

# Preference Reversals in Willingness to Pay and Choice

MICHAEL O'DONNELL  
ELLEN R. K. EVERS

A fundamental contribution of consumer behavior research is to help marketing scholars develop an understanding of how people think about and express their preferences. In this article we find that two commonly used preference elicitation procedures, willingness to pay (WTP) and choice, are consistently associated with different expressed preferences. Specifically, choices are associated with a relatively greater preference for hedonic goods, while WTP is associated with a relatively greater preference for utilitarian goods. We find that this is caused, in part, by the greater reliance on deliberation in determining WTP values, while preferences in choices are determined by an affect heuristic. Unlike other choice and WTP preference reversals, we find that this effect is not caused by mechanical determinants such as scale compatibility, as the effect persists with continuous scale measures that rely on affect and with choice-based scale measures that rely on determining valuation.

**Keywords:** choices, willingness to pay, preferences, judgment and decision-making, affect, deliberation

Michael O'Donnell (mo279@berkeley.edu) is a PhD Student in the marketing group at the Haas School of Business, University of California Berkeley, Student Services Building, 2200 Piedmont Avenue, Room F502, Berkeley, CA 94720. Ellen R. K. Evers (evers@haas.berkeley.edu) is an assistant professor in the marketing group at the Haas School of Business, University of California Berkeley, Student Services Building, 2200 Piedmont Avenue, Room F528, Berkeley, CA 94720. Please address correspondence to Michael O'Donnell. This article is based on Michael O'Donnell's dissertation. Michael O'Donnell and Ellen Evers designed and analyzed all the studies in the article. Michael O'Donnell wrote the article and Ellen Evers provided feedback and revisions. We would like to thank Philippe van den Calseyde, Yoel Inbar, Alex Reese-Jones, Leif Nelson, Alice Moon, Don Moore, Clayton Critcher, and the University of California, Berkeley, Behavioral Lab for their valuable comments, questions, and suggestions throughout this project. We are also grateful for the feedback and comments provided by the reviewers and editors throughout the review process. The authors received funding from a Berkeley-Haas Behavioral Lab Minigrant and from a Berkeley Experimental Social Science Lab (Xlab) researcher grant. Supplementary materials are included in the web appendix accompanying the online version of this article. Preregistration documents, materials, and data can be found at the Open Science Framework page for this article: [osf.io/grxy8](https://osf.io/grxy8).

*Editor:* Vicki G. Morwitz

*Associate Editor:* Kristin Diehl

*Advance Access publication June 6, 2018*

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.

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). 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 pre-registered 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.



**TABLE 1**  
STUDIES 1A–1J PREFERENCE FOR UTILITARIAN GOOD, PARTICIPANT INFORMATION, AND STUDY FEATURES

Study	N	Preference (WTP) <sup>b</sup>	Preference (choice) <sup>b</sup>	Significance test <sup>b</sup>	Study features	Key point
1a <sup>a</sup>	300 (11 indifferent)	86.2% (81/94)	27.7% (54/195)	$\chi^2$ (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 <sup>a</sup>	200 (14 indifferent)	25.6% (23/90)	12.5% (12/96)	$\chi^2$ (1, $N$ = 186) = 5.18, $p$ = .02	Pen (MSRP: \$2.33) and chocolate bar (\$1.82).	Effect generalizes to situations in which there is an absolute preference for one good over the other.
1c <sup>a</sup>	202 (9 indifferent)	89.6% (86/96)	49.5% (48/97)	$\chi^2$ (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)	$\chi^2$ (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 Study 1c (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)	$\chi^2$ (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)	$\chi^2$ (1, $N$ = 192) = 26.47, $p$ < .001	Toaster (MSRP: \$39.99) and pair of movie tickets (MSRP: \$35) (nonconsumable items).	Effect generalizes to non-food domain.
1g	202 (13 indifferent)	85.9% (79/92)	24.7% (24/97)	$\chi^2$ (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)	$\chi^2$ (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 each other (no MSRP info).	Effect persists when the products are substitutes (i.e., the differences are not due to noncomparable products engendering different processing styles).
1i	200 (14 indifferent)	47.8% (43/90)	32.3% (31/96)	$\chi^2$ (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)	$\chi^2$ (1, $N$ = 368) = 27.29, $p$ < .001	A JC Penney store brand four-slice toaster (relatively more utilitarian; MSRP: \$85) and a KitchenAid two-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.

<sup>a</sup>Indicates items were presented in a vertical orientation; all other studies used a horizontal side-by-side orientation.

<sup>b</sup>The denominator and  $df$  for these columns are based on the sample size excluding those who were indifferent. See the [web appendix](#) for details.

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. Therefore, we used two products that are substitutes of each other in study 1h. 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.

## 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,  $M_{\text{age}} = 22.4$ ) 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,  $\chi^2(1, N = 283) = 7.22, p = .007$ .

