli in each study so as to enhance the generalizability of our findings.

Study 3a: Participants and Procedure

Participants were 402 workers (55.5% female M_{age} = 36.4, SD = 11.1) recruited through MTurk. Participants were randomly assigned to one of four conditions in this four-cell design. Two of the four conditions were the traditional WTP and choice conditions used in the earlier studies. In the traditional WTP condition, participants were shown a photo of a bar of Dutch chocolate (the hedonic good) and a can opener (the utilitarian good) and were asked to indicate their maximal willingness to pay to acquire each item. In the traditional choice condition, participants were asked to choose which item they would prefer to receive (indifference was an option). In the new WTP-categorical condition, participants saw the two items and responded to the question, "If you were to buy the two products pictured above, which would you pay more for?" (indifference was an option). In the new continuous choice condition, participants were prompted with, "Imagine that you are offered a choice between the following two options," were showed the two items, and responded to the question, "how likely are you to choose each product?" on a nine-point scale ranging from 1 (extremely unlikely) to 9 (extremely likely). After completing the preference elicitation procedure, all participants provided demographic information and exited the survey.[4]

---

[4] In an earlier version of this study, participants evaluated ice cream and trash bags. Participants in the WTP-categorical condition were instructed to think about their maximum WTP for each item and then select which item they would be willing to pay more money for. Likewise, in the same earlier version of the study, participants in the continuous choice condition were asked to rate how attractive a product they believed each item to be on 21-point scales ranging from 0 (not at all attractive to 20 (extremely attractive). The results in this study the same in direction and statistical significance, but we report study 3a as it is at the suggestion of anonymous reviewers.

Study 3a: Results

As planned, we excluded participants who were indifferent (in the WTP condition, those who indicated the same price for either item, $n = 6$; in the continuous choice condition, those who indicated the same purchase likelihood for each item, $n = 7$; in the choice and WTP-categorical conditions, those who selected the indifference option, $n = 1$ and $n = 7$, respectively), leaving us with data from 381 participants. We determined participants' preferences in each condition in the following ways: in the choice and WTP-categorical condition, we assumed participants preferred the selected option; in the WTP condition, we assumed participants preferred the item with the greater stated WTP; and in the continuous choice condition, we assumed participants preferred the item with the greater self-reported purchase likelihood. A chi-square analysis revealed that preference for the utilitarian good significantly differed by condition, X^2 (3, $N = 352$) = 157.82, $p < .001$ (see figure 1 for results). Preference for the utilitarian good was highest in the WTP and WTP-categorical conditions (90.4%, 85/94 in the WTP condition and 81.9%, 77/94 in the WTP-categorical condition). These proportions do not differ from each other, X^2 (1, $N = 188$) = 2.86, $p = .09$. Preference for the utilitarian good was lowest in the choice and continuous choice conditions (23.0%, 23/100 in the choice condition and 21.5%, 20/93 in the continuous choice condition). These proportions do not differ, X^2 (1, $N = 193$) = .06, $p = .80$.

Planned pairwise comparisons reveal that the proportions preferring the utilitarian good in the choice and WTP conditions are significantly different from each other, X^2(1, $N = 194$) = 89.26, $p < .001$. Similarly, the proportions in the continuous choice and WTP conditions are significantly different from each other, X^2 (1, $N = 187$) = 90.19, $p < .001$. Likewise, the proportion preferring the utilitarian good in the WTP-categorical condition is significantly different from the choice condition, X^2 (1, $N = 194$) = 67.34, $p < .001$ and from the continuous choice condition, X^2 (1, $N = 187$) = 68.34, $p < .001$.

Study 3b: Participants and Procedure

Participants were 402 workers (48.3% female M_{age} = 34.8, SD = 11.0) recruited through MTurk. Participants were randomly assigned to one of four conditions in this four-cell design. The study design and four conditions were identical to those described in study 3a. Participants in study 3b expressed their preferences for Kind Healthy Grains granola bars (the utilitarian good) and Hershey's milk chocolate bars (the hedonic good).

Study 3b: Results

As planned, we excluded participants who were indifferent (we determined which participants were indifferent in study 3b in an identical manner to study 3a; $n = 26$ in WTP, $n = 5$ in choice, $n = 13$ in WTP-categorical, $n = 6$ in continuous choice), leaving data from 352 participants. We determined participants' preferences in study 3b in an identical manner to study 3a. A chi-square analysis revealed that preference for the utilitarian good significantly differed by condition, X^2 (3, $N = 352$) = 50.56, $p < .001$ (see figure 1 for results). Preference for the utilitarian good was highest in the WTP and WTP-categorical conditions (83.8%, 62/74 in the WTP condition and 86.2%, 75/87 in the WTP-categorical condition). These proportions do not differ from each other, X^2 (1, $N = 161$) = .19, $p = .67$. Preference for the utilitarian good was lowest in the choice and continuous choice conditions (56.3%, 54/96 in the choice condition and

44.2%, 42/95 in the continuous choice condition). These proportions do not differ X^2 (1, $N = 191$) = 2.77, $p = .10$.

Planned pairwise comparisons reveal that the proportions preferring the utilitarian good in the choice and WTP conditions are significantly different from each other, $X^2(1, N = 170)$ = 14.62, $p < .001$. Similarly, the proportions in the continuous choice and WTP conditions are significantly different from each other, X^2 (1, $N = 169$) = 27.52, $p < .001$. Likewise, the proportion preferring the utilitarian good in the WTP-categorical condition is significantly different from the choice condition, X^2 (1, $N = 183$) = 19.69, $p < .001$ and from the continuous choice condition, X^2 (1, $N = 182$) = 34.89, $p < .001$.

Studies 3a and 3b: Discussion

Studies 3a and 3b demonstrate that the differences in expressed preferences between choices and WTP are not due to the categorical or continuous nature of the response scales in each elicitation procedure. The results in both of these studies provide evidence that common elicitation procedures create different psychological contexts, which are associated with participants differentially preferring hedonic and utilitarian goods. In both studies, the results from the categorical WTP condition most closely resembled the traditional WTP condition and differed significantly from both the choice and continuous choice conditions; this suggests that choosing or similar processes such as rating choice likelihood, which share similar psychological components but not mechanical components, rely more on affect and are associated with a relatively greater preference for hedonic goods. Likewise, the results from the continuous choice condition were similar to the traditional choice condition and differed from both the WTP and the WTP-categorical conditions, suggesting that processes requiring participants to consider spending money or generating a valuation rely more on deliberation and are associated with a relatively greater preference for utilitarian goods. The results from these studies are consistent with hypothesis 1 and provide initial evidence for hypothesis 2. We directly tested for the role of deliberation in contributing to the different preferences associated with each preference elicitation procedure in study 4.

Figure 1a. Results for Study 3a. Each bar shows the preference for the utilitarian good in either study, organized by preference elicitation method.
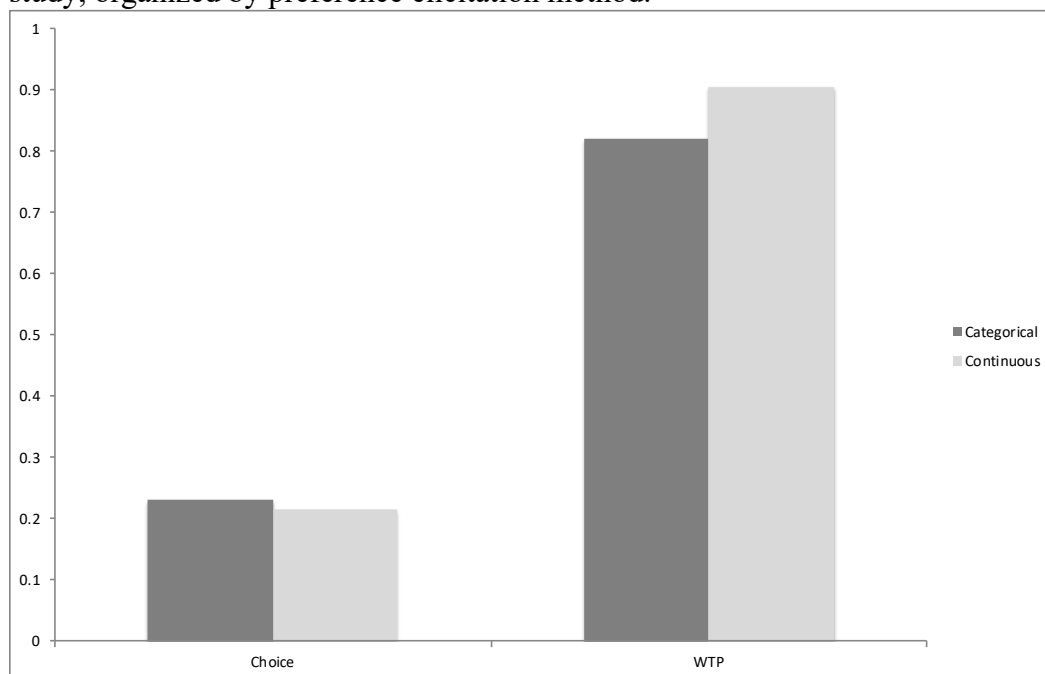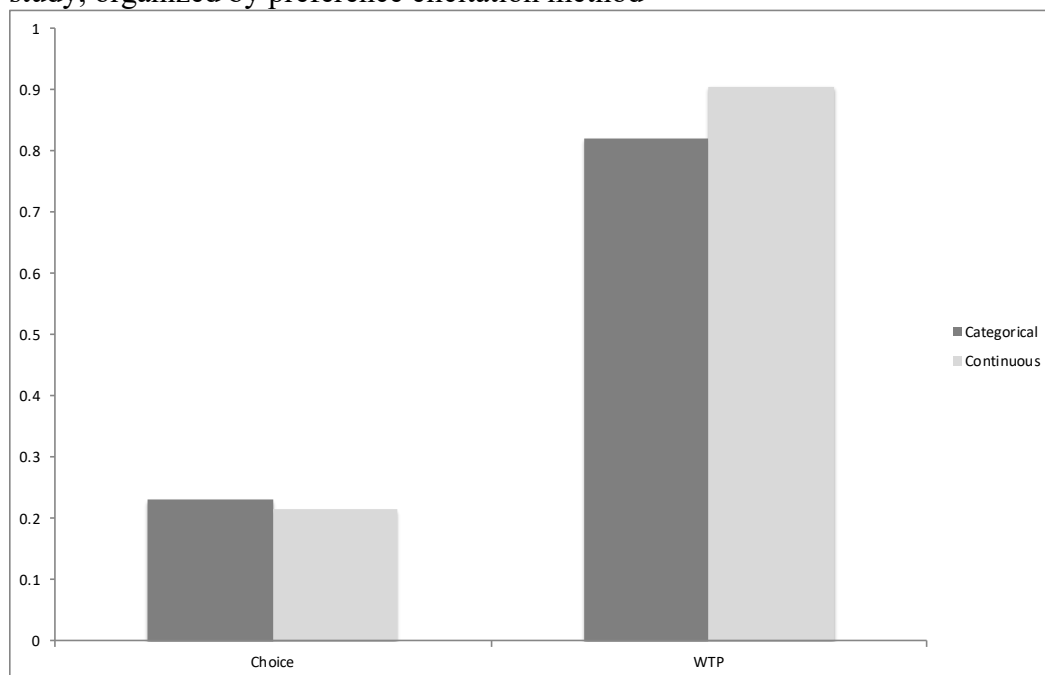

Figure 1b. Results for Study 3b. Each bar shows the preference for the utilitarian good in either study, organized by preference elicitation method

STUDY 4: TESTING FOR THE ROLE OF DELIBERATION

Thus far, we have found that participants who indicate preferences in a choice context are more likely to choose a hedonically pleasing, affectively arousing good than when indicating preferences in a WTP context. We expect this pattern occurs because choosing between two goods can be a relatively low-effort, fairly intuitive process. Each day consumers have to make an innumerable amount of choices (such as what to wear, what to have for breakfast, or which brand of cereal to buy), and having one's feelings inform the decision is a sensible and cognitively cheap heuristic that most of the time will lead to a satisfactory outcome. For WTP, however, merely relying on one's feelings is insufficient and may even feel inappropriate. Features such as quality, amount, and price might be taken into account to a greater degree than in choices. For instance, if one is actually going to put down money for something, it might be more relevant to one's WTP whether the product is higher quality or not, but in an effortless choice, these considerations could loom less largely. Therefore, we argue that if choices are inherently less thoughtful than WTP, inducing participants to deliberate by asking them to consider the value of acquiring each good should influence subsequent choices more than it influences willingness to pay. After all, indicating WTP is already expected to cause participants to think more about the value of the options under consideration, whereas choices are expected to mostly reflect gut feelings.

Consistent with the possibility that more deliberation will lead to a greater preference for the utilitarian good in a choice context, Laran and Wilcox (2011) found that the relative difference in elaboration between choosing and rejecting is associated with a change in preferences. That is, by inducing participants to reject, rather than choose, different alternatives, Laran and Wilcox increased participants' deliberation about different alternatives and found a change in expressed preferences. Therefore, we expect that by inducing participants to deliberate, the increased elaboration over options will cause participants' behavior in the choice condition to look more like their behavior in the WTP condition. We tested this moderating role of deliberation in determining product values in study 4.

Study 4: Participants and Procedure

Participants were 402 workers (43.5% female M_{age} = 32.8, SD = 10.3) recruited through MTurk. Participants were randomly assigned to one of four conditions in this 2 (WTP vs. choice) × 2 (control vs. deliberation) design. Participants in the WTP conditions were presented the same items as used in study 1a (ice cream and trash bags) and were asked to indicate their WTP for each good, while the other half was asked to choose between the same products. For the manipulation of deliberation, we adapted a thought listing procedure used by Dhar and Wertenbroch (2000), which has been shown to attenuate other types of hedonic and utilitarian good preference asymmetries. We believe that by merely requiring participants to engage more deeply with their decisions, this will slow down the faster affective process that underlies choices. In this study, half the participants in the WTP condition were asked to write a sentence listing one positive thing about each product before indicating their WTP, while the other half indicated only WTP. Similarly, half the participants in the choice condition were asked to list a benefit of choosing each item before making a choice between the products, whereas the other half of participants only made a choice between them. The deliberation prompt differed slightly between conditions to so as to use language that would better accommodate the WTP and choice tasks. For all conditions, order of the products was counterbalanced.

Study 4: Results

First, we calculated preferences in the WTP conditions by comparing the indicated WTP for the ice cream with the indicated WTP for the trash bags. Upon excluding participants who indicated indifference, $n = 15$, we were left with 387 responses.

We collapsed across presentation order (there were no order effects; see web appendix) and conducted a logistic regression of the chosen good on dummy variables representing willingness to pay versus choice, deliberation condition, and the interaction between these terms. Consistent with hypothesis 1, there is a main effect of whether participants were asked to indicate their willingness to pay or to directly choose, such that those who made a choice were more likely to select ice cream than those who indicated their WTP, $z = 7.76$, $p < .001$ (see figure 2). There is no main effect of deliberation, $z = .36$, $p = .72$, but the interaction between deliberation and elicitation method was marginally significant; $z = 1.89$, $p = .059$. Probing this interaction, we found that within the WTP conditions, participants did not differ in their preferences, with 81% (81/100) of participants in the nondeliberation condition and 83% (78/94) in the deliberation condition choosing the trash bags X^2 (1, $N = 194$) = .13 $p = .72$). However, consistent with our expectations, in the choice conditions participants who deliberated were significantly more likely to choose the trash bags (44.2%, 42/95) as compared to participants who did not deliberate (21.4%, 21/98) X^2(1, $N = 193$) = 11.39 $p = .001$.

Study 4: Discussion

If consumers are less likely to think about value when choosing between products as compared to indicating WTP, then deliberation should affect choices more than WTP. The results of study 4 provide evidence for this hypothesis showing that deliberation not only attenuated the difference in preferences between WTP and choice but did so by increasing preferences for the more utilitarian option in the choice condition while not affecting WTP. Taken together, the results of studies 3a and 3b and study 4 show that deliberation plays a role in causing participants to prefer hedonic and utilitarian items at different rates in choice and WTP contexts. Specifically, we find that when participants deliberate before making a choice, their preferences look more like those participants who indicate WTP, in that they show a relatively greater preference for the utilitarian good.

Figure 2. Preference for utilitarian good by elicitation method and deliberation, Study 4.
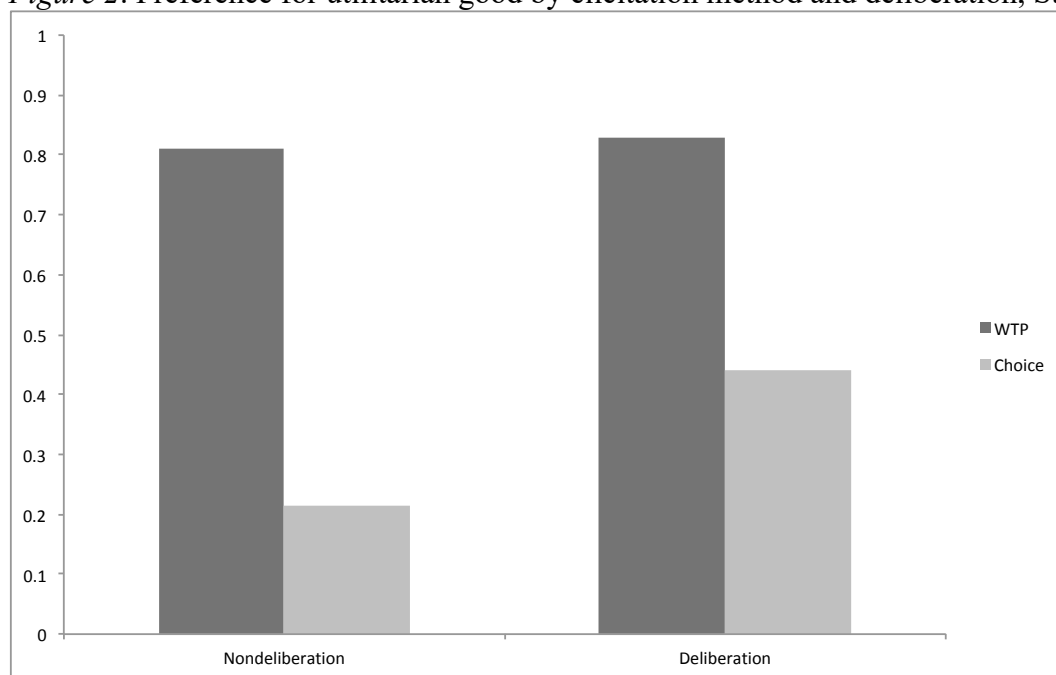
GENERAL DISCUSSION

In a set of 15 experiments, we find strong evidence that using different preference elicitation procedures leads to differential expressions of preference in a consistent way. The use of a more deliberative measure of preferences, such as WTP, appears to lead to a relatively greater preference for a utilitarian rather than a hedonic good, in comparison to the use of a more affective measure of preferences, such as choices. We documented this phenomenon in online scenario experiments (study 1) and in fully incentive-compatible field sample experiments (studies 2a and 2b). We ruled out the possibility that these findings are driven by differences in response scales (studies 3a and 3b), and instead found that the psychological context of different preference elicitation procedures is associated with participants expressing different preferences. Finally, we tested for the role of deliberation in contributing to these results, and we found that deliberation attenuates this difference between choices and WTP (study 4), indicating that giving more attentive thought to the decision under consideration leads preferences in a choice condition to more closely resemble the WTP condition.