### 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  $M_{\text{age}} = 22.2$ ,  $SD = 8.6$ ) 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<sup>1</sup> 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 generated price and enforce the outcomes of the BDM auction.<sup>2</sup> 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 trade-off 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.<sup>3</sup> 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. 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,  $X^2(1, N = 418) = 3.99$ ,

1 Bids were censored to a maximum of \$2.00; 35 participants bid \$2.00 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.

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 \$2.00.

3 For a detailed description of how we calculated extrapolated choices in wave 1, see the [web appendix](#).



$p = .046$ . 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.

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  $M_{\text{age}} = 36.4$ ,  $SD = 11.1$ ) 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.

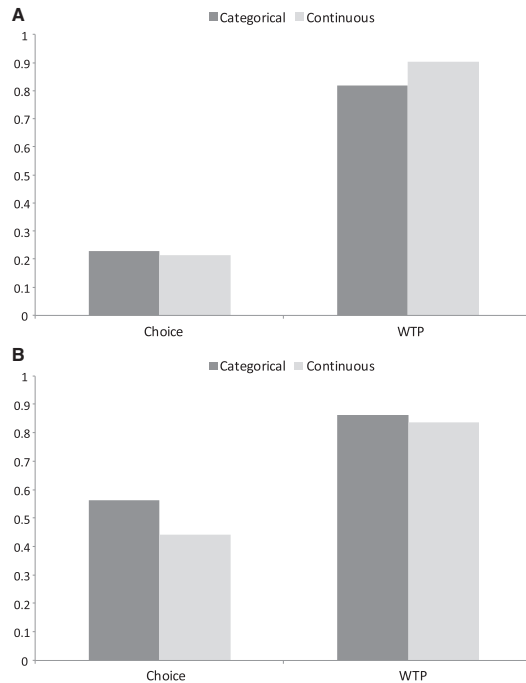
## Study 3a: Results

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

**FIGURE 1**

(A) RESULTS FOR STUDY 3A. EACH BAR SHOWS THE PROPORTION OF PARTICIPANTS WHO PREFERRED THE UTILITARIAN GOOD IN STUDY 3A, ORGANIZED BY PREFERENCE ELICITATION METHOD.

(B) RESULTS FOR STUDY 3B. EACH BAR SHOWS THE PROPORTION OF PARTICIPANTS WHO PREFERRED THE UTILITARIAN GOOD IN STUDY 3B, ORGANIZED BY PREFERENCE ELICITATION METHOD.



$n = 7$ , 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 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.

## 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 cost should be taken into account. 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 about the value of the options under consideration, whereas choices are expected to mostly reflect gut feelings. Consistent with this idea, Laran and

Wilcox (2011) found that the relative difference in elaboration between choosing and rejecting is associated with a change in preferences. Therefore, we expect deliberation to make participants' behavior in the choice condition 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)  $\times 2$  (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, 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. For all conditions, order of the products was counterbalanced.

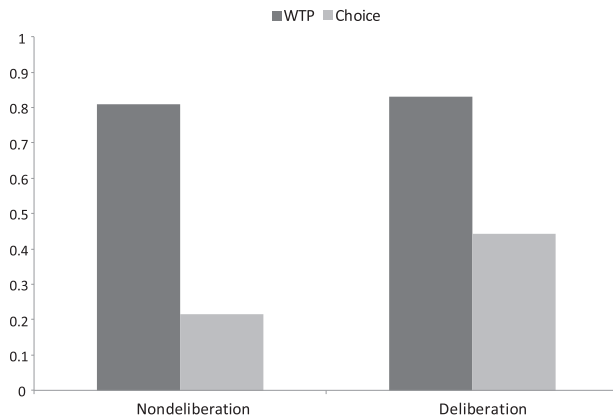
## Study 4: Results

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  $\chi^2(1, N = 194) = .13$ ,  $p = .72$ . However, consistent with our expectations, in the choice conditions participants who deliberated were significantly more likely to

FIGURE 2

RESULTS FOR STUDY 4. EACH BAR SHOWS THE PROPORTION OF PARTICIPANTS WHO PREFERRED THE UTILITARIAN GOOD IN STUDY 4.



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.

### 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. This idea is not incompatible with our data, but in essence is not more than an (incomplete) redescription 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 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 ( $M_s = 9.44$  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.

### DATA COLLECTION INFORMATION

The first author collected data under the supervision of the second author in all studies. For studies 1a–1j, studies 3a and 3b, and study 4, data were collected online through MTurk between early 2016 and late 2017. For study 2a, the first author supervised and collected data with a team of research assistants at UC Berkeley who collected the data in spring 2017. For study 2b, the first author supervised and collected data with a team of research assistants at UC

Berkeley who collected data in spring 2016. The first author analyzed the data for all studies under the supervision of the second author.

## 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, Hyeonmin 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.
- (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.
- (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.