One may wonder to what degree these effects are caused by participants' beliefs about the market value of the goods. There are two ways in which market value could influence the results reported in this article: either market value was incorporated more strongly in preferences expressed through willingness to pay as compared to choices, or our participants misunderstood the willingness-to-pay question and instead just reported the market value of the goods. The latter interpretation seems unlikely based on the results of studies 2a and 2b. Specifically, when participants paid for the goods with their own money and when they were confronted with the outcomes of their answers, we still found evidence for preference reversals.

In the former interpretation, however, perceived market value *could* play a role if participants believe it is more appropriate to incorporate market value in WTP than it is in choices. One possibility is that market value could be a useful starting point for participants to consider in articulating WTP. Studies 2a and 2b would also cut against this possibility that this accounts for the entirety of our effect by matching participants' ideographic incentive compatible WTP values with random selling prices. That this possibility is in some measure responsible for the difference between WTP and choice is not incompatible with our data, but in essence is not more than an (incomplete) re-description of our findings. Specifically, we argue that choice and WTP differ because consumers take into account different considerations when expressing their preferences in the context of these elicitation methods. Market value—as well as other attributes such as quantity, perceived quality, and opportunity costs—is likely to be among the considerations that are weighted more in WTP than in choice. Indeed, if market value truly matters more in WTP than in choice, this is merely stronger evidence that WTP and choice are better understood as different situations, rather than identical measures of preference.

Implications for Understanding Preferences

We focused primarily on choices and WTP because they are ubiquitous measures of preferences in consumer behavior research. Indeed, one does not need to look back much earlier than the recent past to find that choices and WTP are frequently used to measure consumers' preferences. A cursory review of recent consumer behavior articles published in the *Journal of Consumer Research* shows that numerous findings rely on participants' choices between consumer goods (He and Bond 2015; Klesse, Levav, and Goukens 2015; Tully et al. 2015; Yang and Urminsky 2015); their WTP for different goods (Faraji-Rad, Samuelsen, and Warlop 2015; Hsee, Yang, and Ruan 2015; Jiang et al. 2015; Kim and Kramer 2015; Kupor and Tormala 2015; Lee et al. 2016); or sometimes results from both measures (O'Guinn, Tanner, and Maeng 2015; Smith, Newman, and Dhar 2015) as representing their participants' preferences. Given that researchers so frequently use these measures to assess preferences and often use them interchangeably to study behavior, we believe that identifying that they are associated with different expressions of preferences and understanding the different cognitive processes that underlie these differences will be a consequential contribution to consumer behavior research.

One could interpret, based on our findings, that preferences are not stable. We do not believe our participants are violating the axiom of transitivity in a way that undermines rational choice theory. Although it is undoubtedly true that participants' responses to these measures are at least partially constructed on the fly, particularly in a hypothetical setting, there is no reason to doubt that in an incentive-compatible setting these measures are a clear indicator of consumers' preferences. Instead, we believe it is likely that the assumptions that lead us to believe choice and WTP are interchangeable are not valid. For instance, choices appear to be a good measure of preferences as they are face-valid—that is, people choose what they like. Likewise, WTP should also be a sensible measure of preferences because money is fungible and can be exchanged for things that are desirable, meaning it should scale relatively well as a measure of how much a good is wanted. However, while these measures should converge on participants expressing the same preferences, we find they do not, and believe this is because they differ in a meaningful way by creating contexts where participants differently value the attributes under consideration.

One concern that might arise in interpreting these findings is whether the stimulus selection in these preference elicitation procedures could partly be leading us to find these

effects. We attempted to conduct studies using a wide range of stimuli, both within and between different classes of goods as a means of obviating this concern. We do note, however, that our findings focus on divergence between expressed preferences in choices and WTP about relatively inexpensive, quotidian items. Because we find that deliberation attenuates (but does not eliminate) the difference between choice and WTP, we expect that in situations where consumers naturally deliberate more about the decision, the effect will be smaller (but not eliminated). For example, when consumers choose between two very expensive items (e.g., two possible houses to buy), they probably deliberate more and include more features in their evaluations. Therefore, we would expect an attenuation of the difference between choice and WTP if people were evaluating goods that were consequential, expensive, or highly meaningful.

An additional question related to stimulus selection is whether these effects would persist if researchers investigate preferences about products that are both hedonic or both utilitarian. Our effect relies on there being some degree of difference between the products under consideration but does not require one product to be a "typically" hedonic good and the other to be a "typically" utilitarian good (see studies 1i and 1j). We expect that, if participants expressed preferences between two different goods that are equally hedonic or equally utilitarian, choices and WTP align.

Different Explanations for Preference Reversals

While we (to our best knowledge) are the first to show preference reversals for consumer goods based on choices versus WTP, there is of course a large body of work on preference reversals in different contexts. On a surface level, our findings clearly relate to—and are consistent with—Lichtenstein and Slovic's (1971, 1973) work on elicitation effects for decisions under uncertainty. These classic findings show that participants who choose between lotteries with varying payouts and probabilities of winning prefer higher-probability but lower-payout lotteries, while those who indicate WTP prefer lower-probability lotteries with higher payouts. The process believed to underlie these preference reversals, scale compatibility—the notion that WTP participants anchor on the monetary payouts and overprice their WTP—cannot explain the effects documented in this article, however. First, none of the outcomes are on the same scale as WTP; that is, there are no monetary outcomes that can contaminate participants' responses in the WTP condition. Furthermore, if the differences could be explained by anchoring on a monetary value, deliberation should affect preferences in the WTP conditions, *not* in the choice conditions. Thus, the concept of scale compatibility cannot account for the observed preference reversals. Furthermore, studies 3a and 3b rule out the possibility that different response scales are contributing to the effects we observe. Another crucial difference between the studies presented in this article and the previous work on preference reversals is that earlier work focuses on a context where the outcomes are probabilistic. However, in the studies presented here, neither outcome is subject to any uncertainty.

We now turn again to the studies by Nowlis and Simonson (1997) to explain why our results appear to conflict with theirs. Specifically, we believe that because we allowed participants to express indifference, whereas Nowlis and Simonson used a forced choice in the choice conditions and dropped participants who expressed indifference in the rating conditions, the studies are not equivalent. By not allowing participants in the choice condition to express indifference, the study design created an asymmetry between the conditions, where one half can express indifference between unfamiliar products, while the other half is forced to make a

possibly arbitrary choice. Because of this, a participant that is not interested in buying a toaster would indicate a low rating in the rating condition for both toasters and be dropped from the sample, while in forced choice this participant would likely choose the cheaper of the two toasters. We believe this selection effect could explain the seemingly contradictory findings.

Implications for Researchers

The implications for researchers should be self-evident: the conclusions we draw about consumers' preferences are likely to vary greatly based on how these preferences are measured. There does not, however, appear to be a clear answer as to what the "correct" measure of preferences is. In fact, this decision is likely to depend largely on the question the researcher is trying to answer and the context of the behavior the researcher is trying to predict.

Importantly, we believe our effects can inform studies of preference, even when researchers are not looking to find evidence for preference reversals. Given that we have demonstrated reliable and predictable differences in expressed preferences between elicitation procedures, the choice of which procedure to use in a study is likely to lead to differences in preferences, independent of other manipulations. Consequently, the researcher's decision to use WTP or choice paradigms is likely to influence effect size outcomes and effect size estimates (see also Moon and Nelson 2017, who have documented a dissociation between WTP and anticipated enjoyment in decision-making under uncertainty). We think it worth reiterating the advice offered by Frederick and Loewenstein (2008), who find differences between choice and pricing for sequences of outcomes, on the importance of using multiple measurements to assess and describe general phenomena.

Discerning readers might have noticed that, in many studies, the utilitarian good is longer lasting, while the hedonic good offers a larger amount of immediate pleasure. In designing these studies, we initially believed this to be a potential confound. However, we have come to believe that a crucial, but not the only, determinant of what is considered hedonic or utilitarian is the timing of how a consumer experiences utility from each item. Hedonic goods seem to provide a large amount of pleasure immediately, while for utilitarian goods the utility is spread out over time. We conducted a brief study on MTurk. We recruited 99 participants and asked them to consider receiving 10 Hershey's chocolate bars both all at once and once a week for 10 weeks. We asked participants to rank how hedonic relative to utilitarian each of these situations would be on an 11-point scale, and found that participants rated receiving all 10 chocolate bars at once to be significantly more hedonic than receiving them once a week (Ms = 9.44 vs 7.77, paired $t(98) = 5.19$, $p < .001$. Given that choice and WTP seem to engender strong differences in preferences for hedonic and utilitarian goods, we believe the link between the consumption timeframe for a given product and its classification as a hedonic or utilitarian good is a topic that is relevant to the effect we demonstrate here and worthy of further exploration.

Implications for Applications in Industry

Finally, we turn briefly to industry applications and suggest that our findings may have implications for how products are sold and how consumers actually choose products. For instance, it is likely that "impulse" buys might represent something of a choice context, where consumers maximize on what is pleasing to them, with little consideration for other attributes. Likewise, purchasing a home or a car would be more similar to a WTP context even when

consumers are choosing between different options, where they consider various attributes in conjunction with their budgets.

With regard to a purchasing context, in some cases consumers clearly have to determine their willingness to pay for a given good (for instance, in purchasing a product on eBay), while in other cases they clearly make a choice (e.g., choosing between ice cream flavors after having decided to buy a scoop). Developing an understanding of how consumers approach consumption decisions in the context of the purchase is likely to help inform whether consumers approach a decision more like choice or WTP.

Conclusion

In closing, we would like researchers to take away the point that their choice of elicitation matters because different measures of preference activate different psychological contexts for consumers, which affects their preferences. Choices and WTP ought not to be used interchangeably, particularly in comparisons of consumer preferences across product categories that might differ on how hedonic or utilitarian they are perceived to be. We hope that this work increases understanding of the psychological processes underlying how consumers express their preferences, while also contributing to researchers' understandings of how to measure a concept that appears to be the most basic and fundamental in our discipline but is still not fully understood.

STUDY A1: MORE FULLY TESTING THE ROLE OF MARKET PRICE

To more fully account for the role of market price in affecting these WTP and choice preference reversals, we conducted a follow-up study that crossed salience of market price with whether preferences were measured by WTP or choice within the same study (we have found the difference is robust whether market price is salient or not between studies, but have not yet tested for the effects of market price within a single study). If market price were the driving force in these preference reversals, we would expect the magnitude of the difference between WTP and choice to decrease when market price is made salient, relative to when it is not, because consumers would be more likely to choose the more expensive good. However, if the role of market price is more limited, we would expect the magnitude of the difference in expressed preferences between WTP and choice to be similar whether prices are salient or not.

Study A1 Participants and Procedure

Participants were 815 workers (56.7% female M_{age} = 37.8, SD = 12.2) recruited through MTurk. Participants were randomly assigned to one of six conditions, in which we varied whether participants indicated their preference by WTP or choice and whether the market price for the goods was shown or not. We also included as a sub-factor within the WTP condition whether participants indicated their WTP values jointly (side-by-side on the same page) or separately (on two different pages). We randomized the order in which participants saw the products.

All participants evaluated the same products as in studies 1c-1d (toilet paper and ice cream). Participants in the choice condition were told, "Imagine you had a choice between the following two options" and were prompted to make a selection with, "I would prefer the:". Participants in the WTP conditions were told to look at the products and in the joint pricing condition were asked, "How much would you maximally be willing to pay for each of these products?" or in the separate pricing condition were asked, "How much would you maximally be willing to pay for this ice cream/these toilet paper rolls?" In the conditions in which market price was made salient, participants were told the MSRP for the ice cream was $4.99 and the MSRP for the toilet paper was $6.99.

Study A1 Results

First, we calculated preferences in the WTP conditions by comparing the indicated WTP for the ice cream with the indicated WTP for the toilet paper rolls. We classified participants as preferring the good for which they indicated the higher WTP. If participants indicated the same WTP for both goods, they were classified as indifferent. For parsimony, we excluded indifferent participants and were left with 769 responses.

Among participants who indicated their WTP separately and did not see the MSRP, 84.8% (84/99) preferred the toilet paper; among those who indicated their WTP separately and did see the price, 92.2% (83/90) preferred the toilet paper; among those who indicated their WTP jointly and did not see the price, 83.9% (78/93) preferred the toilet paper, while among those who indicated their WTP jointly and did see the price, 88.7% (86/97) preferred the toilet paper. An ANOVA reveals that there was no significant effect of joint or separate pricing $F(1, 375)$ =

.44, no significant effect of whether the price was salient or not, $F(1, 375) = 3.16$, and no significant interaction between these two factors $F(1, 375) = .14$, all ps > .05.

Among participants who made a choice between the two products, when the price was salient, 43.1% (84/195) chose the toilet paper, and when the price was not salient, 47.7% (93/195) chose the toilet paper. An ANOVA reveals there was no significant effect of whether price was salient or not, $F(1, 388) = .84$, p > .05. All means are graphed in Figure 3.

Collapsing across joint or separate pricing within the WTP condition, an ANOVA reveals that there is only an effect of preference elicitation method, $F(1, 765) = 187.79$, $p < .001$. There is no significant effect of market price salience $F(1, 765) = 3.00$, $p = .08$ and no interaction between elicitation method and price salience $F(1, 765) = .05$, $p = .82$.

Figure 2. Preference for utilitarian good by condition, Study A1.
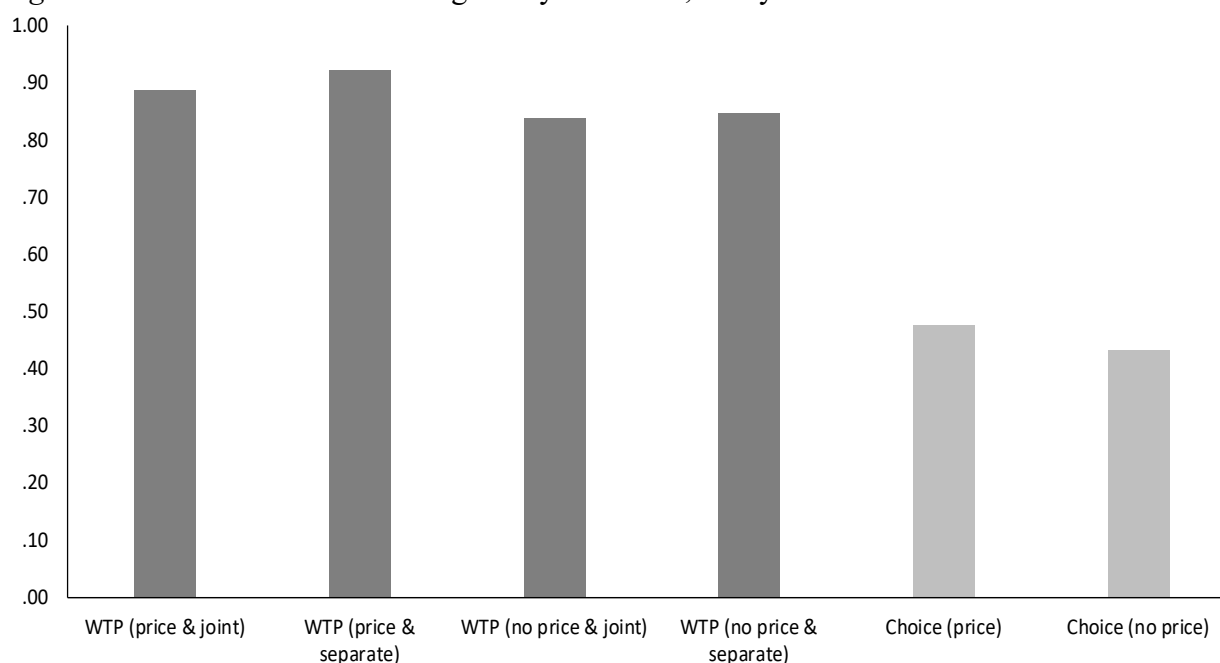
Study A1 Discussion

Study A1 bolsters our theory that WTP and choice preference reversals are not due to the salience of the market price. Participants in the WTP condition preferred the utilitarian good at a similar rate regardless of whether they indicated their WTP values jointly or separately and regardless of whether the market price for each item was salient or not. Likewise, participants in the choice condition chose the utilitarian good at the same rate regardless of whether MSRP was salient or not. If participants were merely responding to market value, one would expect preference for the (more expensive) toilet paper to increase when its higher market value was made salient, and this did not occur. In any case, participants were vastly more likely to select the utilitarian good in a choice context than in a WTP context. This is consistent with our theory that different preference measures create different contexts that allow for expressed preferences to vary between them.

Chapter 2: Variety Seeking Behavior in Bundle Construction: Choice Myopia and Combinatorics

Even when consumers have favorites, they may still want to mix things up. Taco Tuesday is the highlight of some consumers' week, but few of them eat tacos for every meal. With repetition, consumers experience diminished marginal utility; this is why people tend to find less satisfaction in the twenty-first Swedish Fish than the first (e.g., Redden 2007). Anticipating this, consumers add variety to what they consume.

Consumption and choice are frequently separated in time (Read and Loewenstein, 1995). For example, grocery store purchases are essentially choices for eating later in the week. Two psychological mechanisms can lead consumers to prefer more variety at the time of choice than what they would ultimately prefer at the time of consumption. One mechanism rests on the phenomenon that two points in the future seem subjectively closer to each other today than they ultimately will feel. Due to this *time contraction*, people overestimate how long it will be before they will be ready to consume their most preferred option again (Simonson, 1990). Taco Tuesday fans underestimate how ready they will be for Taco Thursday. Second, when choosing multiple options at a single point in time, people fail to appreciate that although they are considering all of these selections at a single time when choosing, they will actually consume them in distinct, individual episodes. This is a demonstration of broad *choice bracketing* (Read, Loewenstein, and Rabin 1999): When all the choices are considered together, the same psychology that discourages people from consuming too much of the same item at once may encourage people to seek more variety for consumption down the road. But given that one's entire grocery cart will not be consumed in one sitting; broad choice bracketing may encourage more diversity of choice than is ultimately desired at the time of consumption.

Although choice and consumption are often temporally separated, we credit Mittelman, Andrade, Chattopadhyay, and Brendl (2014) for noting how even the choice itself can occur all at once (as when a consumer selects a preassembled six-pack of beer) or sequentially and thus with small temporal separations (as when a consumer builds that six-pack herself). Mittelman and colleagues reasonably argue that neither time contraction nor choice bracketing can explain why consumers may differ in their preference for variety in these two contexts. Across four experiments Mittelman et al (2014) provided evidence that constructing bundles through multiple, sequential choices (like the shopper building her own six-pack) prompts more of a selection of variety than single choices of a prepackaged bundle (like the shopper choosing among preassembled six-packs). Although their work was intriguing, the authors provided no evidence of why this *offer framing effect* occurs.

Choice Myopia

Numerous lines of research depict humans as remarkably short-sighted decision-makers. For example, people seem to have surprisingly high discount rates, which imply that they greatly undervalue future states relative to immediate outcomes (Mischel, Shoda, & Rodriguez, 1989; Loewenstein & Thaler, 1989; for a review see Urminsky & Zauberman, 2014). Milkman, Rogers, and Bazerman (2008) discussed how decision-makers are likely to prioritize *want* options over *should* options when making short-term decisions. Seeking to explain the prevalence of gambling, impulsivity, and risky sexual behaviors, Loewenstein (1996) noted that visceral factors aroused in these contexts (e.g., sexual desire) encourage people to satisfy

immediate goals at the expense of future goals. Furthermore, Laibson (1997) has explained that the increased liquidity of financial assets has allowed consumers' truly short-sighted preferences to emerge, which has led to a rise in immediate spending and a reduction in the savings rate. Taken together, these findings depict human decision makers as myopic—considering what the immediate, and not the long-term, effects of their choices will be.

This tendency to focus on what is immediately in front of the self and not to give full thought to how the present relates to a fuller, longer-term picture can be observed over much short time scales. Consider a function as basic as linguistic sentence processing. The horse raced past the barn fell. Many readers will need to read that sentence more than once. This is because the mind does not wait until completing the sentence to assign grammatical roles to each word. Instead, most first decide that "raced" is the verb long before the word "fell" is encountered. Psycholinguists would say the reader got garden-pathed: He preemptively committed to a single interpretation of the sentence before learning that this led to a dead-end (Bever, 1970; Ferreira & Henderson, 1991).

Consider again the construction of bundles sequentially (choosing each item one by one) or simultaneously (choosing among prepackaged bundles). Simultaneous choosers are directly selecting the end-product they would like. In contrast, the choices made by sequential choosers constrain the set of final bundles they may form. But if sequential choosers are myopic, then they may make choices in the present that keep them from achieving bundles that—were they more forward-looking in their decision approach—they would prefer. For example, a dozen roses are a "bouquet of roses," whereas eleven roses and one tulip seem more like a mistake. But the chooser who is indifferent between roses and tulips can garden paths himself from building a bouquet of either the moment he adds the second type of flower to his bundle. Of course, the eager listener had no choice but to wait until the end of the ambiguous sentence or guess (perhaps incorrectly) prematurely. But those constructing bundles myopically do not have the same excuse; they may simply be shortsightedly misled by failing to think through the full implications of their decisions. How should we expect such extreme choice myopia to influence consumers' bundle construction?

Pathways to Variety

In the casino dice game craps, players take turns shooting (throwing) two dice. If a player throws a 1 and a 3, this is called an "Easy 4." But when a player throws a 2 and a 2, this is called a "Hard 4." It is quite literally easier (twice as easy, in fact) to achieve a 4 by throwing two different numbers than it is to throw the same number twice. This is because there are multiple pathways to variety (the first throw can be either a 1 or a 3), but only a single way by which a no-variety 4 can be achieved.

Building a bundle sequentially is analogous to throwing dice one at a time. If one wishes to build a bouquet of two tulips and two roses, there are six different ways to achieve that outcome. In the language of combinatorics, there are 6 unique permutations of 2 tulips and 2 roses: $4! / (2! * 2!)$. But achieving a bouquet with no variety—much like achieving a hard 4—can occur in fewer ways. One must either repeatedly select tulips or repeatedly select roses. This means that when choosing sequentially, there are many distinct pathways that produce a high-variety bundle, but many fewer that produce an unvaried bundle. As craps jargon anticipates, sequential choice quite literally makes it easier to achieve variety.

In contrast, when selecting among prepackaged bundles, one typically observes unique combinations, not all of the possible permutations that could define them. That is, a florist might have on display the five different combinations of tulips and roses that could compose a set of four such flowers. But it's extremely unlikely that the florist would show the 256 unique permutations ($4 \wedge 4$) that could emerge. This means variety is relatively overrepresented among the permutations that choosers of prepackaged bundles see. And because sequential choosers are likely to be myopic—not predetermining the bundle they would like and then making the choices to get them there—we begin to see why sequential and simultaneous choosers may differ in their choices of variety.

The fact that sequential choice offers more pathways to variety may be sufficient to explain why consumers will choose more variety when constructing bundles in this way. More generally, previous work has found that as the composition of a response set changes—even in arbitrary ways—judgments and decisions shift to spread across them (Benartzi and Thaler 2001; Fox, Ratner and Lieb 2005). In a classic example, survey respondents reported vastly slimmer TV-watching habits when selecting their daily viewing patterns from among shorter options (under 1/2 hour, ½ to 1 hour, 1 to 1.5 hours, 1.5 to 2 hours, 2 to 2.5 hours, or over 2.5 hours) instead of longer options (under 2.5 hours, 2.5 to 3 hours, 3 to 3.5 hours, 3.5 to 4 hours, 4 to 4.5 hours, or over 4.5 hours; Schwarz, Hippler, Deutsch, and Strack, 1985). Following a shift in the partitions along the response continuum, judgments spread across those (artificial) categories.

Some evidence extends this phenomenon to choice. Benartzi and Thaler (2001) identified how it applied to investment decisions. As retirement providers offered different sets of funds that skewed toward riskier vs. safer options, they found investors continued to spread their money relatively evenly across the available choices. This meant that investors seemed to follow something of a 1/n rule, placing 1/n of their money in each of n funds. Such decisions did not reflect stable attitudes about risk, but instead a tendency for choices to conform to the distributions of available choices. Of course, by changing the available funds, the actual set of constructible portfolios shifted as well. We extend beyond these phenomena by arguing that (sequential) choosers will show more variety in their choices merely because they have more pathways to reach such outcomes, even though the actual final bundles they and simultaneous choosers are selecting among are identical.

Does this solve the mystery of why the offer framing effect emerges?

Mittelman et al. (2014) offer four studies in support of the idea that people prefer more variety when constructing bundles sequentially than when selecting among prepackaged bundles.

Let us consider how the present ideas could explain their Study 3. In that study, participants selected a three-flower bouquet that could be composed of some combination of red and orange flowers. When participants constructed the bundles sequentially, 70% of them selected a varied bundle (two flowers of one color, one of another). When participants selected among prepackaged bundles, a slim majority (52.7%) actually selected flowers of only one color. If choosers were not myopic—i.e., if sequential choosers were simply selected among which final bundle they wanted—it would be difficult to imagine why the decision structure would matter. But also, because more pathways in sequential choice produce variety (6 of 8; 75%) than in simultaneous choice (2 of 4; 50%), our account correctly anticipates sequential choosers' interest in variety.

Readers with more knowledge of Mittelman et al.'s (2014) other three studies would be right to be skeptical of our account. In fact, combining choice myopia with pathways to variety cannot anticipate the remaining three findings in that paper. Of course, Mittelman et al.'s effects could be multiply determined, though this makes our above arguments suspect on grounds of parsimony. That said, we identify an artifact that we suspect may account for two of Mittelman et al.'s studies. We test whether that is the case in Studies 1a and 1b. For the third study, we considered the possibility that the original findings—which we had trouble explaining—might not be robust. We conducted several direct replications of it to determine whether the findings should guide the field's thinking on the topic.

In this way, the present paper follows something of an unusual approach. We are motivated by a previous paper whose effect was documented but incompletely explained. We offer that explanation. But through gaining this mechanistic understanding, we come full circle by showing that the originally documented effect was mis-specified. We hope this cycle can offer one guide by which scientific findings can inspire follow-up work, which ultimately leads to revisions to the motivating work.

STUDIES 1A AND 1B: A MISUNDERSTOOD MEASURE?

In Mittelman et al. (2014, Experiment 1) participants had to select a two-soda bundle of Coke and/or Sprite. As predicted, those who constructed these bundles sequentially were more likely to choose a varied bundle (one Coke, one Sprite) compared to those who selected among prepackaged bundles. But notably, choice myopia combined with pathways to variety cannot anticipate this result. For participants who selected among prepackaged bundles, the authors were careful to show *both* permutations of high-variety bundle: one Coke, one Sprite; one Sprite, one Coke. In other words, there were just as many pathways to variety for sequential choosers (2 of 4; 50%) as simultaneous choosers (2 of 4; 50%).

Instead, we considered whether participants in the sequential choice condition may have been confused about what the experimenters were asking. In that condition, participants specified their "first choice" and their "second choice." Although participants were permitted to select either soft drinks at each decision point, it seemed possible that some fraction of them may have misinterpreted the question as asking for their favored soda (first choice) and their disfavored one (second choice). But even if there was some confusion, could it have been enough to explain the impressively large effect the authors observed? In their study, while only 34% of simultaneous choosers selected a bundle of one Coke and one Sprite, 62% constructed this variety bundle when making two separate choices. Mittelman et al.'s Experiment 2 was conceptually identical but involved creating two-item bundles of Snickers and Twix.

In Studies 1a and 1b, we directly replicate Mittelman et al.'s (2014) Experiments 1 and 2. That is, participants constructed bundles of soda (Study 1a) or candy (Study 1b) through two, sequential choices or by directly selecting among prepackaged bundles. Half of participants saw Mittelman et al.'s original measures. Half saw measures that clarified the potentially confusing wording by making it explicit that participants could choose either two of the same item or two different items. This would allow us to test whether some or even all of the originally documented effect was due to this measurement artifact.

Participants and Procedure

We *a priori* decided to collect data from 200 participants per condition. This compares to two hundred and eighty-nine participants in total in Experiment 1 and one hundred eleven in total in Experiment 2 of Mittelman et al. (2014). Eight hundred twelve participants were recruited from Amazon's Mechanical Turk (MTurk) for Study 1a. Eight hundred eight, for Study 1b. Participants were required to be Americans with an approval rating greater than or equal to 95% for all MTurk Human Intelligence Tasks (HITs) completed prior to ours. In both studies, participants were randomly assigned to one of four conditions in a 2 (Instructions: original or clarified) ✕ 2 (Bundle: prepackaged or constructed) full-factorial design.

All participants were asked to consider going to a convenience store (Study 1a) or a supermarket (Study 1b) to buy two cans of soda (Study 1a) or two candy bars (Study 1b). In Study 1a, participants were told the convenience store sells only Coke and Sprite. In Study 1b, participants were told the supermarket sells only Snickers and Twix.

Original. Those in the original instructions condition saw prompts that were copied from Mittelman et al. (2014). In the prepackaged bundle condition, participants selected among four prepackaged bundles: Coke-Coke, Coke-Sprite, Sprite-Coke, Sprite-Sprite (Study 1a) or Snickers-Snickers, Snickers-Twix, Twix-Snickers, Twix-Twix (Study 1b). Each bundle was represented by a unique image of two soda cans or two candy bars side-by-side. Those in the constructed bundle responded to the prompt "My first choice would be," before also responding to "My second choice would be." Following each prompt were the two unique soda cans (Study 1a) or the two unique candy bars (Study 1b).

Clarified. Those in the clarified instructions condition first received additional information designed to clarify that they were free to choose "any combination" of drinks (Study 1a) or candy bars (Study 1b). We clarified this meant "two different drinks [candy bars] or two of the same drink [candy bar]." Just to make sure this was perfectly clear, we added, "Feel free to choose two of the same soda [candy bar] or, if you like, two different sodas [candy bars]." In this way, it should have been clear to everyone that participants could choose two of the same sodas or one of each. Keep in mind that our worry about ambiguity in the wording applied only to the constructed bundle condition, not the prepackaged bundle condition. That said, if the clarification (unintentionally) pushed choosers toward the low-variety or high-variety bundle, then this should have been the same for both clarified instructions conditions. Instead of asking sequential choosers for their "first choice" and "second choice," we instead said, "Please choose one of your two beverages…" and "Please choose the second of your two beverages…"

Results and Discussion

We began by coding all selections for whether they reflected a varied bundle (one soda or one candy bar of each type) or not. Using Mittelman et al.'s (2014) original instructions, we replicated their effects. In Study 1a, 66.17% of participants (133 / 201) constructed a varied bundle, but only 41.06% (85 / 207) selected a prepackaged bundle of both Coke and Sprite, χ^2 (1, N = 408) = 24.80, $p < .001$. In Study 1b, 67.49% of participants (137 / 203) constructed a bundle with both Snickers and Twix, but only 49.75% (100 / 201) choose a prepackaged bundle of that composition, χ^2(1, N = 404) = 13.10, p < .001. These numbers are similar to Mittelman et al. (2014), wherein 62.1% of participants selected variety in Experiment 1 (which corresponds to

Study 1a) and 73.6% of participants selected variety in Experiment 2 (which corresponds to Study 1b). Given participants in both conditions had the same number of pathways to variety, these findings cannot be explained by simple combinatorics.

Nevertheless, neither difference persisted with the clarified phrasing. With this clarification in Study 1a, only 33.83% of those constructing the bundle sequentially (68 / 201) chose two different sodas, a proportion quite similar to those who chose among prepackaged bundles: 37.43% (76 / 203). And in Study 1b, 53.69% of participants (109 / 203) constructed a bundle of two different candy bars, which was no greater than the 52.74% (106 / 201) who selected a prepackaged bundle of both Snickers and Twix. In both studies, addressing the measurement concern eliminated the connection between how the bundle was selected and a preference for variety, $\chi^2 < 1$.

Given this pattern of results, we tested for the interaction between bundle condition and instructions. We found that the Instructions X Bundle interaction was significant in both Study 1a, $z = 4.00$, $p < .001$ (see Figure 1a), and Study 1b, $z = 2.72$, $p = .007$ (see Figure 1b). The clarified instructions completely eliminated the difference in apparent preference for variety between the simultaneous and sequential choice conditions.

Having found that both of Mittelman et al.'s (2014) first two experiments stemmed from measurement confusion, not a greater preference for more variety when constructing bundles sequentially, it is actually easier to interpret other unusual features of their results. In their Experiment 1, participants who had a strong preference for one soda over the other chose both bundles only when constructing the bundle sequentially (i.e., when they were likely confused about what they were asked to do). In their Experiment 2, coders read participants' thought-listing protocols in which they described their thoughts when choosing their bundles. Coders could relatively easily identify who had selected two of the same candy bars. But coders frequently misidentified who was selecting two different candy bars, but only when they constructed those bundles sequentially. Coders presumably had trouble because many of these participants actually had no interest in selecting two different candy bars.

Mittelman et al. (2014) interpreted these (then-)cryptic findings as demonstrating the power of the offer framing effect: It pushed people to sequentially construct high-variety bundles even when their true preferences seemed incompatible. But in light of our findings, these findings are now more easily understandable. Many sequential choosers who sought variety did not realize that is what they were doing. They believed they were indicating their favored and their disfavored item, explaining why their choices did not follow from their preferences. Mittelman et al. (2014) report that one participant who selected a Snickers and a Twix bar stated, "I don't care for Snickers." Presumably this participant did not realize what he or she was selecting when indicating Snickers as the "second choice" candy bar.

Figure 1a. Proportion choosing variety (i.e., different soft drinks) in Mittelman et al. (2014) Experiment 1 and the present manuscript's Study 1a by condition.
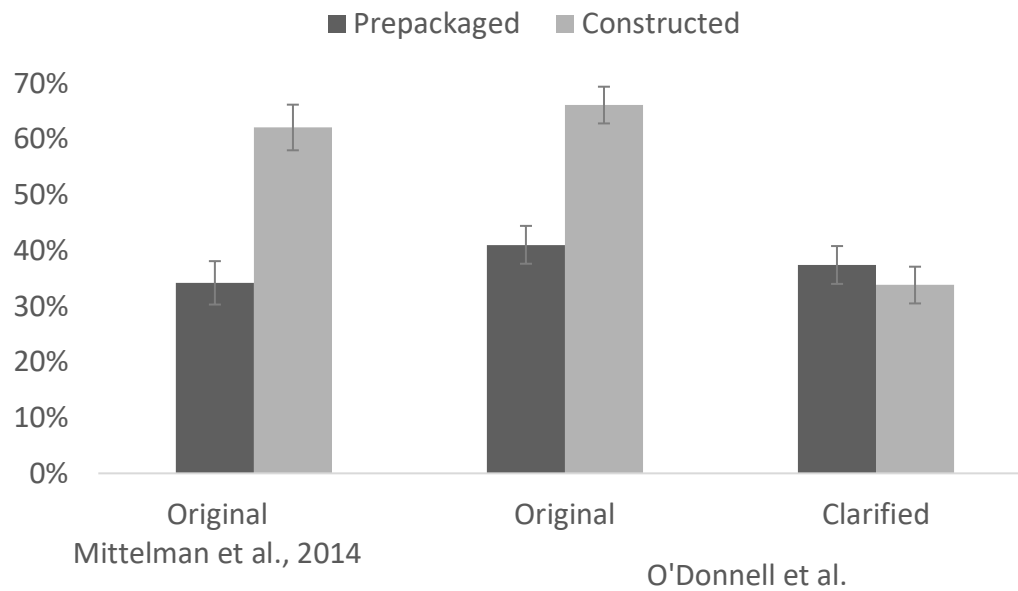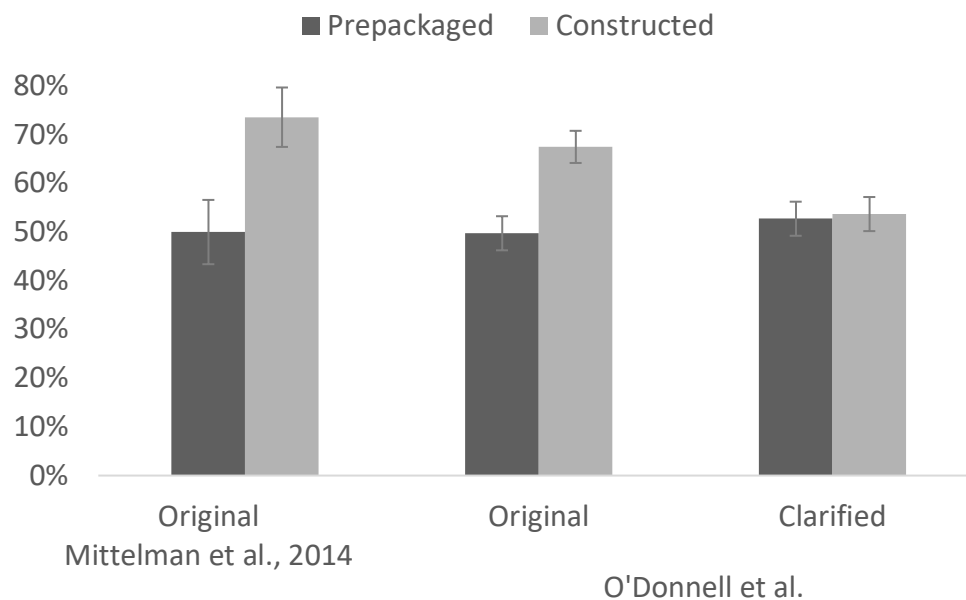

Figure 1b. Proportion choosing variety (i.e., different candy bars) in Mittelman et al. (2014) Experiment 2 and the present manuscript's Study 1b by condition.

STUDY 2: EXPLORING PATHWAYS TO VARIETY

**Study 2a**: **Expanding Pathways to Variety**

Studies 1a and 1b demonstrated that the previously published understanding of how the bundle selection process changes a preference for variety is wrong, but it is not yet time to be fatalistic. After all, the theoretical account we advance anticipated those effects. In Study 2, we move to a more direct test of our model. We move to a context—one tested in Mittelman et al. (2014)—in which there actually are more pathways to variety when constructing bundles sequentially instead of choosing among prepackaged bundles. We then make a minimal change that eliminates this difference in order to see if it eliminates the different preference for variety as well.

In Mittelman et al. (2014, Experiment 3) participants constructed a bouquet of flowers from a large supply of yellow and orange roses. For those who selected among prepackaged bouquets, they saw four such bundles (OOO, OOY, OYY, YYY), 50% of which contained variety. But for those who constructed the bundles sequentially, they confronted 8 unique pathways, 75% of which led to variety. Only 2 (25%) of the pathways led to an unvaried bundle: OOO, YYY.

Note that this difference is a property of the choice architecture, not a difference in the psychological orientation of the choosers themselves. But when combined with the property of choice myopia—suggesting those constructing a bundle sequentially are not far-sighted enough to merely preemptively select which bundle they are building toward—this can lead to different apparent preferences for variety. More generally, when there is less redundancy (i.e., more variety) in a bundle, there will be more unique pathways by which that bundle could be constructed. That said, this structural confound can be corrected for experimentally.

We accomplished this by adding a third condition in which we offered a prepackaged bundle for each unique permutation, not merely each unique combination. For example, instead of merely offering a prepackaged bouquet of two yellows and one orange (as in the standard prepackaged condition), we offered YYO, YOY, and OYY. In this *expanded prepackaged* condition, we retained the crucial choice process feature of the prepackaged condition (i.e., selecting among prepackaged bundles), but increased the pathways to variety (from 50% to 75%) to match the constructed bundle condition. If our pathways to variety account is accurate, we should find: 1) those in the constructed bundle condition select more varied bundles than those in the prepackaged condition, and 2) those in the expanded prepackaged condition select *less* variety than those in the (standard) prepackaged condition. See Figure 2 for a visualization of the pathways to variety available to participants in each condition.

Figure 2. An illustration of how Study 2a's constructed bundle participants (those who made three distinct choices) faced more unique pathways to variety than did prepackaged bundle participants (those who selected directly from the four boxed bundles), but the same number as expanded prepackaged participants (those who selected directly from all eight bundles).
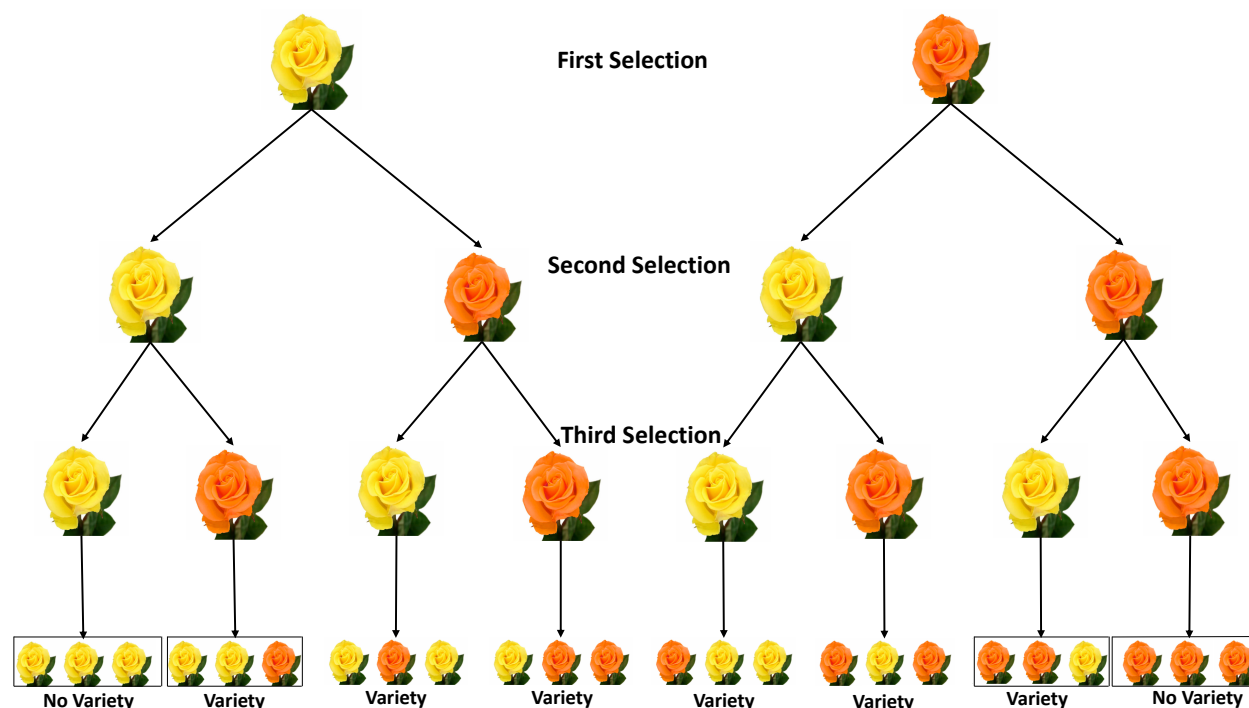
Participants and Procedure. We decided *a priori* to collect data from 200 participants per condition. This compares to one hundred seventy-nine participants in total in Experiment 3 of Mittelman et al. (2014). Eight hundred six Americans were recruited from Amazon Mechanical Turk. All were required to have an approval rating of at least 95%. Each participant was randomly assigned to one of four conditions. Three of these were crucial for testing our ideas: constructed, prepackaged, or expanded prepackaged. Merely in an effort to include all three conditions that Mittelman et al. (2014, Experiment 3) did, we also included a fourth unbalanced constructed condition. We give brief discussion to it below.

All participants were asked to consider "shopping for flowers." Their task would be "to select a bouquet of roses." Those in the constructed bundle condition constructed their bundle sequentially. They saw an array of 6 yellow and 6 orange roses, the supply from which they would construct their bouquets. Participants had to select three roses from this set of twelve in order to construct a bundle. Note that because there were only two colors of roses, and because we did not ask for a "second choice" or "third choice" rose, there is not the same possibility of misinterpretation that Studies 1a and 1b addressed. Participants in the unbalanced constructed condition confronted a nearly equivalent choice. But the array from which they chose had 8 yellow and 5 orange roses. Our account does not anticipate that this imbalance in supply should matter.

In contrast, those in the (original) prepackaged bundle condition selected from the four unique combinations of roses. However, those in the (new) expanded prepackaged condition

selected directly from every unique permutation of roses. This meant that when choosing a bouquet with two roses of one color and one rose of another, participants in the prepackaged condition could select the lone rose to be in the left, middle, or right position. Note that this more closely parallels the experience of those constructing a bundle: They could select the one different-colored flower first, second, or third.

Results and Discussion. We began by coding participants' chosen or constructed bundles for whether they contained variety—i.e., two roses of one color and one of the other. Participants' likelihood of selecting a varied bundle strongly differed across the four conditions, χ^2 (3, N = 806) = 44.60, $p < .001$, see Figure 3. We proceeded to test targeted pairwise comparisons that would permit us to test the role of pathways to variety in partially or fully explaining the preference for variety when constructing bundles sequentially instead of selecting among prepackaged bundles.

First, we found that those who constructed the bouquet themselves were more likely to put together a varied bouquet (74.38%; 151 / 203) than those who chose among the standard set of prepackaged bouquets (44.55%; 90 / 202), χ^2 (1, N = 806) = 37.39, $p < .001$. This directly replicates the pattern observed by Mittelman et al. (2014, Experiment 3), but still leaves us with no evidence as to why this pattern emerges. Did the difference in pathways to variety play a role? We compared the expanded prepackaged condition to the (original) prepackaged condition. And indeed, those in the expanded prepackaged condition were more likely to select a varied bundle (59.59%; 118 / 198) than those who considered the standard set of prepackaged bouquets, χ^2 (1, N = 400) = 9.06, $p = .003$. In other words, those considering among prepackaged bundles displayed a lower preference for variety in part because the choice context offered them fewer pathways to variety. Note that those who constructed the bundle sequentially did show more of a preference for variety than those who considered the expanded set of prepackaged bundles, χ^2 (1, N = 401) = 9.93, $p = .002$. This could suggest that choice myopia combined with pathways to variety only *partially* explains the observed effect. We return to this question in later studies.

Before continuing, let us also consider the unbalanced constructed condition—those who drew from a supply of flowers that skewed yellow. These participants selected just as much variety (69.46%; 141 / 203) as those who constructed bundles from the color-balanced supply, χ^2 (1, N = 406) = 1.22, $p = .27$. Mittelman et al. (2014) mentioned including this condition as a way to explore an idea that has superficial similarities to our own pathways to variety account. That is, they noted that if participants selected their roses randomly from the total available supply, then participants would create more varied bouquets. The authors hoped to rule out this account by showing that by skewing the supply toward yellow roses, that this would not increase the selection of yellow roses. Just like Mittelman et al. (2014), we also found no significant increase ($M = 1.33$, $SD = 0.91$) compared to when participants drew from the unskewed supply ($M = 1.21$, $SD = 0.84$), $t(404) = 1.36$, $p = .17$. But note that this test fails to consider how to determine whether pathways to variety plays a role in the offer framing effect. The claim is not that randomness influences people's selections from the source array—the bed of orange and yellow flowers from which one chooses. Instead, the claim is that there is some haphazardness in how people make each local choice of either an orange or a yellow flower (irrespective of how many such identical flowers exist in the source array). The present study indicates that the greater number of pathways to variety does play a role.

Figure 3. Proportion choosing variety (i.e., a multi-colored bouquet) in Study 2a by condition (Study 2a).
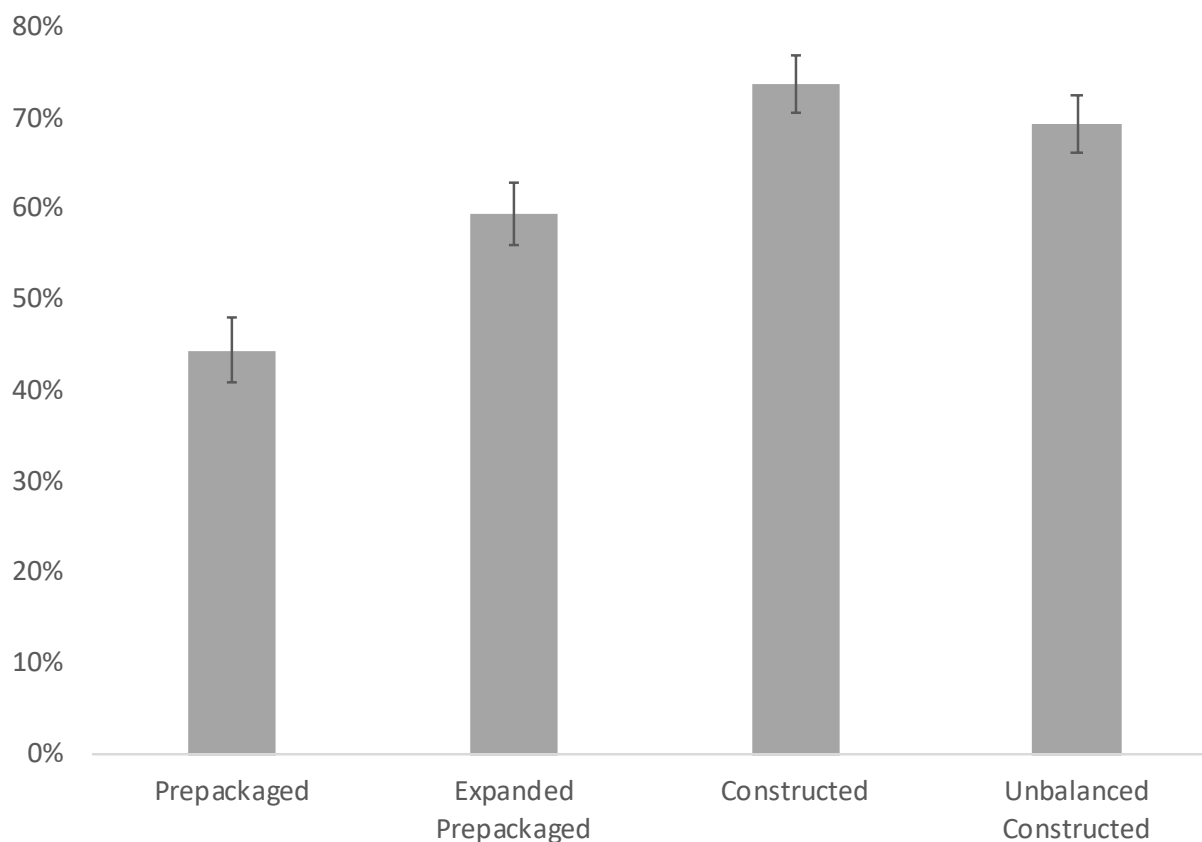

Study 2b: Restricting Pathways to Variety

Whereas Study 2a expanded the number of pathways to variety in the prepackaged bundle condition to match the (sequentially) constructed bundle condition, Study 2b takes the reverse approach to offer a convergent test of our theoretical account. But Study 2b adds a new condition in which we retain the key choice process feature of the constructed bundle condition (i.e., putting together the bundle sequentially) but eliminates the key element that we hypothesize promotes variety-seeking (i.e., the increased number of pathways to variety). This means that even myopic choosers should no longer be led to more varied bundles.

We accomplish this reduction by exploiting a simple fact about the two-color, three-flower bouquet paradigm: All bouquets must have at least two flowers of the same color. Thus, in a new *restricted constructed* bundle condition, we first had participants select two roses of the same color. In this way, when participants were selecting their third and final rose, they were making a choice that would create a high-variety bundle (by selecting the opposite color) or a low-variety bundle (by selecting the same color again). In other words, just as in the prepackaged bundle condition, only 50% of pathways led to variety (instead of 75% as in the original constructed bundle condition). See Figure 4 for a visualization of the pathways to variety available to participants in the restricted constructed bundle condition.

Finally, we once again included our expanded prepackaged bundle condition—our key addition in Study 2a. In that study, we found that expanding the pathways to variety for those choosing among prepackaged bundles increased their choice of variety, but not so much that it matched sequential choosers' choice of variety. Because in Study 2b we had access to a much larger sample, we could determine whether that difference was robust. Stated differently, we could: 1) attempt to replicate Study 2a's finding and 2) have a clearer understanding of whether pathways to variety only partially or instead fully accounts for the present effects.

Figure 4. An illustration of how Study 2b's restricted constructed bundle participants (those who made two distinct choices) faced an identical number of pathways to variety as those in the prepackaged bundle conditions.
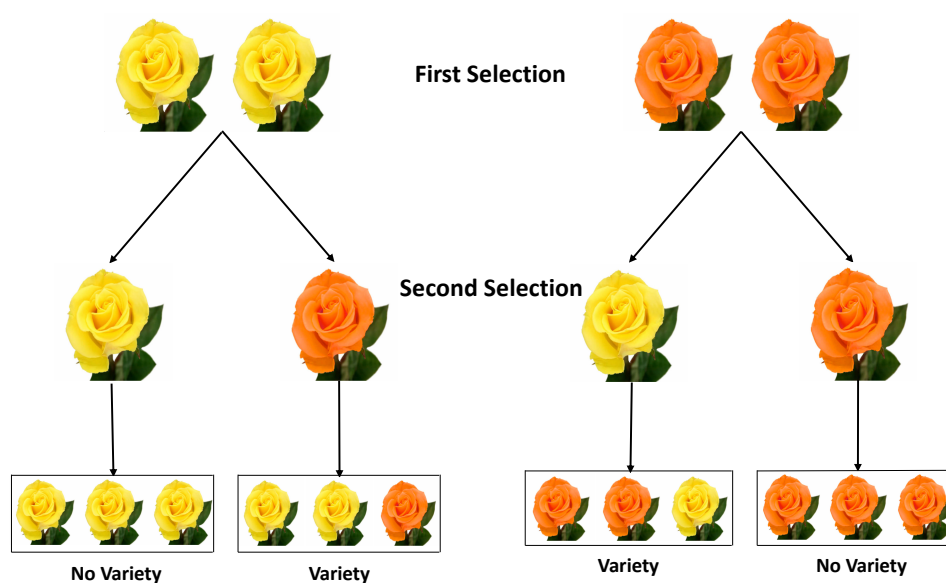
Participants and Procedure. One thousand eight hundred seventy-one participants were recruited to complete the study online by Luth Research. Luth collects large online samples that are demographically representative of the United States adult population. Participants completed this study as one of several unrelated studies that composed a 20-minute research session. Each participant was randomly assigned to one of four conditions. Three of those bundle conditions were identical to three of the conditions used in Study 2a: constructed, prepackaged, and expanded prepackaged. We added a restricted constructed condition.

Participants considered a similar choice scenario to the one used in Study 2a. How participants selected their three-flower bouquets of orange and/or yellow roses varied by condition. As before, those in the constructed bundle condition added flowers one-by-one by drawing from a supply of yellow and orange roses. Those in the prepackaged bundle condition selected among the four unique combinations of yellow and/or orange roses that can form three-flower bouquets. And those in the expanded prepackaged bundle condition saw all eight possible permutations of bouquets.

In the restricted constructed bundle condition, participants made two choices. First, participants were told, "We will now have you choose your first two roses. What would you like your first two roses to be?" They could select two orange roses or two yellow roses. Crucially, because every final bouquet has at least two roses of one color, this restricted choice does not foreclose the ability to reach any of the possible bouquets. These two roses moved to a box that said, "Your bouquet so far." Participants then selected either an orange or a yellow rose to complete their bouquet. Participants were able to see what the final bouquet would look like (in order to match the experience of participants in the other condition) before confirming their final selection. If they were not pleased, they could return to the first step. In this way, restricted constructed bundle condition had two (out of four) pathways to variety (50%), the same as those in the prepackaged bundle condition.

Results and Discussion. As before, we began by classifying participants' selection in terms of whether they selected a high-variety bouquet (two of one color, one of the other) or a low-variety bouquet (all the same color). A chi-square analysis revealed that participants' preferences for variety significantly differed among the four conditions, $\chi2$ (3, N = 1,871) = 98.56, $p < .001$, see Figure 5. We began by testing pairwise comparisons that would allow us to potentially replicate Study 2a's results. First, those in the standard constructed bundle condition were more likely to choose a varied bundle (66.26%; 324 / 489) than those who selected among the four prepackaged bouquets (47.45%; 242 / 510), $\chi^2(1, N = 999) = 35.96$, $p < .001$. Second, we found that by expanding the number of prepackaged bundles that displayed variety, we increased choice of variety (64.02%; 274 / 428), $\chi^2(1, N = 938) = 25.81$, $p < .001$. With a much larger sample than that used in Study 2a (meaning we had even greater power to test for real differences), we found no difference between the constructed bundle and expanded prepackaged bundle conditions, $\chi^2(1, N = 917) < 1$. This lends support to the idea that our proposed theoretical account (choice myopia combined with different pathways to variety) *fully* accounts for the observed effect of the choice process on preference for variety. But before embracing this conclusion too strongly, it is useful to move to the new test that Study 2b offers.

With the pathways to variety restricted, sequential choosers (in the restricted constructed bundle condition) were relatively unlikely to choose a varied bundle (38.51%; 171 / 444). That is, this restriction prompted a diminished interest in variety compared to the standard constructed bundle condition, $\chi^2(1\ N = 933) = 71.92$, $p < .001$. But did pathways to variety fully or only partially explain our basic effect? In actuality, restricting sequential choosers' pathways to variety led to even *less* of a preference for variety than those in the (standard) prepackaged bundle condition, $\chi^2(1, N = 954), = 7.72$, $p = .005$. This offers an independent confirmation that different pathways to variety explained the entirety of the effect of the bundle choice process on a selection of variety.

In summary, participants' interest in variety was predicted by the number of pathways to variety they could encounter, not whether they constructed their bundle sequentially or chose among prepackaged bundles. When people have more opportunities to choose variety, their myopic decision-making leads them to select bundles that contain more variety without consideration for the outcome. That is, in the two conditions in which 75% of participants' choice pathways led to variety (constructed; expanded prepackaged), their selection of variety was high. In the two conditions in which only 50% of participants' choice pathways led to variety (prepackaged; restricted constructed), their selection of variety was much lower. Furthermore, this appeared to account for the entirety of the effect, even in a very well-powered

study that had the potential to detect even small, lingering effects.

Figure 5. Proportion choosing variety (i.e., a multi-colored bouquet) in Study 2b by condition (Study 2b).
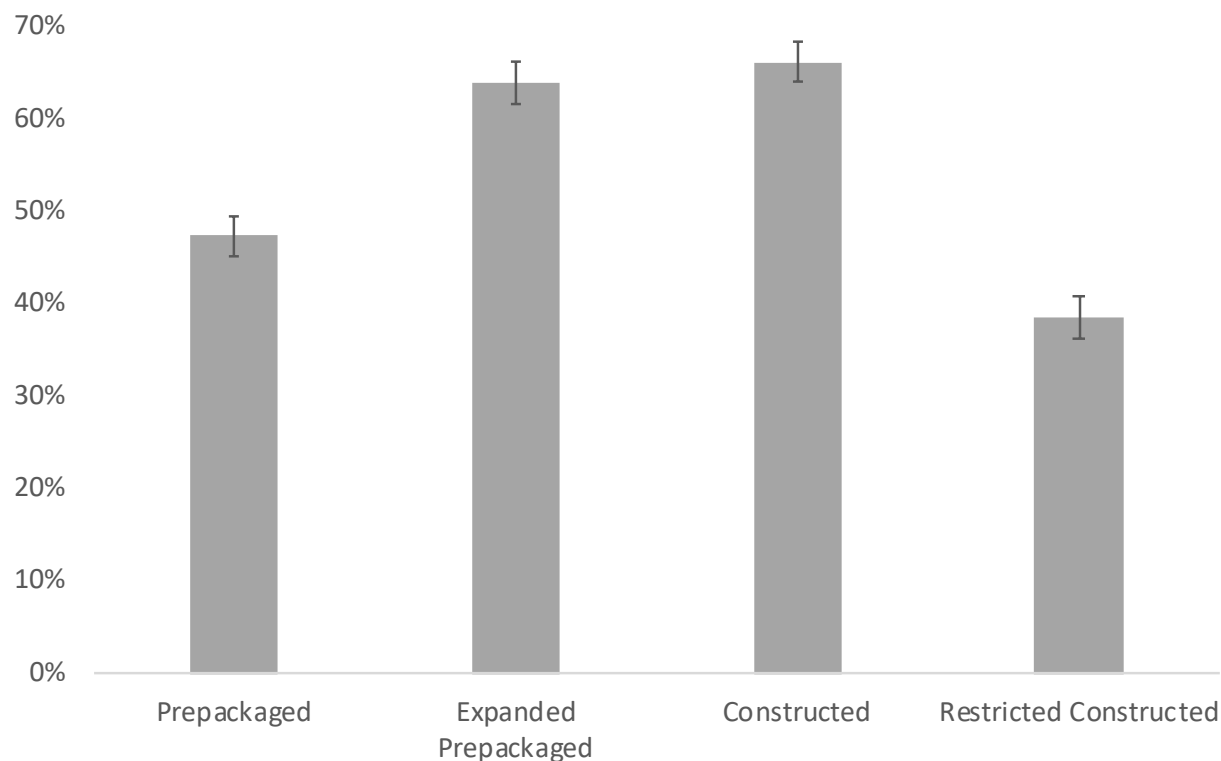
STUDY 3: DO PATHWAYS TO VARIETY AFFECT CONSEQUENTIAL CHOICES?

The results of Studies 2a and 2b suggest that people's selection of bundles is in part determined by the number of options made available to them, and that more pathways to variety lead to more variety. However, one limitation of these studies is that the participants' choices are entirely inconsequential – our respondents are selecting pictures of flowers that they will never actually use. Although we believe those findings do illuminate the important role of pathways to variety in the decision-making process, we concede that consumers might behave differently if their choices had meaning. Therefore, we extended the findings of Studies 2a and 2b to an incentive compatible setting.

In this study, we offered participants the opportunity to allocate a small donation to charity. Participants could each allocate $0.15 to one of two charities, Doctors Without Borders/ Médecins Sans Frontières (MSF) or Save the Children (STC). We actually donated $120 to these two charities in accordance with the participants' allocations and set up a public website where participants could verify the donations upon completion of the study. Because participants' choices were actually meaningful, they were incentivized to be honest about their preferences for how to allocate the money to charity, which should obviate any concerns about whether randomly selecting imaginary stimuli is illustrative of a true decision-making process.

Participants and Procedure

Eight hundred fifteen Americans were recruited from MTurk. All were required to have an approval rating of at least 95%. After the study finished collecting data, we noticed that 31 responses were from duplicate worker IDs, and we dropped these responses. A further 29 participants were excluded because they incorrectly responded to a pre-registered attention check item. This yielded a final sample size of 755 participants. Each participant was randomly assigned to one of four conditions. These four conditions are identical to those used in Study 2b.

Participants were first told that they could allocate a $0.15 donation between two charities, MSF or STC. We explained that we would recruit 800 participants and donate $120 in total to the two charities. We included a public link to an OSF page where participants could verify that we actually donated to the charities upon conclusion of the study. All participants next saw a one-page description of each charity in a random order. Next, participants were told that they could allocate $0.15 to charity by choosing a set of three logos, where each logo represented a $0.05 donation to the charity whose logo they selected. How participants selected their bundles of charity logos/donations varied by condition. As in Studies 2a and 2b, those in the constructed bundle condition made their donations one at a time by selecting from a set of 6 MSF logos and 6 STC logos. Those in the prepackaged bundle condition selected among the four unique combinations of MSF and STC logos that can form three-logo bundles, while those in the expanded prepackaged bundle condition saw all eight possible permutations of bundles. Participants In the restricted constructed bundle condition, participants made two choices as before, first selecting two logos of the same type and then selecting a third logo on a separate screen. At the end of the study, participants completed an attention check item, some brief demographic items, and were again shown the link to verify their charity donations

Results and Discussion

As before, we began by classifying participants' selection in terms of whether they selected a high-variety bundle (two of one charity, one of the other) or a low-variety bundle (all the same charity). A chi-square analysis revealed that participants' preferences for variety significantly differed among the four conditions, $\chi2$ (3, N = 755) = 11.40, p = .010. We then tested the pairwise comparisons between conditions. First, those in the standard constructed bundle condition were more likely to choose a varied bundle (66.85%; 123 / 184) than those who selected among the four prepackaged bouquets (56.08%; 106 / 189), $\chi^2(1, N = 373) = 4.56, p = .033$. Second, we found that by expanding the number of prepackaged bundles that displayed variety, we increased choice of variety (66.67%; 128 / 192), $\chi^2(1, N = 381) = 4.50, p = .034$. We found no difference between the constructed bundle and expanded prepackaged bundle conditions, $\chi^2(1, N = 376) < 1$. This again lends support to the idea that our proposed theoretical account (choice myopia combined with different pathways to variety) *fully* accounts for the observed effect of the choice process on preference for variety. Once again, with the pathways to variety restricted, sequential choosers (in the restricted constructed bundle condition) were relatively unlikely to choose a varied bundle (53.68%; 102 / 190). Again, in an incentive compatible context, this restriction prompted a diminished interest in variety compared to the standard constructed bundle condition, $\chi^2(1\ N = 374) = 6.76, p = .009$. Unlike in Study 2b, participants in this restricted sequential choice condition did not differ in preference for varied charity donations from those in the standard pre-packaged bundle condition, $\chi^2(1, N = 379) < 1$.

Participants' preferences for allocating money to charity was predicted by the number of pathways to variety they could encounter, not whether they constructed their bundle sequentially or chose among prepackaged bundles. Pathways to variety continues to be a meaningful predictor of choice even when participants are incentivized to be honest about and aware of their preferences.

STUDIES 4A-4C: A LINGERING CONUNDRUM OR A FALSE POSITIVE?

Our studies suggest that the choice procedure can influence the choice of variety, but it does so through offering more pathways to variety and not through a change in preferences. With that identified, it is worth returning to the original evidence presented by Mittelman et al. (2014). They presented three demonstrations that seem to contradict our account. Two of those seem to be explained by participant confusion with measurement wording (as demonstrated in our Studies 1a and 1b). Their third demonstration needs an explanation.

In their Experiment 4, participants either constructed a bundle of six candies or chose among two prepackaged six-candy bundles. When participants constructed a bundle themselves, they first received three different-flavored candies (ABC). They could then complete their bundle by adding the same three candies (low-variety choice) or by adding three more A's (high-variety choice). Note that Mittelman et al. defined variety by variability in what was chosen (ABC and then AAA) instead of variability in the outcome (the bundle ABCAAA is predominately an A bundle). Other participants merely indicated whether they preferred ABCAAA or ABCABC. When choosing sequentially, participants showed an increased interest in ABCAAA, the bundle created by showing variety in one's choices (63.3% of participants selected ABCAAA in the sequential choice condition, compared to 33.3% of participants who directly selected their bundle).

Our theoretical account rooted in choice myopia combined with pathways to variety cannot account for this result. Does this signal that our theoretical account is incomplete? Before embarking on such a search, we first wanted to determine whether those results were replicable. We conducted three replications, all with more than three times the sample size of the original study (Mittelman et al. 2014, Experiment 4).

We describe these efforts briefly below. To foreshadow, we never replicated the originally reported findings. Because we did not start out knowing that the original result would not replicate, each effort included additional conditions that had the potential to help us make sense of the effects. Given the basic effects never replicated, these conditions became moot; the datasets including these additional conditions and the associated study materials are included on the Open Science Framework page for this project.

Studies 4a and 4b: Replications with Original Stimuli

We *a priori* decided to collect data from 100 participants per condition. This compares to sixty participants in total in Experiment 4 of Mittelman et al. (2014). Participants (N_{3a} = 205; N_{3b} = 200; Mechanical Turk) considered the three candy flavors used in Mittelman et al. (Experiment 4). In these bundles, A = cherry, B = grape, and C = apricot. For those constructing a bundle, they began with ABC and then had to decide whether to add AAA or ABC. Those choosing among prepackaged bundles selected either ABCAAA or ABCABC.

Study 4a. Among those constructing a bundle, participants were relatively unlikely to make a varied choice by selecting AAA (31.07%; 32 / 103). Among those participants who selected directly between the two prepackaged bundles there was a directional, but not statistically significant increase in participants selecting ABC + AAA. The proportion who selected ABCAAA directly was 41.18% (42 / 102). This did not significantly differ from the sequential choice condition, $\chi^2(1, N = 205)$, = 2.27 $p = .13$.

Study4b. Study 4a suggested that the original result was not replicable, but we did not want to dismiss it just yet. We conducted a second direct replication with a sample drawn from the same population. This time, we found an effect in the opposite direction of the original finding. That is, those constructing bundles themselves were relatively unlikely to make a varied choice by selecting AAA (36%; 36 / 100), a percentage that was actually *smaller* than the proportion who selected ABCAAA directly (52%; 52 / 100), $\chi^2(1, N = 200)$, = 5.19 $p = .02$.

Study 4c: Conceptual Replication

Studies 4a and 4b provided no support for the possibility that constructing a bundle sequentially would increase the varied nature of participants' choices. We considered that perhaps this was related to the fact that Mittelman et al.'s study was conducted in Argentina, where candy flavor preferences likely differ from those of the American sample. We showed 100 Americans from American Turk 12 candy flavors and asked them to rate their liking for each (1 = not at all, 7 = very much). We selected the 5th (lemon), 6th (peach), and 7th (apple) ranked candies to be A, B, and C. In our main study, which we decided to power with 2.5 times the sample size of Studies 4a and 4b, ($N = 503$), those first endowed with ABC were again relatively unlikely to make a varied choice by selecting AAA (27.09%; 68 / 251). This percentage was similar to the percentage of prepackaged bundle participants who chose ABCAAA directly (27.78%; 70 / 252), $\chi^2(1, N = 503) = .03$, $p = .86$. Once again, we find no differences in varied choice when the pathways to variety were equated.

GENERAL DISCUSSION

We both revisited and built upon past literature on how the bundle selection process influences a preference for variety. Previous researchers argued that when consumers construct bundles sequentially (instead of selecting among prepackaged bundles), they prefer more variety (Mittelman et al., 2014). In seeking to offer the first explanation for this effect, we instead found that the original claim needed to be revised. That is, we found that three of the four studies purporting to show the effect should not factor into our theorizing. Two studies' results could be chalked up to participant misunderstanding of measures (Mittelman et al 2014, Experiments 1 and 2). One study did not replicate in three high-powered direct and conceptual replications (Mittelman et al 2014, Experiment 4). Crucially, we propose setting aside these findings not in an effort to dismiss the importance of the bundle selection process, but to allow us to focus on those findings that are robust and thus amenable to theoretical explanation.

We argue that by appreciating the short-sightedness of decision makers—a quality we call choice myopia—we can understand why those who construct bundles sequentially sometimes arrive at different bundles than those who choose among prepackaged bundles. As decision makers construct a bundle themselves, they often have more choice pathways by which to arrive at a high-variety than a low-variety bundle. As an example, consider the construction of

a six-pack of beer. A low-variety bundle (e.g., 6 Chimay Blues) can be constructed in one way: the selection of a Chimay Blue six times in a row. If the bundle had just a touch of variety (e.g., 5 Chimay Blues, 1 Chimay White), the consumer now has six unique pathways to arrive at such a bundle; the lone Chimay White could be selected first, second, third, fourth, fifth, or sixth. When looking at a prepackaged bundle, the all-Blue bundle and the five-blue, one-white bundle are merely two options on equal footing; it does not matter that there were many more pathways by which the latter bundle could have been constructed. Simply because variety can be more easily achieved through the sequential construction of a bundle, it more frequently is. By experimentally disentangling the process by which bundles are selected from the number of pathways to variety these choice processes offer, we showed that it is the latter feature that explains when and why sequential choosers construct more varied bundles.

The present research is the first to recognize the importance of choice architecture in producing different numbers of pathways to variety and thus variety-seeking behavior, we earlier drew parallels to previous research that has examined how judgments and decisions tend to spread over decision sets. Our work offers two primary advances. First, we hold the possible set of final bundles constant, in contrast to research that has examined what happens when the set of available options expands (Benartzi and Thaler 2001). Second, we did not unpack superordinate categories into subordinate categories, a move that could signal the relative importance or scope of those subcategories (Fox et al. 2005). In contrast, our work focuses on the role of choice architecture not in modifying the choice outcomes themselves, but merely in affecting the pathways to get to them.

Choice Myopia Revisited

In considering the role of choice myopia, we focused on a mechanism (pathways to variety) that enhances variety seeking among those constructing bundles sequentially. But might there also exist a mechanism that encourages the selection of *low* variety among those considering the prepackaged bundles? That is, we have argued that sequential choosers are myopic (i.e., not carefully envisioning the final bundles toward which they are heading). Are they missing something that their prepackaged-considering counterparts find appealing about low-variety bundles?

One possibility is suggested by research on set-fit effects (Evers, Inbar, and Zeelenberg 2014). That research stresses that utility offered by whole bundles is distinct from the utility granted by the sum of its parts. More specifically, people prefer bundles that are all-different (i.e., including all distinct elements) or all-same (e.g., a single-color bouquet) over less gestalt-pleasing bundles (i.e., those that include only some repetition of elements). Bundles with the least variety (i.e., all-same bundles) necessarily benefit from set fit.

Myopic sequential choosers may not be forward looking enough to identify and move toward this unique value. For example, those selecting among prepackaged bouquets may more easily identify the simple beauty of a three-orange bouquet; those constructing bundles sequentially may be too shortsightedly focused on their next choice instead of on whether they are moving toward a final bundle with the most pleasing gestalt. In the present paper, we focused on testing the pathways to variety account given previous research has already established the importance of set-fit to choice. Whether pathways to variety still encourages more choice of variety even when the bundles' resonance with set fit is held constant (e.g., the selection of 3-

orange, 2-yellow vs. 4-orange, 1-yellow bouquets—two bundle compositions that differ in their pathways to variety but are similar in their lack of resonance with the rules of set fit) is a more nuanced question that awaits empirical investigation.

Implications for Marketing Practice

Consider the decision of how to structure the bundle choice process. Taking the perspective of the consumer, we can ask whether one structure is likely to lead to *better* choices than the other. This is of course a difficult question. By one perspective, we might think it best to consider which choice process unfolds most similarly to how the actual consumption episodes ultimately will. Through this lens, it might seem that decision makers should construct their bundles sequentially. After all, the beer drinker will ultimately consume those beers in sequence. Just as pathways to variety are exaggerated when constructing the bundle sequentially, the sequential unfolding of reality actually offers the same elevated opportunity for variety in consumption. On the other hand, previous research has suggested that decision makers select more variety for their future than their future selves ultimately would prefer (Ratner, Kahn, and Kahneman 1999; Read and Loewenstein 1995). For this reason, pushing decision makers to select among prepackaged bundles (at least when those bundles restrict pathways to variety) may nudge them toward more optimal choices.

Marketers themselves may have their own incentives for wanting consumers to select more or less variety. Consider the beer manufacturer who is attempting to promote sales of a new varietal. To encourage trial, the manufacturer may wish to include the beer in certain bundled offerings. The seller may have more luck getting customers to add the new varietal to their bundles when they construct their own six-packs (thereby offering six chances for the new varietal to be chosen) instead of having them select among prepackaged bundles. Furthermore, this approach may prove economically savvy, especially compared to a more traditional approach like an introductory price promotion. Examination of this possibility awaits relevant field research.

CONCLUSION

Even when scientists work in isolation, they take part in a collaborative enterprise. Researchers look to the efforts of others for guidance on what questions need asking, and which answers are most plausible. As the behavioral sciences have aimed to understand which of its own published literature is robust and replicable (Open Science Collaboration, 2015; Dreber et al., 2016), it can be easy to slip into a certain dichotomous thinking: Is a finding real or not? We hope the present paper offers a more nuanced answer to this question. A single paper may contain some findings which are seemingly not replicable, others that are perhaps artifactual, but still others that are novel, important, and yet imperfectly understood.

We greatly benefited from Mittelman et al.'s (2014) recognition that the bundle selection process is an interesting but otherwise neglected topic in consumer research. But surprisingly, moving the theoretical and empirical understanding forward required recognizing that some of the original evidence needed to be reconsidered and others challenged more directly. The published literature involves a mix of replicable, artifactual, and non-replicable findings. Only by carving out these three categories could we identify the most vital of the previous contributions and lay a foundation for others to now build further.

Chapter 3: Spending Guilt

Money is fundamentally fungible, meaning it can be exchanged for goods and services on demand. As a consequence, consumers should treat money as an exchange medium that allows them to maximize their outcomes by trading it for the goods they desire the most. In that sense, the exchange of money for products and services ought to be a cold and calculative affair. In reality, however, people have standards, or rules, about how they believe they ought to spend their money. A consumer who both wants to purchase and can afford a massage might be hesitant to do so because she could feel like this is not a good use of her money. In this instance, the consumer's desires to avoid violating her standards are limiting the fungibility of her money.

A number of situations have been documented in which consumers do not treat money as sufficiently fungible, resulting in what are considered classic examples of "irrational" behavior. The consequences of mental accounting (i.e., people partition money into distinct accounts and only use it for the purposes associated with each account, Thaler, 1985), pain of paying (i.e., the idea that paying for something with cash is painful, thereby reducing the likelihood of purchase, Zellermayer, 1996), and preference reversals (e.g., violations of procedure invariance associated with willingness-to-pay and choice, Lichtenstein and Slovic, 1971; O'Donnell and Evers, 2018) all represent instances in which money is treated as less than perfectly fungible. For example, O'Donnell and Evers (2018), found that when consumers were offered a choice between chocolate and toothpaste, many chose (preferred) the chocolate. However, when consumers were offered a chance to buy chocolate or toothpaste, many paid more (preferred) the toothpaste. Although money is thought to scale reliably with preferences, in practice, its use is subject to psychological rules about how it ought to be spent and we believe that violating these is likely to cause people to feel guilty.

Guilt is a negative, unpleasant emotion that is associated with falling short of a standard (Smith and Ellsworth, 1985; Kugler and Jones, 1992). It is distinguished from other negative evaluative emotions such as shame, because it is self-evaluative, rather than reliant on the negative evaluations of others (Tangney, Wagner, Fletcher, and Gramzow, 1992). We argue that spending guilt, or the feelings of guilt associated with spending money in ways that violate one's personal standards, plays a role in shaping consumers' spending behavior. We believe that understanding the role of spending guilt in the marketplace will allow us to predict when people will buy what they want and when the anticipation of guilt will prevent them from doing so.

We propose that consumers have standards about how they ought to spend their money. Children are socialized to save money (e.g., put money in a piggy bank), adults are told to follow strict budgets, and there are entire industries dedicated to managing savings and planning for retirement. The mere existence of these social norms is liable to create a standard for spending that consumers can meet or violate. This idea is consistent with work by Dahl, Honea, and Manchanda (2003), who identified three typographies for guilt in the marketing context (guilt related to societal standards, others, or oneself) and found that personal transgressions (like violating one's own spending standards) account for the largest share of consumer guilt.

Guilt is an aversive emotion and the experience of guilt leads people to atone for their transgressions (Tangney, Miller, Flicker, and Barlow, 1996). Here, we mostly focus on *anticipated* guilt. Because guilt is such an aversive emotion, the mere anticipation of experiencing guilt can lead consumers to alter their behavior so as to avoid feeling guilty

(Baumeister, Vohs, DeWall, and Zhang, 2007). Moreover, Duke and Amir (2018) demonstrated that consumers who take action to reduce felt guilt (e.g., by temporally separating decision and action guilt) are more likely to behave indulgently and less likely to attempt to make amends for their guilt-inducing behavior. These findings suggest that consumer behavior is susceptible to being influenced by the unpleasantness of feeling guilt.

Combining these causes and consequences of spending guilt, we expect that, conditional on a spending standard existing, a consumer who is more guilt prone (i.e., someone who has a greater dispositional likelihood to experience guilt, see Cohen, Wolf, Panter, and Insko, 2011) is more likely to experience the emotion of guilt and will feel worse about violating these standards. Additionally, the tendency for people to experience anticipatory guilt suggests that a consumer who is more spending-guilt-prone will be more likely to suppress her behavioral intentions to violate spending standards. As a consequence, the spending guilt prone consumer who opts not to buy a massage is sacrificing some of the fungibility of money to avoid the disutility associated with feeling guilty and the diminished enjoyment of the massage that occurs as a result of this disutility. A consumer who is less spending guilt prone could more easily spend the money on the massage and enjoy it.

We are not the first to posit that consumers may experience (dis)utility in transactions. For example, Thaler (1985) suggested that people evaluate potential transactions and make a decision to engage in those that seem like a good deal and avoid those that seem like bad deals. Similarly, others have posited that the spending of money can lead to negative feelings, or even pain ("pain of paying"; Prelec and Loewenstein, 1998; Zellermayer, 1996, Mazar, Plassman, Robitaille, and Lindner 2016). Prelec and Loewenstein (1998) suggest that the pain of paying for a good reduces the pleasure associated with the consumption of that good. Conceptually, we agree with the notion that transactions can cause disutility or reduce consumer enjoyment. However, our theoretical model differs from these other concepts in three crucial ways.

First, we discriminate between guilt and negative affect more broadly. We do not think the anticipation of spending money will activate sadness or anger, but instead, that it will activate guilt specifically. This is an important distinction because different emotions have distinct behavioral intentions associated with them (see Zeelenberg, Nelissen, Breigelmans, and Pieters, 2008). Second, we do not expect consumers to experience negative emotions from spending money *per se*. Instead, it is spending that violates one's standards which results in spending guilt. Finally, we expect these feeling of guilt to arise when consumers feel like they are spending money in a way that violates their standards, and not as a result of the literal exchange of cash. Whereas a pain of paying account would predict that spending cash on either ice cream or toilet paper would both feel painful, but buying either with a credit card would not (Raghubir and Srivastava, 2008), we instead expect feelings of guilt to arise when consumers buy ice-cream (with cash or a credit card), but not when buying toilet paper.

In sum, our theory can be distilled into the idea that spending money in specific ways that violate one's standards (or the anticipation of doing so), such as on frivolous goods, specifically activates the emotion of guilt. The aversiveness of (anticipatory) guilt then causes consumers to alter their spending behavior. In the studies that follow, we establish that consumers have generalizable standards for how money ought to be spend, develop and validate a measure for spending guilt, and use this measure to predict consumers' behavior.

STUDY 1: ESTABLISHING CONSUMERS' STANDARD FOR MONEY USAGE

All our predictions rely on the assumption that consumers set certain standards about how they *should* and *should not* spend money. In Study 1 we first asked participants on Mechanical Turk (MTurk) to describe some of their personal beliefs about how they should and should not spend their money (Study 1a). Then, using a new sample of participants, we measured the degree to which consumers agree with these standards, to determine if the standards participants generated in Study 1a are held by other people (and to demonstrate that there are broadly held standards for money usage, Study 1b). Finally, using yet another sample of participants, we measured consumers' anticipated emotional responses to situations in which they imagined spending money in ways that are believed to be *should* or *should nots* for money usage (Study 1c).

Study 1a: Participants and Procedure

We recruited 200 workers from MTurk (50.5% female, 0.50% non-binary, M_{age} = 35.61, SD = 11.68). All participants were first told that we were interested in learning about how people think about their money and what people think they should or should not do with their money. Next, all participants were asked to list two things that they thought they should do with their money and two things they thought they should not do with their money. Participants then completed some demographic measures and exited the survey.

Study 1a: Results and Discussion

We first read through each of the participants' responses for what they believed they should do with their money and what they believed they should not do with the money. After reading these responses, we developed a coding scheme for each set of items. For the *should* items, we coded participants' responses as reflecting a belief that they should save or invest money, pay their bills, or spend on necessities. For the *should not* items, we coded participants' responses as reflecting a belief that they should not gamble, spend money unnecessarily or frivolously, or spend money on junk food. Responses that did not fit these categories were coded as "other". Among the "other" responses, many in the *should* category expressed a belief that money should be "spent wisely." The coding results are presented in table 1.

Table 1. "Participants' coded responses for how they should and should not use their money. Items that did not fit into the pre-defined categories were coded as 'other'."

	Shoulds	Should Nots
Save/Invest Money	50.25%	-
	(205/400)	
Pay Bills	13.25%	-
	(53/400)	
Spend on Necessities	14%	-
	(56/400)	
Gamble	-	18.75%
		(75/400)
Spend Money Unnecessarily/Frivolously	-	29.75%
		(119/400)
Buy Junk Food	-	7.25%
		(29/400)
Other	21.5%	44.25%
	(86/400)	(177/400)

Study 1b: Participants and Procedure

Because Study 1a used an open answer format, many of the *should* and *should nots* that were most often reported were still only reported by a minority of the participants. To test the degree to which consumers actually hold these standards, we next showed each of the top three *should* and *should nots* to a new group of participants to estimate the degree to which consumers agree with them. We recruited a total of 201 workers from MTurk (47.26% female, 0% non-binary, $M_{age} = 35.63$, $SD = 11.60$) to participate in this study. A total of 18 participants responded incorrectly to our pre-registered attention check item, and were thus dropped from the remaining analyses, yielding a final sample of 183 participants. All participants were told that we were interested in learning how people think about the various ways they use their money and were asked to indicate the extent to which they believed they should do each of the top 3 *shoulds* and top 3 *should nots* generated from Study 1a. Participants responded on a 6-point scale (so as to avoid an indifferent option) anchored at 1 = "Definitely should not do" and 6 = "definitely should do." The *should* items were "save or invest your money", "pay your bills", and "purchase groceries and other necessities." The *should not* items were "gamble with your money", "purchase items that you want but do not necessarily need", and "buy junk food. Participants responded to the six items in a random order.

Study 1b: Results and Discussion

First, to demonstrate that each of these uses of money represents a standard that consumers generally hold, and to demonstrate that there are generalizable standards for how

consumers believe they ought to use their money, we investigated the proportion of participants who agreed that they should or should not do each behavior. We expected that the majority of participants would report a belief that they should do each of the *should* behaviors (i.e., select a 4 or higher on the six-point scale) and likewise that a majority of participants would report a belief that they should not do each of the *should not* behaviors (i.e., select a 3 or lower on the six-point scale). Our findings were consistent with this hypothesis, and we report the proportion who believe they should or should not do each behavior in table 2.

To further examine the generalizability of these beliefs, we next conducted a repeated measures ANOVA to test for the difference in means between the six behaviors. A repeated measures ANOVA revealed a significant main effect of behavior, $F(5, 910) = 477.74$, $p < .001$. We report the means for each behavior in Table 2. A *post hoc* contrast of the difference in means between the three *should* behaviors (coded 1/3) and the three *should not* behaviors (coded -1/3) likewise reveals a significant difference in participants' beliefs about whether they should engage in these behaviors, $F(1, 910) = 2,179.25$, $p < .001$.

Table 2. "Proportion of participants who indicated a belief that they should or should not engage in each spending behavior and mean scores for responses to the should not/should item for each behavior."

Behavior	Proportion Should	Proportion Should Not	Mean (out of 6) (SD)
Save or invest your money	93.44%	6.56%	5.21
	(171/183)	(12/183)	(.92)
Pay your bills	96.17%	3.83&	5.64
	(176/183)	(7/183)	(.77)
Purchase groceries and other necessities	97.27%	2.73%	5.55
	(178/183)	(5/183)	(.85)
Gamble	14.21%	85.79%	2.03
	(26/183)	(157/183)	(1.29)
Purchase items that you want, but do not necessarily need	41.53%	58.47%	3.26
	(76/184)	(107/183)	(1.16)
Buy junk food	19.67%	80.33%	2.56
	(36/183)	(147/183)	(1.25)

Study 1c: Participants and Procedure

We argue that consumers who violate their standards for spending behavior will not experience increased negative affect generally but will instead experience heightened guilt specifically. We recruited 151 workers from MTurk ($M_{age} = 33.62$, $SD = 9.20$, 49.7% female, 0% non-binary) and randomly assigned each participant to imagine how they would feel if they engaged in one of the six *should* or *should not* behaviors generated by respondents in Study 1a.

Participants predicted what they believed their emotions would be if they engaged in one of the behaviors by responding to the PANAS, a list of 10 positive and 10 negative emotions, on a 5-point scale ranging from 1 = "Not at all" to 5 = "Very much" (Watson, Clark, and Tellegen, 1988). We expected that the largest differences between the *should* and *should not* behaviors would be on guilt. Note though that, because the measures for each emotion were only 1-word items, we also expected participants to have difficulty discriminating between each emotion and predicted that although the greatest increase would be in anticipated guilt, this would not necessarily be the only difference between sets of behaviors.

Study 1c: Results and Discussion

For this study, we opted to focus on descriptive, rather than inferential statistics. We collapsed across the three *should* and three *should* not scenarios and calculated the average value for each emotion across these scenarios. We then computed the difference scores for each emotion between the sets of scenarios. The difference scores for the 20 emotions are reported in table 3. The difference between these sets of scenarios is largest for guilt, consistent with our theory that engaging in behavior that violates one's standards is likely to activate feelings of guilt.

Table 3. "Mean scores for the 20 PANAS items collapsed across the three *should* and *should not* behaviors as well as the difference in scores between these scenarios. Differences denoted with an * are significant at the .05 level."

Emotion	Should	Should Not	Difference
Interested	3.16	3.13	-.03
Distressed	1.76	2.59	.83*
Excited	2.73	2.99	.25
Upset	1.53	2.28	.74*
Strong	3.01	2.11	-.90*
Guilty	1.32	3.11	1.80*
Scared	1.53	2.26	.73*
Hostile	1.39	1.67	.28
Enthusiastic	2.80	2.59	-.21
Proud	3.33	1.88	-1.45*
Irritable	1.83	2.23	.40
Alert	3.16	3.13	-.03
Ashamed	1.25	2.43	1.18*
Inspired	2.60	2.04	-.56*
Nervous	1.71	2.68	.98*
Determined	3.59	2.53	-1.06*
Attentive	3.55	3.11	.44*
Jittery	1.59	2.25	.66*
Active	3.04	2.80	-.24
Afraid	1.63	2.20	.57*

While feelings of perceived guilt showed the largest difference between the *should* and *should not* behaviors, we wanted to test if this increase was greater than the next largest increase, which was for shame. To do this, we conducted a hierarchical linear regression predicting participants responses to the PANAS item from a dummy variable for behavior type (coded 1 for *should* and 0 for *should not*), a dummy variable indicating emotion (coded 1 for guilt and 0 for shame), and an interaction between these two variables. Because this is a mixed design, we included a random intercept term for respondents. The regression revealed a significant main

effect of guilt, meaning people tended to feel guiltier overall, $b = .68$, $z = 7.89$, $p < .001$, a significant main effect of *should*, indicating people felt less of both emotions in the *should* scenarios, $b = -1.18$, $z = -6.92$, $p < .001$, and a significant interaction effect, $b = -.62$, $z = -5.02$, $p < .001$, indicating that the increase in shame was smaller than an increase in guilt. Notably, although we expected lay people to have difficulty distinguishing shame and guilt, there was still a strong enough degree of specificity such that the increase in anticipated guilt was significantly larger than the increase in anticipated shame. Finally, the other emotion that shows a distinct difference in anticipated experience between the *should* and *should not* behaviors is pride. This is noteworthy, because in some ways pride is the complement to guilt. Like guilt, pride is self-evaluative, and is related to feelings of accomplishment and achievement (Tracy and Robins, 2007). Participants who imagined behaving in ways that violate standards feel reduced less, proud, which is consistent with feeling more guilt – rather than achieving, they imagine falling short, and the two emotions move in opposing direction.

STUDY 2 – DEVELOPING A MEASURE OF SPENDING GUILT

Having established that consumers indeed hold beliefs and have standards about how they feel they should and should not spend their money, we next turn to the measurement of anticipated spending guilt.

Item Generation and Content Analysis

Based on the responses collected in Study 1a, existing measures of guilt proneness (GASP; Cohen, Wolf, Panter, and Insko, 2011) and a money usage scale (Furnham, 1984), we developed 22 candidate items for a spending guilt scale.

Initial Sample

Participants and procedure. We administered these 22 items to a sample of 505 workers (51.29% female, 0.40% non-binary, $M_{age} = 36.52$, $SD = 11.96$) recruited from MTurk. Participants indicated their agreement with the extent to which they believed each of the 22 guilt scale items was true of them on a 7-point scale. We also included the "Spendthrift/Tightwad Scale" (4 items summed and scored with a range of 4 to 26; Rick, Cryder, and Loewenstein, 2008) and a one item measure for pain of paying, "How painful would spending money frivolously feel" (measured on an 11-point scale, 1 = painful and 11 = pleasurable) to begin to establish discriminant validity.

Results and Discussion. We factor analyzed participants' responses to these 22 spending guilt items using an iterated factor process. The unrotated factor analysis strongly suggested a 1-factor solution, with an eigenvalue of 10.14 associated with the first factor and 65.97% of the variance explained by this factor (see figure 1). Given that the items were written to support a single factor, the high proportion of variance associated with the one-factor solution, and the large difference in eigenvalues associated with the first and second factors, we opted to retain a one-factor solution.

To reduce the number of items in the scale, we eliminated items with factor loadings less than .50. This criterion led to the elimination of three items from the scale (see table 4).

We calculated participants' scores for spending guilt by averaging the 19 items with factor loadings greater than .50, $M_{\text{spending guilt}}$ = 4.61, SD = 1.27, Cronbach's α = .95.

Spending guilt did not differ by gender (male vs. other, $t(503)$ = .11, p = .92) but was negatively associated with age, β = -.12, $t(503)$ = -2.68, p = .008.

The 19-item composite correlated moderately negatively with the pain of paying item (as expected, higher spending guilt is associated with more pain of paying) r = -.31 and the Spendthrift/Tightwad scale r = -.26 (also as expected, higher spending guilt is associated with being a "tightwad"). These modest correlations are encouraging because they are directionally consistent with theoretically relevant constructs, but not so large as to render the spending guilt scale redundant.

Table 4. "Factor loadings associated with the single spending guilt factor for each of the 22 items included in the initial set of spending guilt scale items."

Variable	Factor1		Variable	Factor1
guilt18	0.8269		guilt8	0.7039
guilt7	0.8055		guilt10	0.6946
guilt6	0.8037		guilt14	0.6724
guilt12	0.7942		guilt17	0.6493
guilt15	0.7926		guilt21	0.5871
guilt22	0.7831		guilt13	0.5841
guilt20	0.7762		guilt16	0.5539
guilt19	0.7689		guilt11	0.5268
guilt5	0.7611		guilt3	0.421
guilt2	0.7303		guilt4	0.3578
guilt9	0.7123		guilt1	0.2182

Figure 1. Scree plot displaying eigenvalues associated with each factor after factor analyzing participants' responses to the 22 spending guilt items.
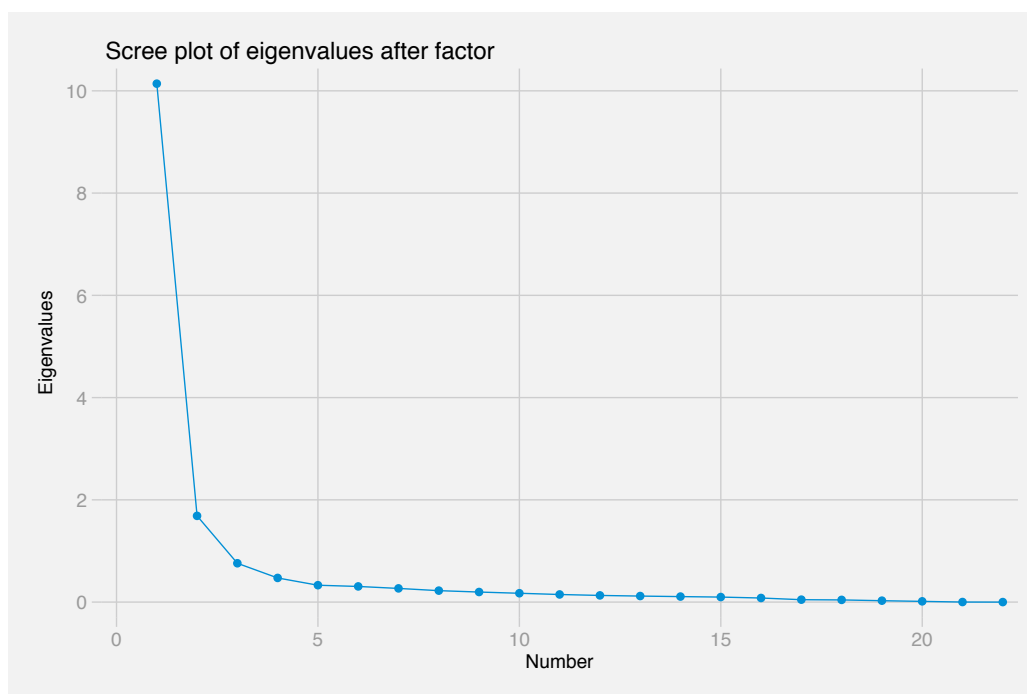
Nationally Representative Sample

Participants and Procedure. We next administered the 19 retained items and a number of personality measures to a large nationally representative sample maintained by Luth Research, an online marketing research firm. Luth Research recruited 2,283 participants to complete the survey, however we excluded 230 participants who incorrectly responded to a pre-registered attention check item. The remaining sample consisted of 2,053 completed responses (62.83% female, 0.39% non-binary, M_{age} = 52.69[1], SD = 13.91). Participants completed the 19 guilt scale items, the Spendthrift/Tightwad Scale, the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, and Swann, 2003) and the Guilt and Shame Proneness Scale (GASP) with the responses to the latter two surveys measured on 7-point scales.

Results and Discussion. We factor analyzed participants' responses to the 19 spending guilt items using an iterated factor process. The unrotated factor analysis again strongly supported a one factor solution, with an eigenvalue of 10.58 associated with the first factor and 79.22% of variance explained by this factor. To further reduce the number of items on this scale, we eliminated items with a factor loading of less than .70, in accordance with our pre-registered item removal criterion. This led to the removal of 4 items, leaving 15 items remaining for the scale. We *post hoc* opted to apply a more stringent criterion of retaining items with factor loadings greater than .80 to further reduce the number items included in the scale, which yielded 5 items for a short form of the spending guilt scale. We report factor loadings for each item in table 5.

---

[1] We recoded one response of -64 and another response of 8 as missing given the unlikeliness of the former age and an age restriction put in place by Luth Research that would prohibit the latter.

We calculated participants' scores for spending guilt by averaging the 15 items with factor loadings greater than .70, $M_{\text{spending guilt}} = 3.73$, $SD = 1.55$, Cronbach's $\alpha = .96$. Spending guilt did not differ by gender (male vs. other, $t(2051) = 1.68$, $p = .09$) and was again negatively associated with age $\beta = -.22$, $t(2051) = -10.43$, $p < .001$. We calculated correlations between the participants' composite scores on the 15-item spending guilt scale, openness, conscientiousness, agreeableness, neuroticism, the spendthrift/tightwad scale, the negative behavior evaluation and repair subscales of the GASP, and the pain of paying item. We report the correlation matrix in table 6.

Importantly, the spending guilt scale correlates negatively with emotional stability (positively with neuroticism) and negatively with both pain of paying and the spendthrift/tightwad scale, rs = -.30, -.17, -.28, respectively. These correlations are again directionally consistent with these theoretically relevant constructs, but not so large as to render the spending guilt scale redundant. The spending guilt scale correlates positively with both the Guilt – Negative Behavior Evaluation and Guilt – Repair subscales of the GASP, rs = .07 and .10, respectively. The correlation with the GASP guilt measures is somewhat low, but this could be because guilt proneness is measured by participants' responses to a number of scenarios that they are unlikely to encounter in day-to-day life (e.g., "You are privately informed that you are the only one in your group that did not make the honor society because you skipped too many days of school. What is the likelihood that this would lead you to become more responsible about attending school?").

Table 5. "Factor loadings associated with the single spending guilt factor for each of the 19 spending guilt scale items administered to the nationally representative sample."

Variable	Factor1
guilt6_1	0.8462
guilt18_1	0.8322
guilt5_1	0.8279
guilt22_1	0.8189
guilt19_1	0.8049
guilt20_1	0.7874
guilt7_1	0.7844
guilt15_1	0.7831
guilt12_1	0.7748
guilt8_1	0.7712

Variable	Factor1
guilt10_1	0.7677
guilt9_1	0.7494
guilt2_1	0.7493
guilt21_1	0.7257
guilt17_1	0.717
guilt14_1	0.6656
guilt16_1	0.6301
guilt13_1	0.5607
guilt11_1	0.4624

Table 6. "Correlation matrix for spending guilt, the Big Five Personality Factors, Spendthrift/Tightwad Scale, the GASP subscales, and the pain of paying measure."

	Spending Guilt	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability	Spendthrift/ Tightwad	GASP – NBE	GASP – Repair	Pain of Paying
Spending Guilt	1									
Openness	-.1111	1								
Conscientiousness	-.1892	.1820	1							
Extraversion	-.1061	.2954	-.0377*	1						
Agreeableness	-.1662	.2671	.3865	.0760	1					
Emotional Stability	-.2978	.2076	.4471	.1266	.4389	1				
Spendthrift/Tightwad	-.1683	.0552	-.3100	.1683	-.0368*	-.0776	1			
GASP – NBE	.0727	.0362*	.2319	-.0405*	.2848	.1310	-.0872	1		
GASP – Repair	.0974	.0822	.2980	-.0605	.3266	.1725	-.1318	.6833	1	
Pain of Paying	-.2849	.0670	-.1941	.1693	-.0730	.0239*	.4305	-.0791	-.0812	1

All correlations are significant except those indicated with an asterisk (*).

Test-Retest Reliability

Participants and Procedure. We re-contacted the initial sample of MTurk workers 55 days after they first completed the original set of spending guilt scale items. We were able to contact 500 out of the 505 original respondents because we did not have MTurk IDs for the remaining sample. Of the 500 participants whom we were able to re-contact, 253 responded within three days of being re-contacted. From this sample, we were able to match up the new set of responses with the initial responses for 247 workers. Three workers were excluded from the test-retest reliability analysis because they failed to correctly respond to a pre-registered attention check item. The final[6] sample consisted of 243 workers re-contacted from the MTurk sample pool (54% female, 0.40% non-binary, M_{age} = 39.54, SD = 12.89). Participants completed the 15-item spending guilt measure, the Spendthrift/Tightwad scale, the one item pain of paying measure, the TIPI, and the GASP guilt negative behavior evaluation and repair subscales.

Results and Discussion. First, we calculated a composite score for spending guilt by averaging participants' responses to the 15 items on the spending guilt scale, $M_{spending\ guilt}$ = 4.32, SD = 1.53, Cronbach's α = .96 (this is similar to participants' responses at time 1, $M_{spending\ guilt\ time\ 1}$ = 4.48, SD = 1.43). Next, we calculated a Cronbach's alpha reliability index between participants' scores on the spending guilt scale at time 1 and time 2 and found strong reliability between time 1 and time 2, α = .79. The raw correlation coefficient between the two time points was also high, r = .65.

Bland and Altman (1986) recommend supplementing statistical indices of temporal reliability with a visual approach. Thus, we constructed a Bland-Altman plot, which plots the average of participants' spending guilt scores at time 1 and time 2 against the difference in spending guilt scores between time 2 and time 1. Data points that fall more than ±2 standard errors outside of the mean difference are considered outside of the limits of agreement. Bland and Altman recommend pre-committing to an acceptable percentage of data points outside the limits of agreement. Although we did not pre-register such a limit, we believe the plot is still informative and find a relatively small number of points (18/244, 7.38%) outside the limits of agreement (see figure 2).

Finally, we calculated correlation coefficients for participants' spending guilt scale time 2 responses with the other personality measures included in this survey. This is a significantly smaller sample than the initial MTurk sample and the nationally representative Luth Research sample and thus the point estimates of the correlations are less reliable, but still directionally consistent with the other samples for each of the theoretically relevant constructs. See table 7.

---

[6] Participants may respond by 5/7/2019 and participants who respond by this deadline will be included in future analyses.

Figure 2. A Bland-Altman plot, depicting the average of participants' spending guilt scores at time 1 & time 2 plotted against the difference in spending guilt scores between time 2 & time 1.
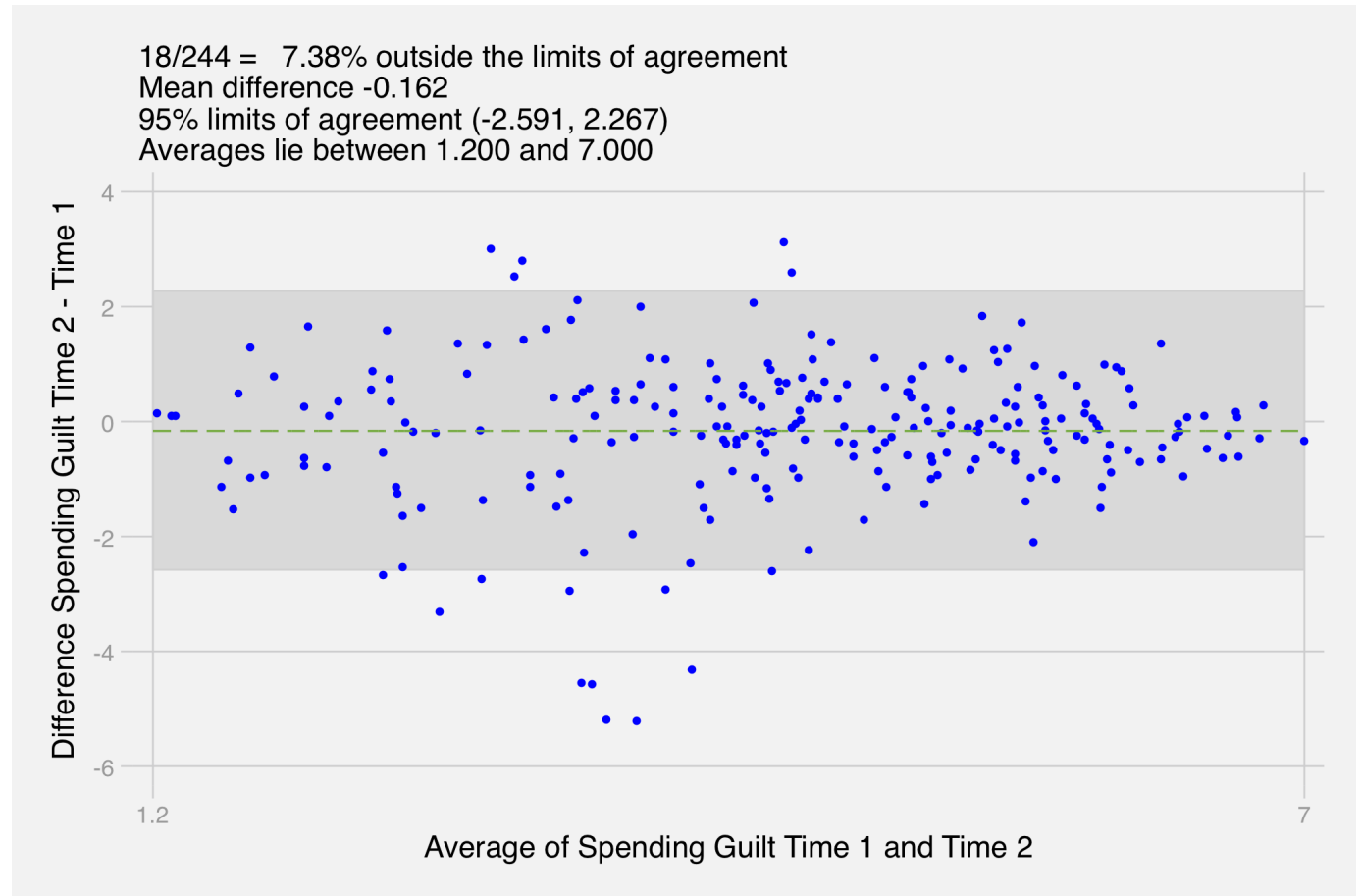

18/244 =   7.38% outside the limits of agreement
Mean difference -0.162
95% limits of agreement (-2.591, 2.267)
Averages lie between 1.200 and 7.000

Table 7. "Correlation matrix for spending guilt, the Big Five Personality Factors, Spendthrift/Tightwad Scale, the GASP subscales, and the pain of paying measure.

	Spending Guilt	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability	Spendthrift/ Tightwad	GASP – NBE	GASP – Repair	Pain of Paying
Spending Guilt	1									
Openness	-.1436	1								
Conscientiousness	-.1383	.2158	1							
Extraversion	-.2456	.2872	.1716	1						
Agreeableness	-.1478	.3417	.2518	.1632	1					
Emotional Stability	-.3432	.1989	.4810	.2423	.3061	1				
Spendthrift/Tightwad	-.1883	.0548*	-.2422	.1839	.0771*	-.0444*	1			
GASP – NBE	.1035*	.0175*	.1747	.0360*	.3278	.0353*	.0101*	1		
GASP – Repair	.1230*	.1381	.2369	.0322*	.3389	.1347	-.0339*	.5981	1	
Pain of Paying	-.4598	.0336*	-.1533	.1805	.0707*	.0581*	.5125	-.0462*	-.0580*	1

All correlations are significant except those indicated with an asterisk (*).

STUDY 3: PREDICTIVE VALIDITY OF THE SPENDING GUILT MEASURE

After developing a satisfactory 15-item measure of spending guilt, we next sought to test the predictive validity of the measure. Based on the large number of participants' text responses indicating that they would feel guilty spending their money on frivolous or unhealthy items, we reasoned that participants who experience a high degree of spending guilt are less likely to be willing to purchase hedonic products after controlling for their liking of each product. We do not expect spending guilt to relate to willingness to purchase utilitarian products, given the high proportion of participants who indicated that they believe spending money on necessities and functional items is a good use of money.

Participants and Procedure

We recruited 502 workers from MTurk, however we excluded responses from 45 participants who failed one or both of two pre-registered attention check items. The final sample consisted of 455 participants (52.97% female, 1.76% non-binary, M_{age} = 37.50, SD = 12.12).

All participants first completed the 15-item spending guilt scale, measured on a 7-point Likert-type scale. Next, participants were presented with 10 products in random order. Participants evaluated 5 utilitarian items (a pen, a can opener, trash bags, toilet paper, and a toaster oven) and 5 hedonic items (a chocolate bar, a pair of movie tickets, a pint of ice cream, a voucher for a relaxation massage, and a new video game). For each item, participants rated how much they liked the item on a 7-point scale (anchored at 1 = dislike a great deal and 7 = like a great deal) and indicated whether they would buy the item at the current MSRP (participants' responses were binary yes/no responses). We varied between subjects whether participants rated liking first or whether they indicated if they would buy the item first, but we did not expect an order effect.

Results and Discussion

We first calculated participants' scores on the spending guilt composite measure by averaging their responses to the 15 items ($M_{spending\ guilt}$ = 4.43, SD = 1.38, Cronbach's α = .96). Next, we coded participants' responses to the purchase item (yes = 1 and no = 0) and coded each of the items to be hedonic or utilitarian, as indicated above and in our pre-registration document (hedonic = 1 and utilitarian = 0). We conducted a hierarchical linear regression analysis by regressing participants responses to the purchase item on the spending guilt composite, the hedonic dummy variable, and an interaction term between the hedonic dummy variable and the spending guilt composite. We also included the continuous liking measure as a covariate in the model and allowed for a random slope to account for participant effects.

The hierarchical regression yielded a strong main effect of liking such that participants' are more likely to indicate that they will buy an item if they like it more, b = .14, z = 36.62, p < .001, and a main effect of hedonic product type, such that participants are more likely to indicate they would buy hedonic products than utilitarian products, b = .15, z = 3.96, p < .001. The regression results showed a non-significant main effect of spending guilt, b = -.00, z = -.19, p = .85, however because we coded products as 1 = hedonic and 0 = utilitarian, this is only the slope of the relationship between spending guilt and buying for utilitarian products. Most importantly, we did find a significant interaction between product type and spending guilt, b = -.03, z = -3.44,

$p = .001$. This interaction is the difference between the slope for hedonic and utilitarian items and indicates that there is a significant relationship between spending guilt and buying only for hedonic products, as predicted.

Because the outcome variable is binary, the regression output can be interpreted as the predicted probability of a participant indicating that they would buy a given item. To help interpret the regression output, one can consider the predicted probabilities of buying a hedonic or utilitarian item for a participant with an average liking score (collapsed across items) and either a high (+1 SD) or low (-1 SD) spending guilt score. Figure 3 shows these plotted probability values and clearly illustrates both the negative slope for hedonic items and the relatively flat slope for utilitarian items. This plot suggests that if two consumers, one who scores high on spending guilt proneness and the other who scores low on it, are considering buying a can opener, they are equally likely to buy it, so long as they like the product equally. However, if these same consumers are considering buying a massage, the consumer who has a high spending guilt proneness score is significantly less likely to purchase the massage, even if she likes it as much as the person who has a low spending guilt proneness score.

Figure 3. Predicted purchase likelihood of hedonic and utilitarian goods for a consumer who has a -1SD score (low) and +1SD score (high) on the spending guilt scale. Purchase likelihood scores are plotted at the mean value of product liking, collapsed across all products.
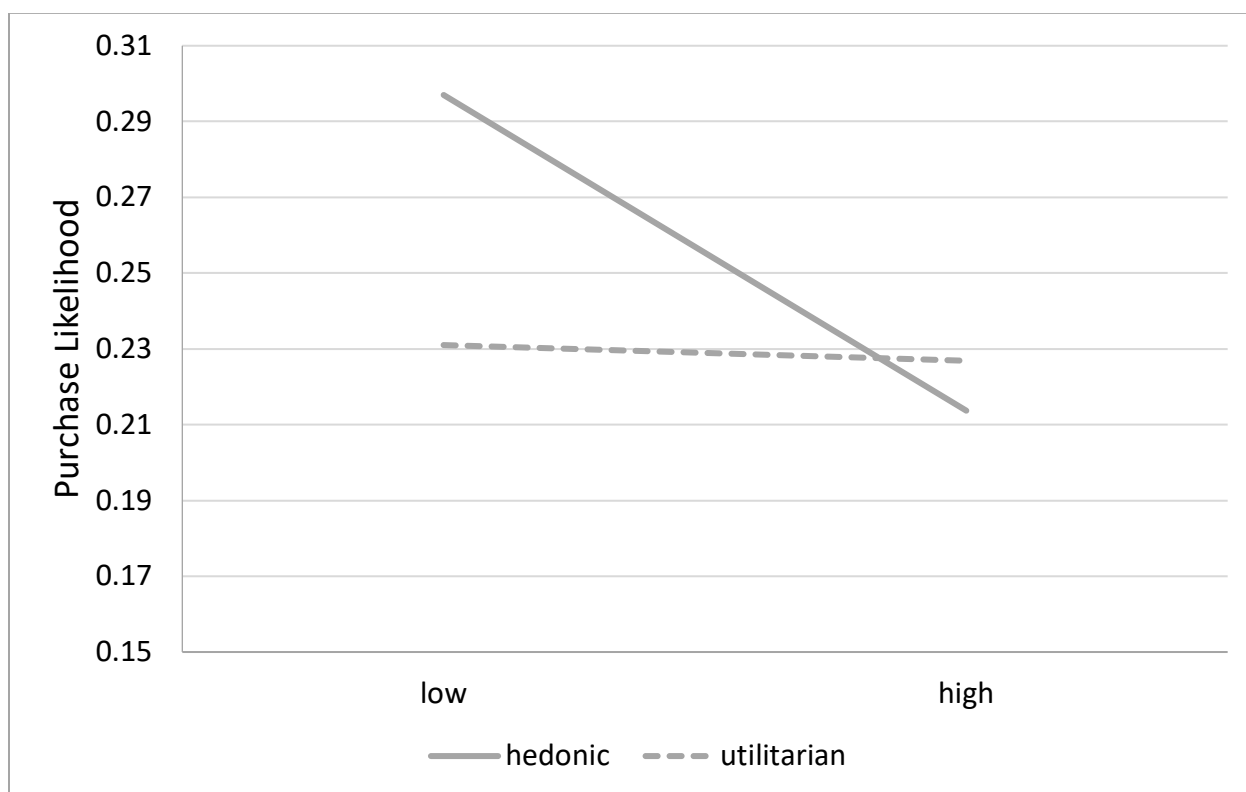
The results of this study suggest that spending guilt is a valid predictor of participants' willingness to buy hedonic items over and above the effects of product liking. This is consistent with our theory that spending guilt should be activated for items that are considered frivolous or unnecessary and should reduce willingness to buy these products.

GENERAL DISCUSSION

Consumers have standards and rules for how they spend their money. When they violate these standards, they are likely to feel guilty. Consumers who are more spending guilt prone, we argue, will anticipate this aversive state and behave in ways to prevent it from occurring. We validated a scale to measure this construct in large, nationally representative samples, showed strong test-retest reliability and predicted relevant behavioral outcomes from spending guilt proneness.

Our work in understanding spending guilt is far from complete. We have planned a number of follow-up studies that will test for the role of spending guilt in violations of fungibility. For instance, we predict that consumers who are more spending guilt prone will be more likely to show WTP and choice preference reversals due to spending guilt affecting WTP values. Furthermore, we would expect that if WTP is measured in non-monetary currencies such as points, spending guilt proneness will be less relevant and preference reversals will occur to a lesser extent. We also plan to study how spending guilt will affect monetary exchange. For instance, does spending guilt predict the extent to which consumers will exchange their money for hedonic or utilitarian products? We expect it will.

In sum, we believe the aversive self-evaluative emotion of guilt can be applied to participants' beliefs about how they spend their money. We showed that spending guilt is related to, but distinct from related concepts such as pain of paying and the extent to which consumers are spendthrifts or tightwads. The construct relates in sensible ways to other personality measures and is non-redundant with existing measures of guilt. The spending guilt scale is useful in that it predicts behavioral intentions over and above liking. We believe we have made a significant first step in uncovering and measuring a concept that could help predict when consumers will treat money as less than fungible and when they are less likely to purchase products, even if they both like and can afford to buy them.

Spending Guilt Items

Bolded items are included in the 15-item version of the spending guilt scale, bolded and italicized items are included in the 5-item short form of the scale.

1.) Relative to my income I tend to save quite a lot of money.
2.) **Even when I have enough money for the necessities, I often feel guilty about spending money on fun or pleasurable things.**
3.) In making any purchase, I always strongly consider how useful whatever I buy will be to me.
4.) I feel good about myself when I save money.
5.) ***If I spend my money on unnecessary purchases, I feel like I am violating my standards.***
6.) ***If I spend my money on unnecessary purchases, I often feel quite bad afterwards.***
7.) **When I spend my money unwisely, I feel like I have left myself down.**
8.) **I tend to feel guilt and remorse when I buy things.**
9.) **I often have a strong sense of regret when I make purchases.**
10.) **If I spend my money on unnecessary items, I feel like a bad person.**
11.) I have strict standards for how I use my money.
12.) **If I spend my money in ways I think are wrong, I will feel terrible about it, whether or not anyone knows about it.**
13.) I limit how I spend my money, so I don't feel bad about my purchases.
14.) Sometimes I won't buy something that I can afford because I think spending money on it would be wrong.
15.) **If I spend money unwisely, I judge myself harshly.**
16.) If I spend my money on unnecessary purchases, I try to make up for it by saving more.
17.) **If I spend my money frivolously, I feel like I should punish myself.**
18.) ***If I spend my money on unnecessary purchases, I feel like I'm doing something wrong.***
19.) ***If I spend my money on unnecessary purchases, I feel like I'm hurting my future self.***
20.) **I feel anxious after spending my money on frivolous purchases.**
21.) **If I spend money frivolously, I feel like I'm borrowing from the future.**
22.) ***When I spend money on unnecessary purchases, I feel like I'm not living up to my expectations of myself.***

Chapter 1 References

Ailawadi, Kusum L., Scott A. Neslin, and Karen Gedenk (2001), "Pursuing the Value-Conscious Consumer: Store Brands versus National Brand Promotions," *Journal of Marketing*, 65 (1), 71–89.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak (1964), "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9 (3), 226–32.

Chaiken, Shelly (1980), "Heuristic versus Systematic Information Processing and the Use of Source versus Message Cues in Persuasion," *Journal of Personality and Social Psychology*, 39 (5), 752.

Dhar, Ravi and Klaus Wertenbroch (2000), "Consumer Choice between Hedonic and Utilitarian Goods," *Journal of Marketing Research*, 37 (1), 60–71.

Ding, Min, Rajdeep Grewal, and John Liechty (2005), "Incentive-Aligned Conjoint Analysis," *Journal of Marketing Research*, 42 (1), 67–82.

Evers, Ellen R., Yoel Inbar, and Marcel Zeelenberg (2014), "Set-Fit Effects in Choice," *Journal of Experimental Psychology: General*, 143 (2), 504.

Faraji-Rad, Ali, Bendik M. Samuelsen, and Luk Warlop (2015), "On the Persuasiveness of Similar Others: The Role of Mentalizing and the Feeling of Certainty," *Journal of Consumer Research*, 42 (3), 458–71.

Frederick, Shane and George Loewenstein (2008), "Conflicting Motives in Evaluations of Sequences," *Journal of Risk and Uncertainty*, 37 (2–3), 221–35.

Grether, David M. and Charles R. Plott (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, 69 (4), 623–38.

He, Stephen X. and Samuel D. Bond (2015), "Why Is the Crowd Divided? Attribution for Dispersion in Online Word of Mouth," *Journal of Consumer Research*, 41 (6), 1509–27.

Hsee, Christopher K. (1996), "The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives," *Organizational Behavior and Human Decision Processes*, 67 (3), 247.

Hsee, Christopher K. and France Leclerc (1998), "Will Products Look More Attractive When Presented Separately or Together?" *Journal of Consumer Research*, 25 (2), 175–86.

Hsee, Christopher K., Yang Yang, and Bowen Ruan (2015), "The Mere-Reaction Effect: Even Nonpositive and Noninformative Reactions Can Reinforce Actions," *Journal of Consumer Research*, 42 (3), 420–34.

Jiang, Yuwei, Gerald J. Gorn, Maria Galli, and Amitava Chattopadhyay (2015), "Does Your Company Have the Right Logo? How and Why Circular- and Angular-Logo Shapes Influence Brand Attribute Judgments," *Journal of Consumer Research*, 42 (5), 709–26.

Johnson, Michael D. (1989), "The Differential Processing of Product Category and Noncomparable Choice Alternatives," *Journal of Consumer Research*, 16 (3), 300–9.

Kahneman, Daniel (2003), "A Perspective on Judgment and Choice: Mapping Bounded Rationality," *American Psychologist*, 58 (9), 697.

Kim, Hyeongmin Christian and Thomas Kramer (2015), "Do Materialists Prefer the 'Brand-as-Servant'? The Interactive Effect of Anthropomorphized Brand Roles and Materialism on Consumer Responses," *Journal of Consumer Research*, 42 (2), 284–99.

Klesse, Anne-Kathrin, Jonathan Levav, and Caroline Goukens (2015), "The Effect of Preference Expression Modality on Self Control," *Journal of Consumer Research*, 42 (4), 535–50.

Kupor, Daniella M. and Zakary L. Tormala (2015), "Persuasion, Interrupted: The Effect of Momentary Interruptions on Message Processing and Persuasion," *Journal of Consumer Research*, 42 (2), 300–15.

Laran, Juliano and Keith Wilcox (2011), "Choice, Rejection, and Elaboration on Preference-Inconsistent Alternatives," *Journal of Consumer Research*, 38 (2), 229–41.

Lee, Hyojin, Kentaro Fujita, Xiaoyan Deng, and H. Rao Unnava (2016), "The Role of Temporal Distance on the Color of Future Directed Imagery: A Construal-Level Perspective," *Journal of Consumer Research*, 43 (5), 707–25.

Lichtenstein, Sarah and Paul Slovic (1971), "Reversals of Preference between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89 (1), 46.

Lichtenstein, Sarah and Paul Slovic (1973), "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas," *Journal of Experimental Psychology*, 101 (1), 16.

Moon, Alice and Leif D. Nelson (2017), "The Uncertain Value of Uncertainty: When Consumers Are Unwilling to Pay for What They Like," working paper, Haas School of Business, University of California, Berkeley, Berkeley, CA 94720, https://ssrn.com/abstract.2676699.

Nowlis, Stephen M. and Itamar Simonson (1997), "Attribute-Task Compatibility as a Determinant of Consumer Preference Reversals," *Journal of Marketing Research*, 34 (2), 205–18.

O'Curry, Suzanne and Michal Strahilevitz (2001), "Probability and Mode of Acquisition Effects on Choices between Hedonic and Utilitarian Options," *Marketing Letters*, 12 (1), 37–49.

O'Guinn, Thomas Clayton, Robin J. Tanner, and Ahreum Maeng (2015), "Turning to Space: Social Density, Social Class, and the Value of Things in Stores," *Journal of Consumer Research*, 42 (2), 196–213.

Petty, Richard E., John T. Cacioppo, and Rachel Goldman (1981), "Personal Involvement as a Determinant of Argument-Based Persuasion," *Journal of Personality and Social Psychology*, 41 (5), 847.

Pham, Michel Tuan (1998), "Representativeness, Relevance, and the Use of Feelings in Decision Making," *Journal of Consumer Research*, 25 (2), 144–59.

Shaffer, Victoria A. and Hal R. Arkes (2009), "Preference Reversals in Evaluations of Cash versus Non-Cash Incentives," *Journal of Economic Psychology*, 30 (6), 859–72.

Shiv, Baba and Alexander Fedorikhin (1999), "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making," *Journal of Consumer Research*, 26 (3), 278–92.

Slovic, Paul, Melissa Finucane, Ellen Peters, and Donald G. MacGregor (2002), "Rational Actors or Rational Fools: Implications of the Affect Heuristic for Behavioral Economics," *Journal of Socio-Economics*, 31 (4), 329–42.

Slovic, Paul, Melissa Finucane, Ellen Peters, and Donald G. MacGregor (2007), "The Affect Heuristic," *European Journal of Operational Research*, 177 (3), 1333–52.

Slovic, Paul, Dale Griffin, and Amos Tversky (1990), "Compatibility Effects in Judgment and Choice," in *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, ed. Robin M. Hogarth, Chicago: University of Chicago Press, 5–27.

Smith, Rosanna K., George E. Newman, and Ravi Dhar (2016), "Closer to the Creator: Temporal Contagion Explains the Preference for Earlier Serial Numbers," *Journal of Consumer Research*, 42 (5), 653–68.

Tully, Stephanie M., Hal E. Hershfield, and Tom Meyvis (2015), "Seeking Lasting Enjoyment with Limited Money: Financial Constraints Increase Preference for Material Goods over Experiences," *Journal of Consumer Research*, 42 (1), 59–75.

Tversky, Amos, Paul Slovic, and Daniel Kahneman (1990), "The Causes of Preference Reversal," *American Economic Review*, 80 (1), 204–17.

Tversky, Amos and Richard H. Thaler (1990), "Anomalies: Preference Reversals," *Journal of Economic Perspectives*, 4 (2), 201–11.

Wegener, Duane T., Richard E. Petty, Kevin L. Blankenship, and Brian Detweiler-Bedell (2010), "Elaboration and Numerical Anchoring: Implications of Attitude Theories for Consumer Judgment and Decision Making," *Journal of Consumer Psychology*, 20 (1), 5–16.

Yang, Adelle X. and Oleg Urminsky (2015), "The Foresight Effect: Local Optimism Motivates Consistency and Local Pessimism Motivates Variety," *Journal of Consumer Research*, 42 (3), 361–77.

Zajonc, Robert B. (1968), "Attitudinal Effects of Mere Exposure," *Journal of Personality and Social Psychology*, 9 (2, Pt. 2), 1.

Chapter 2 References

Benartzi, Shlomo and Richard H. Thaler (2001), "Naive Diversification Strategies in Defined Contribution Saving Plans," *American Economic Review*, 91(1), 79-98.

Bever, Thomas G. (1970), "The Cognitive Basis for Linguistic Structures," *Cognition and the Development of Language*, 279(362), 1-61.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, ... & Emma Heikensten (2016), "Evaluating Replicability of Laboratory Experiments in Economics," *Science*, 351(6280), 1433-1436.

Evers, Ellen R., Yoel Inbar and Marcel Zeelenberg (2014), "Set-fit Effects in Choice," *Journal of Experimental Psychology: General*, 143(2), 504.

Ferreira, Fernanda and John M. Henderson (1991), "Recovery from Misanalyses of Garden-Path Sentences," *Journal of Memory and Language*, 30(6), 725-745.

Fox, Craig R., Rebecca K. Ratner, and Daniel S. Lieb (2005), "How Subjective Grouping of Options Influences Choice and Allocation: Diversification Bias and the Phenomenon of Partition Dependence," *Journal of Experimental Psychology: General*, *134*(4), 538.

Laibson, David (1997), "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112(2), 443-478.

Loewenstein, George (1996), "Out of Control: Visceral Influences on Behavior," *Organizational Behavior and Human Decision Processes*, 65(3), 272-292.

Loewenstein, George and Richard H. Thaler (1989), "Anomalies: Intertemporal Choice," *Journal of Economic Perspectives*, 3(4), 181-193.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman (2008), "Harnessing Our Inner Angels and Demons: What We Have Learned about Want/Should Conflicts and How that Knowledge Can Help Us Reduce Short-sighted Decision Making," *Perspectives on Psychological Science*, *3*(4), 324-338.

Mischel, Walter and Bert Moore (1973), "Effects of Attention to Symbolically Presented Rewards of Self-Control," *Journal of Personality and Social Psychology*, 28, 172–179. doi:10.1037/h0035716

Mittelman, Mauricio, Eduardo B. Andrade, Amitava Chattopadhyay, and C. Miguel Brendl (2014), "The Offer Framing Effect: Choosing Single Versus Bundled Offerings Affects Variety Seeking," *Journal of Consumer Research*, 41(4), 953-964.

Open Science Collaboration (2015), "Estimating the Reproducibility of Psychological Science." *Science* 349, no. 6251: aac4716.

Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman (1999), "Choosing Less-preferred Experiences for the Sake of Variety," *Journal of Consumer Research*, *26*(1), 1-15.

Read, Daniel, and George Loewenstein (1995), "Diversification Bias: Explaining the Discrepancy in Variety Seeking Between Combined and Separated Choices." *Journal of Experimental Psychology: Applied* 1.1: 34.

Read, Daniel, George Loewenstein, and Matthew Rabin (1999), "Choice Bracketing," *Journal of Risk and Uncertainty*, 19 (1–3), 171–97.

Redden, Joseph P. (2007), "Reducing Satiation: The Role of Categorization Level," *Journal of Consumer Research*, *34*(5), 624-634.

Schwarz, Norbert, Hans-J. Hippler, Brigitte Deutsch, and Fritz Strack (1985), "Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments," *Public Opinion Quarterly*, *49*(3), 388-395.

Simonson, Itamar (1990), "The Effect of Purchase Quantity and Timing on Variety-seeking Behavior," *Journal of Marketing Research*, 150-162.

Urminsky, Oleg and Gal Zauberman (2015), "The Psychology of Intertemporal Preferences," The Wiley Blackwell Handbook of Judgment and Decision Making, 141-181.

Chapter 3 References

Baumeister, Roy F., Kathleen D. Vohs, C. Nathan DeWall, and Liqing Zhang (2007), "How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, Rather Than Direct Causation," *Personality and Social Psychology Review*, 11(2), 167-203.

Bland, J. Martin and Douglas G. Altman (1986), "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement," *The Lancet*, 327(8476), 307-310.

Cohen, Taya R., Scott T. Wolf, Abigail T. Panter, and Chester A. Insko (2011), "Introducing the GASP Scale: A New Measure of Guilt and Shame Proneness," *Journal of Personality and Social Psychology*, 100(5), 947.

Dahl, Darren W., Heather Honea, and Rajesh V. Manchanda (2003), "The Nature of Self-Reported Guilt in Consumption Contexts," *Marketing Letters*, 14(3), 159-171.

Duke, Kristen E. and On Amir (2018), "Guilt Dynamics: Consequences of Temporally Separating Decisions and Actions," *Journal of Consumer Research*, 45(6), 1254-1273.

Furnham, Adrian (1984), "Many Sides of the Coin: The Psychology of Money Usage," *Personality and Individual Differences*, 5(5), 501-509.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann Jr. (2003), "A Very Brief Measure of the Big Five Personality Domains," *Journal of Research in Personality, 37, 504-528.*

Kugler, Karen and Warren H. Jones (1992), "On Conceptualizing and Assessing Guilt," *Journal of Personality and Social Psychology*, 62(2), 318.

Lichtenstein, Sarah and Paul Slovic (1971), "Reversals of Preference between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89(1), 46.

Mazar, Nina, Hilke Plassmann, Nicole Robitaille, and Axel Lindner (2016), "Pain of Paying?—A Metaphor Gone Literal: Evidence from Neural and Behavioral Science."

O'Donnell, Michael and Ellen R. Evers (2018), "Preference Reversals in Willingness to Pay and Choice," *Journal of Consumer Research*, 45(6), 1315-1330.

Prelec, Drazen and George Loewenstein (1998), "The Red and the Black: Mental Accounting of Savings and Debt," *Marketing Science*, 17(1), 4-28.

Raghubir, Priya and Joydeep Srivastava (2008), "Monopoly Money: The Effect of Payment Coupling and Form on Spending Behavior," *Journal of Experimental Psychology: Applied*, 14(3), 213.

Rick, Scott I., Cynthia E. Cryder, and George Loewenstein (2007), "Tightwads and Spendthrifts," *Journal of Consumer Research*, 34(6), 767-782.

Smith, Craig A. and Phoebe C. Ellsworth (1985), "Patterns of Cognitive Appraisal in Emotion," *Journal of Personality and Social Psychology*, 48(4), 813.

Tangney, June P., Patricia Wagner, Carey Fletcher, and Richard Gramzow (1992), "Shamed into Anger? The Relation of Shame and Guilt to Anger and Self-reported Aggression," *Journal of Personality and Social Psychology*, 62(4), 669.

Tangney, June P., Rowland S. Miller, Laura Flicker, and Deborah H. Barlow (1996), "Are Shame, Guilt, and Embarrassment Distinct Emotions?" *Journal of Personality and Social Psychology*, *70*(6), 1256.

Tracy, Jessica L. and Richard W. Robins (2007), "Emerging Insights into the Nature and Function of Pride," *Current Directions in Psychological Science*, *16*(3), 147-150.

Thaler, Richard (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4(3), 199-214.

Watson, David, Lee Anna Clark, and Auke Tellegen (1988), "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales," *Journal of Personality and Social Psychology*, *54*(6), 1063.

Zeelenberg, Marcel, Rob M. A. Nelissen, Seger M. Breugelmans, and Rik Pieters (2008), "On Emotion Specificity in Decision Making: Why Feeling Is for Doing," *Judgment and Decision Making*, 3(1), 18.

Zellermayer, Ofer (1996), "The Pain of Paying," unpublished dissertation, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA.