Elicitation-based preference reversals in consumer goods

Michael O’Donnell¹
Ellen R. K. Evers¹
Iris C. Lew²

1. University of California, Berkeley: Haas School of Business
2. University of California, Berkeley: Psychology Department

Correspondence regarding this article should be addressed to Michael O’Donnell, michael_odonnell@haas.berkeley.edu
Student Services Building
2200 Piedmont Ave, Berkeley CA 94720

This paper is based on Michael O’Donnell’s dissertation. Michael O’Donnell and Ellen Evers designed and analyzed all the studies in the paper. Michael O’Donnell wrote the paper and Ellen Evers provided feedback and revisions to the manuscript. Iris Lew designed Study 5a in collaboration with Michael O’Donnell and Ellen Evers.

Acknowledgements: We would like to thank Philippe van den Calseyde, Yoel Inbar, Alex Reese-Jones and the University of California, Berkeley Behavioral Lab for their valuable comments, questions and suggestions throughout this project.
CONTRIBUTION STATEMENT

We demonstrate clear and consistent evidence for large and predictable preference reversals over consumer goods between choices and willingness-to-pay. While earlier research has demonstrated the existence of preference reversals between these two methods of preference elicitation, the mechanisms offered have been incomplete, non-parsimonious, or irrelevant to the stimuli we use in these studies. Importantly, we believe we have demonstrated evidence for a broader phenomenon that applies to nearly any domain of consumer research. We believe this work contributes to the field of consumer research because as the two gold standard measures of preference elicitation, many researchers use them interchangeably and expect equivalence. We show that willingness-to-pay and choice create different psychological contexts wherein participants reliably indicate different preferences. Specifically, we argue that choices and WTP engender inherently different levels of decision deliberation, and have a different level of monetary salience, leading consumers to behave differently in each case. Thus, any inference drawn based on the assumption of equivalence between these measures is flawed. We argue that an understanding of preferences must take into account the different psychological states that each elicitation method engenders in the process of eliciting preferences.
TITLE: Elicitation-based preference reversals in consumer goods

ABSTRACT

A key assumption in the empirical application of rational decision theory is that of procedure invariance; preferences are independent of how they are elicited. This means that measuring preferences among a bundle of goods should yield the same ordinal rankings, regardless of whether preferences are measured using a valuation strategy (e.g., willingness-to-pay) or choices. In 16 studies, the authors show that consumers more strongly prefer affective over functional goods in choices as compared to willingness-to-pay. These preference reversals result from a combination of two processes. The first is that decision-makers are more likely to rely on affect when making a choice. The second is that the concept of monetary exchange is more salient in a WTP setting. Contrary to the necessary empirical assumption that preferences are consistent across measurements, the authors find that participants treat different measurement techniques as entirely different situations.

Keywords: Preferences, Choice, Willingness to Pay, Framing, Preference Reversals
Consumer behavior researchers are interested in understanding what consumers like and don’t like and develop models of behavior and cognition, known as preferences. As researchers, we use the concept of preferences to describe our findings, and we assume that preferences in a given situation are stable as models of behavior for an individual, all other things being equal. It is necessarily true that different ways of measuring preferences over the same products should be equal, and it would be highly consequential for researchers if this assumption were incorrect. Here, we provide evidence that this assumption is incorrect, and that different preference elicitation procedures reliably yield divergent expressions of preference. We demonstrate that two gold-standard measures of preferences, choices and willingness-to-pay (WTP), reliably and predictably show different preferences among consumers for the same goods. Furthermore, we show that these findings are likely multiply-determined, and driven by the different levels of deliberation inherent in choices and WTP, as well as the different psychological context created by each measure.

At the center of consumer psychology and economics is rational choice theory, which is a normative prescription for how people behave (Von Neumann & Morgenstern, 1944; 2007). Because preferences are inherently idiosyncratic, single decisions cannot be compared to a normative standard. However, within each individual we do have a normative standard; all else being equal, rankings of outcomes should be consistent (Thaler, 1987). This means that the way in which preferences are measured (most commonly through willingness to pay or choice) should not affect the individuals ordering of preferences.

In the work that follows, we demonstrate consistent evidence for preference reversals that violate the logic of procedure invariance. Specifically, we demonstrate that preference reversals between choice and willingness-to-pay are robust, large in magnitude, consistent in direction
(and thus predictable), and logically inconsistent with normative choice theory. We are not merely demonstrating a psychological trick designed to be demonstrated in an unusual context, instead, we focus specifically on choice and willingness-to-pay because they are two gold standard measures of preference elicitation in consumer behavior research. Just as vision scientists study optical illusions to understand how sight works, we believe that understanding how and why these preference reversals occur allows us to develop a fuller understanding of what we as a field of consumer behavior researchers are measuring, and to more properly contextualize the conclusions we draw from our research (see Thaler, 1987).

Research in the field of consumer behavior involves the (explicit or implicit) measurement of preferences through choices or willingness-to-pay. A review of consumer behavior research papers published in the *Journal of Consumer Research* for 2015 shows that numerous findings rely on participants’ choices between consumer goods (Tully et al., He & Bond, Yang & Urminsky, Klesse et al.), their WTP for different goods (Kim & Kramer, Lee et al., Jiang et al., Kupor & Tormala, Hsee et al., Faraji-Rad et al.), or sometimes results from both measures (O’Guinn et al., Smith et al.) as representing their participants’ preferences. This non-exhaustive review indicates that consumer psychologists non-controversially believe that choices and WTP accurately assess preferences, and that occasionally researchers will use these measures interchangeably and assume that their findings across studies document convergent validity for a theory. Choices are often used because they are face valid. The simplest way to determine which item someone prefers among a number of goods is to ask the person to choose the one she prefers. Moreover, through a sequence of choices, researchers can determine an individual’s rank ordering among a series of goods. Similarly, willingness-to-pay ought to be an ideal measure of preferences (Jevons, 1875). Because money is readily transferred for goods and
services, it can be used to generate valuation, and valuations can be rank-ordered to represent preferences. Since choices and willingness to pay are both supposed to be accurate measures of preferences, they should be equivalent in that both of them reveal the same rank-order of preferences. To illustrate why this is the case, imagine a consumer who is willing to pay up to $10 to receive a box of trash-bags right now, and $5 to receive a pint of ice cream right now. This means that this consumer is indifferent when asked to choose between trash bags and $10, but would choose trash bags over $8. The same consumer would be indifferent between ice cream and $5 but would choose $8 over ice cream. This hypothetical consumer’s rank-ordered preferences would thus be [trash bags > $8 > ice cream] and therefore this consumer should choose trash bags over ice cream.

However, as we will show, there are many situations in which consumers will express preference reversals, such that they will reliably pay more for one good over the other but subsequently choose the other good over the first. In other words, participants reliably violate procedure invariance (Tversky & Thaler, 1990). While there is evidence that people violate procedure invariance when evaluating gambles (Lichtenstein and Slovic 1971; 1973) or in other narrowly defined contexts, no work has investigated, nor would predict, consumers to do the same thing in the face of valid, incentive compatible evaluations of a series of mundane and well-known consumer products. Specifically, the commonly understood explanation for preference reversals over gambles does not apply (scale compatibility, see; Tversky & Thaler, 1990; Tversky, Slovic, & Sattath, 1988; Tversky, Slovic, & Kahneman, 1990). Instead, the effects we document here appear to represent a fundamental violation of the assumption that choice and willingness-to-pay are measuring identical constructs, and suggest that consumer behavior researchers are drawing flawed conclusions from their research by assuming
equivalence among what turn out to be disparate measures of consumer preference.

In initially observing the differences in expressed preferences between choice and WTP, and then conducting numerous follow-up studies to determine why these differences occur, we have come to believe that the effect occurs because of predictable, but multiply determined processes. Thus, we believe this paper contributes to our understanding of consumer behavior by first documenting an effect and then delving deeper into its underlying mechanisms. We find that WTP is an inherently more deliberative process than choice, and that the relative difference in deliberation between these two elicitation procedures contributes to the differences observed between them. Additionally, consumers appear to read special importance into the concept of exchanging money for products, and this salience of monetary exchange, which can differ across elicitation procedures, leads to different expressions of preference.

**DIFFERENCES IN REQUIRED DELIBERATION**

When a consumer chooses between two goods, an approach that requires minimal cognitive effort (i.e., a heuristic) is to choose whatever feels best. Especially when the choice options are common and not very valuable goods, such an affect-based heuristic is a fast and low-risk approach to make simple choices (Slovic, 2002). However, although affective responses may help the consumer make 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 requires to consumer to consider a greater amount of information before revealing her preference, is likely to lead to a different expressed preference than simple choice.

The idea that choices are mostly influenced by immediate affective responses is
consistent with the arguments brought forward by Slovic et al. (2002) that 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 (1981) that source attractiveness is a source of persuasion when the target is evaluating in a low-effort manner. Furthermore, Evers, Zeelenberg, & Inbar (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.

While a fast-paced, heuristic reliance on affect is likely to be helpful and informative to consumers who are asked to make choices between goods, this process alone cannot be a suitable strategy for deriving the subjective monetary value of a product. Stated differently, when a consumer makes a choice between two options, she is faced with only one trade-off, and can easily maximize on immediate happiness. Generating WTP, however, requires a consumer to make a trade-off with respect to a larger menu of options — i.e., the universe of alternative goods that can be purchased with the same amount of money. We expect this tradeoff to require consideration of many attributes, such as the durability of the good, the market value of the product, and how immediately the good will be needed, as well as any number of other possibilities.

We do not believe that a more effortful decision-making process causes consumers to change their 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., 2009), and with the notion that different measurement processes are likely to call upon different motives for evaluating products (Frederick & Loewenstein, 2008).

**SALIENCE OF MONETARY EXCHANGE**

While the different levels of deliberation involved between making choices and indicating willingness-to-pay explains some of the difference in expressed preferences, we also find evidence for a second mechanism. Specifically, it appears that consumers treat money as being less fungible than it actually is because there are strong societal and cultural norms which dictate how money ought to be spent (Thaler, 1985; Thaler and Shefrin, 1981; for other studies suggesting that money is less fungible than it normatively should be, see Di Muro and Noseworthy, 2012; Raghubir and Srivastava, 2009). Cultural norms about the importance of money (e.g., that one should spend money wisely) suggest that money should be spent on more useful or practical items. Given this sequence of events, we believe that the increased salience of exchange is a contributing factor worthy of being highlighted separately from a broad deliberation-affect distinction.

If consumers indeed approach decisions about WTP and choices differently, this distinction allows us to derive some straightforward predictions.

**H1:** Consumers are more likely to indicate preferring a relatively functional good, to a more affective one when their preference is expressed through WTP as compared to choices.
Furthermore, because we expect these effects to occur as a result of the more deliberative process underlying the generation of WTP, from this, we can generate **H2**.

**H2:** Deliberation attenuates the difference in preferences by increasing choices for the functional goods in the choice condition.

More specifically, the attenuation of the difference in expressed preferences between WTP and choices should mostly result from a greater preference for the functional good when making choices in conjunction with deliberative thought, while preferences in the WTP condition should not change much as a result of deliberation.

Additionally, we believe that choice and WTP create different psychological contexts, which are associated with different preferences. From this, we generate **H3**.

**H3:** The tendency for expressed preferences to diverge between choice and WTP is partially the result of participants’ beliefs that money should be spent on useful goods rather than more hedonic goods.

Finding such robust and consistent differences in revealed preferences between choice and WTP should lead us to question conclusions drawn from earlier studies in which only one method of preference elicitation is used. After all, if the elicitation procedure by itself can strongly influence the preferences, one can question the extent to which stated preferences are actually reflective of consumers’ true preferences, or whether the concept of true preferences even exists. As researchers, we need to be mindful of how this instability can manifest itself in how we measure and document these preferences.

Our aim is greater than just making a mere, albeit highly important, methodological point. We believe the differences in expressed preferences between choice and WTP can also teach us about the psychological processes that underlie how consumers make decisions. If there
is a natural tendency for preferences to follow a hedonic course, such that people try to maximize their utility in the here and now, as suggested by the affect heuristic and hedonic discounting (Laibson, 1997), perhaps elicitation methods that rely on choice or affective processes are tapping into a direct, relatively unfiltered measure of hedonic maximization. Likewise, by inducing participants to undergo a more deliberative process – that is by having participants consider their valuation of the products under consideration, to rank any number of attributes upon which the products can be compared, and to take into account a wider range of foregone alternatives, it is possible that willingness-to-pay and other more intentional preference elicitation processes are capturing a more refined expression of preferences. Indeed, we demonstrate that different elicitation methods lead to the use of distinct strategies, each with their own underlying psychological processes. As such, the implication of our findings are that varying elicitation methods can lead consumers to maximize on distinct dimensions, leading to differences in expressed preferences.

It is worth noting that we do not propose that the two different elicitation methods that we focus on fully reflect the myriad different ways in which scholars measure preferences, nor do we claim that these two approaches perfectly reflect the ways consumers commonly make purchase decisions. We focus largely on choice and willingness-to-pay in this paper as they are by far the most common methods used by scholars in preference measurement.

OVERVIEW OF STUDIES

In what follows we demonstrate preference reversals that call into question the necessary assumption we hold as researchers that choice and WTP yield consistent preferences. More specifically, in Studies 1a-1c we test H1 and present ten hypothetical scenario studies in which participants either choose between or indicate their WTP for affective (or hedonic) and
functional (utilitarian, see; Dhar & Wertenbroch, 2000) products. Then in Studies 2a & 2b, we move away from hypothetical situations and employ two fully incentivized experiments. Consistent with the previous experiments, even with real money and real products on the line, we still find the same type of preference reversals. In Study 3, we more explicitly test the underlying mechanism and find that inducing participants to weakly deliberate for making a choice leads choices to look more like willingness-to-pay. In Study 4, we find evidence for H3 and demonstrate that participants believe that choice and WTP, while both measuring preferences, measure fundamentally different types of preference, rather than the same concept. Finally, in Study 5, we further test H3 by manipulating the psychological context of a preference-elicitation procedure. More specifically, we make choices feel more or less like a monetary exchange, and find that merely framing a choice as a monetary exchange increases preferences for the utilitarian good.

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 pre-determined a minimum sample size of 100 participants per condition; specifically, for most of the studies we collected 100 per cell, but for those which we expected a smaller effect size, we pre-registered a larger sample size and did not analyze any data until we met our pre-determined sample size. Most of the experiments that follow were preregistered. Preregistration documents, materials, and data can be found at the Open Science Framework: osf.io/grxy8. 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
Study 1 consists of three sets of experiments (10 scenario studies) in which we test H1; that the manner in which preferences are elicited can lead to preference reversals. Specifically, we expect that preferences for the utilitarian products are higher when measured by WTP as compared to choices. The experiments in Study 1a document the basic effect and test for possible boundary conditions dependent on the mechanical design of the study (e.g., visual presentation of the stimuli). The experiments in Study 1b vary some of the product characteristics (e.g., average cost of the products) to test whether the effect generalizes to different types of products. Finally, the experiments in Study 1c vary features of the product-comparison (e.g., whether the products are substitutes for each other) to test whether any of those features affects the observed shifts in preferences.

Participants and Procedure

All experiments in Study 1 use the same basic design; participants either indicate their WTP for or make a choice between a pair of goods. In each scenario, one good is more hedonic (e.g., a chocolate bar) while the other is more utilitarian (e.g., trash bags). 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 Amazon’s Mechanical Turk platform. All participants were paid ~$0.12 to take part in the experiment and were randomly assigned to either a willingness-to-pay or a 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 to the question, “How much would you maximally pay for …[this good]?” into a text box. 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:” These participants could indicate a preference for either item or select an option indicating no preference between the two.

Study 1a

The first four experiments were conducted to demonstrate the basic effect and used relatively cheap, well-known product pairs such as ice cream and toilet paper. 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. We varied the latter because choices are usually presented in side-by-side orientation while elicitation of willingness to pay is typically elicited one product at a time, with the two items appearing vertically, one on top of the other. We were concerned, however that this presentation could affect a participant’s decision-making style, such that vertically presented products are evaluated separately, while horizontally presented products are more likely to be evaluated compared to each other (Hsee, 1996). We therefore included experiments in which the products were presented side-by-side both in the choice and WTP conditions.

Results & Discussion 1a

We calculated ordinal 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 for no study the results change in a meaningful way.

As can be seen in Table 1, 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, scenario 1.4, the preference for toilet paper increased by 55.4 percentage points from 36% to 91.4%. Notably, we find that this effect is robust to design-features of the experiment itself such as including or not including information about the regular retail price. Additionally, we exclude the possibility that the documented differences are merely a consequence of the presentation of the stimuli. In Study 1b we present three additional conceptual replications. In these experiments, rather than varying mechanical features of the experiment, we vary features of the products participants are evaluating, allowing us to further testing the degree to which these effects generalize.

*Study 1b, testing product features*

Study 1b consists of three additional scenario studies. In the experiments in Study 1a, 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. That is why, in the next three experiments, we tested whether the same effects emerge when we use products about which consumers do not have a lot of knowledge (a foreign bar of chocolate and a can opener in scenario 1.4), products which are substantially more expensive (a toaster oven and a fancy dinner
for two in scenario 1.5), and non-food products (a toaster and movie-tickets for two in scenario 1.6).

Results & Discussion 1b

Data were analyzed in the same way as for Study 1a. Despite considerably changing key-features of the products under evaluation, we find the same effect in these scenarios as in scenarios 1.1-1.4. When asked to indicate their WTP participants express a much stronger preference for the more utilitarian good as compared to when expressing their preference through choices. Results of scenarios 1.5-1.7 suggest that this type of preference-reversal is not limited to cheap, well-known products, but emerges across a large variety of different kinds of products.

Study 1c, testing for comparison features

In Study 1b we tested whether certain features of the products could moderate or eliminate the effect. In Study 1c we test whether the preference reversals documented in section 1a also generalize to decisions with different comparison features. 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, the other highly hedonic with very little functional benefits. Therefore, we use two products that are substitutes of each other in scenario 1.8. In scenarios 1.9 and 1.10 we also use substitutes, but in these scenarios we use products that vary in the degree to which they are hedonic and utilitarian within the same product category.

Results & Discussion 1c
The results for these scenarios are presented in Table 1, and are similar to those of scenarios 1.1-1.7; participants reliably exhibit a stronger preference in the WTP conditions as compared to the choice conditions. Crucially, in each of these scenarios, the items are close substitutes of each other. Because Johnson (1989) has demonstrated that evaluating non-comparable 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 non-comparable items. Given that the items we used in scenarios 1.8-1.10 are so readily substitutable, and in fact, in scenarios 1.9-1.10 are members of the exact same product category, the stimuli alone are unlikely to be driving the pattern of results that we consistently observe. Instead, we believe 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 decisions.
Table 1. “Study 1a participant information and study features.”

<table>
<thead>
<tr>
<th>Scenario</th>
<th>N</th>
<th>Preference Utilitarian Good (WTP)*</th>
<th>Preference Utilitarian Good (Choice)*</th>
<th>Significance Test*</th>
<th>$M_{Age}$ (SD)</th>
<th>Proportion Female</th>
<th>Study Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1†</td>
<td>300</td>
<td>86.2% (81/94)</td>
<td>27.7% (54/195)</td>
<td>$X^2 (1, N = 289) = 87.13$, $p &lt; .001$.</td>
<td>33.1 (10.6)</td>
<td>39.3%</td>
<td>Trash bags (MSRP: $11.99) and two pints of ice cream ($7.98). Counterbalanced order of stimuli in choice condition.</td>
</tr>
<tr>
<td>2†</td>
<td>200</td>
<td>25.6% (23/90)</td>
<td>12.5% (12/96)</td>
<td>$X^2 (1, N = 186) = 5.18$, $p = .02$</td>
<td>32.6 (10.7)</td>
<td>38.5%</td>
<td>Pen (MSRP: $2.33) and chocolate bar ($1.82).</td>
</tr>
<tr>
<td>3†</td>
<td>202</td>
<td>89.6% (86/96)</td>
<td>49.5% (48/97)</td>
<td>$X^2 (1, N = 193) = 36.55$, $p &lt; .001$</td>
<td>31.4 (9.4)</td>
<td>38.12%</td>
<td>12-pack of Cottonelle toilet paper and pint of ben and Jerry’s ice cream (No MSRP info).</td>
</tr>
<tr>
<td>4</td>
<td>203</td>
<td>91.4% (85/93)</td>
<td>36.0% (36/100)</td>
<td>$X^2 (1, N = 193) = 63.23$, $p &lt; .001$</td>
<td>32.7 (10.0)</td>
<td>37.93%</td>
<td>12-pack of Cottonelle toilet paper and pint of ben and Jerry’s ice cream. Direct replication of Scenario 3 (no MSRP info).</td>
</tr>
<tr>
<td>5</td>
<td>202</td>
<td>45.9% (39/85)</td>
<td>27.3% (27/99)</td>
<td>$X^2 (1, N = 184) = 6.89$, $p = .009$</td>
<td>37.8 (12.0)</td>
<td>54.95%</td>
<td>Toaster oven (MSRP: $64.99) and dinner for two (MSRP: $55) (more expensive items).</td>
</tr>
<tr>
<td>6</td>
<td>205</td>
<td>79.3% (73/92)</td>
<td>43.0% (43/100)</td>
<td>$X^2 (1, N =192) = 26.47$, $p &lt; .001$</td>
<td>36.7 (12.3)</td>
<td>54.15%</td>
<td>Toaster (MSRP: $39.99) and pair of movie tickets (MSRP: $35) (Non-consumable items).</td>
</tr>
<tr>
<td></td>
<td>202</td>
<td>85.87%</td>
<td>24.74%</td>
<td>( \chi^2 (1, N = 189) = 71.15, p &lt; .001 )</td>
<td>33.33% (10.4)</td>
<td>38.61%</td>
<td>Can opener and Dutch chocolate bar (items that participants were unlikely to have lay beliefs about cost; no MSRP info).</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>---------</td>
<td>--------</td>
<td>-----------------------------------------</td>
<td>----------------</td>
<td>--------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>202</td>
<td>81.3%</td>
<td>51.5%</td>
<td>( \chi^2 (1, N = 179) = 17.15, p &lt; .001 )</td>
<td>35.62% (12.1)</td>
<td>50.25%</td>
<td>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).</td>
</tr>
<tr>
<td>9</td>
<td>200</td>
<td>47.8%</td>
<td>32.3%</td>
<td>( \chi^2 (1, N = 186) = 4.65, p = .03 )</td>
<td>31.7% (9.1)</td>
<td>36.5%</td>
<td>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).</td>
</tr>
<tr>
<td>10</td>
<td>403</td>
<td>77.3%</td>
<td>51.0%</td>
<td>( \chi^2 (1, N = 368) = 27.29, p &lt; .001 )</td>
<td>35.7% (11.2)</td>
<td>52.36%</td>
<td>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).</td>
</tr>
</tbody>
</table>

† 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 Web Appendix for details.
STUDY 2A

The effects documented in Study 1 all occur in hypothetical situations. While this is common in many domains of consumer research, one can question whether the documented effects are a result of the hypothetical nature of the previous experiments, or whether the same effects emerge when people make choices involving real products and bid on these products using their own money (see for example, Ding et al 2007). Thus, we conducted two fully incentive compatible studies.

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 2 cells in the study.

Participants were randomly assigned to either indicate their WTP or choose between a 1oz. 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 in cash and given a brief primer on the Becker-DeGroot-Marschak procedure (Becker, DeGroot, & Marschak, 1964) for indicating their incentive-compatible WTP. After completing a comprehension-test, participants indicated their WTP independently for each product (so they could bid up to $1.00 for each product). To guarantee independent bids, participants were informed that only one of the two auctions would be played out for real, 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 $0 - $1. If the participant’s WTP was greater than or equal to the rolled price, we sold the item to the participant at that
price, and returned their change from the initial $ endowment. If the participant’s WTP was less than the rolled price, the participant kept the dollar. Participants in the choice condition simply chose which item they preferred to have, and were given that item.

Results

Upon excluding participants who were indifferent in the WTP condition (those who indicated the same price for either item; 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 significant, $X^2 (1, N = 283) = 7.22, p = .007$.

STUDY 2B

While Study 2a demonstrates preference reversals in an incentive compatible way, we designed an additional study to more conclusively demonstrate the effect in a setting where participants’ choices were consequential. Because it is known that 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 & 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 only describe 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.
Experiment 2b was conducted in person on a central plaza of a large American university. To be able to control the size of the experienced windfall, we conducted this study in 2 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 does result in a more complicated design, it allows us to more directly equate the utility that participants are receiving in the two conditions enabling a more conservative test of the effect.¹

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.

Wave 1 – WTP: Participants and Procedure

Participants were 306 individuals (51.2% female $M_{age} = 22.2$, $SD = 8.6$) approached on campus of a large American university in April 2016 by research assistants. Participants were immediately given $2 in cash upon agreeing to complete the study and were informed that the $2 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 BDM² 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

¹ Note that, like Study 2a, the design of this study was completely pre-registered.
² Bids were censored to a maximum of $2; 35 participants bid $2 for chocolate, 66 for the toothpaste. If censoring had any effects they would have worked against our hypothesis since it affected bids for toothpaste more than for chocolate.
Complete Whitening + Scope Toothpaste. The participants were further told that one of the items would be randomly selected and have the auction actually be played out (i.e., the participants’ bid would be compared to a randomly generated price and the auction would be enforced). Participants then indicated their WTP using the BDM procedure for both goods (counterbalanced). Upon completing both rounds, one of the two products was selected randomly and the outcome for that product was enforced.

Wave 2 – Choice: Participants & 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 containing Toblerone chocolate plus some money, the other 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, by including money in the bundles, the tradeoff between money and products was 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, we aimed to keep the average experienced windfall at $2 (as in Wave 1). To achieve this, these participants chose between [the chocolate bar + $0.82] and [the toothpaste + $0.66]. For the other half of the participants in wave 2, we kept the windfall below $2 (a lower windfall than that experienced in Wave 1), to achieve this, these participants chose between [the chocolate bar + $0.27] and [the toothpaste + $0.11].

---

3 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.
If the tendency for participants to choose relatively more hedonic goods in choice condition is purely the result of windfall effects, we would expect preferences for chocolate to be lowest in the low 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.

Results

As planned, we excluded data from 13 participants who indicated a nut allergy. This left 146 participants in Wave 2a and 147 participants in Wave 2b, for a total of 293 participants. We compare choices between the bundles in Wave 2, with extrapolated WTP in Wave 1. Overall, 174 participants preferred a bundle containing chocolate (59.4% vs 47.6% in Wave 1), whereas 119 participants preferred a bundle containing toothpaste (40.6% vs 52.4% in Wave 1). This is significantly different from the distribution of preferences in Wave 1, \(X^2\) \((1, N = 564) = 7.86, p = .005\). 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, \(X^2\) \((1, N = 293) = .299, p = .585\). Comparing Wave 2a separately to Wave 1, we found that 89 participants preferred chocolate (61%) whereas 57 (39%) participants 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 more conservative test, yields 85 preferring chocolate (57.8%) and 62 participants preferring toothpaste (42.2%). 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.

Discussion

Studies 1a-1c and Studies 2a-2b clearly reveal that using choice as a preference elicitation procedure leads to a greater preference for a hedonic good than in a WTP setting. 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. We believe the tendency to prefer hedonic goods in choice is, at least partially, 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 3.

STUDY 3

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 among 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, having one’s feelings inform the decision is a sensible and cognitively cheap heuristic which 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 willingness-to-pay 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 to look more like their behavior in the WTP condition. This moderating role of deliberation about value is tested in Study 3.

Participants and Procedure

Participants were 402 workers (43.5% female $M_{age} = 32.8, SD = 10.3$) recruited through Amazon’s Mechanical Turk Platform. Participants were randomly assigned to one of four conditions in this 2 (WTP vs. choice) X 2 (control vs. deliberation) design. Participants in the WTP conditions were presented the same items as used in scenario 1.1 of Study 1 (ice cream and trash bags), and were asked to indicate their WTP for each good, while the other half was asked to make a choice 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 only indicated WTP. Likewise, 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.

Results
First, we calculated ordinal preferences in the WTP conditions by comparing the indicated WTP for the ice cream with the indicated WTP the trash bags. Upon excluding participants who indicated indifference, 387 responses remained.

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 vs. choice, deliberation condition, and the interaction between these terms. Consistent with H1, there is a main effect of whether participants are asked to indicate their willingness to pay or to directly choose, such that those who choose directly are more likely to select ice cream than those who indicate their willingness to pay, $z = 7.76, p < .001$ (see Figure 3). Moreover, the interaction between deliberation and elicitation method was marginally significant; $z = 1.89, p = .059$. Probing this interaction, we find that within the WTP conditions, participants did not differ in their preferences with 81% (81/100) of participants in the non-deliberation condition and 82.9% (78/94) in the deliberation condition choosing the trash bags $X^2 (1, N = 387) = .13, p = .72)$. 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) $X^2 (1, N = 193) = 11.39, p = .001$.

Figure 3: Preference for utilitarian good by elicitation method and deliberation, Study 3.
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 3 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. Besides Study 3, we conducted several additional studies using different methods of deliberation ranging from subtle to conspicuous, and while these studies often showed an attenuation of the preference reversal affect, the difference was never fully attenuated. If the difference between choice and WTP were caused only by a reliance on an affect-based heuristic, deliberation should completely eliminate, rather than merely attenuate the effect. Therefore, we investigated the additional processes driving the differences between WTP and choices in Study 4 and 5.
STUDY 4

To this point, we have consistently found that participants’ preference for hedonic relative to utilitarian items is dependent on how their preferences are elicited, despite normative economic theory suggesting that preference reversals should not occur. Yet, these preference reversals have not only repeatedly surfaced, they also appear quite resilient to both varying the context and to explicit attempts to reduce them. For instance, requiring participants to more deeply engage with their decisions only reduced the effect by ~33-50%. This suggests that, besides deliberation, there are other processes at work. In other words, the documented preference reversals appear to be multiply determined. To identify a mechanism beyond elaboration, we employ a method previously used by Frisch (1983) to see if participants view choice and WTP as non-equivalent processes. More specifically, we presented participants with willingness-to-pay and choice elicitation procedures over the same goods. Then, any participant who gave inconsistent answers about their preferences (i.e., chose a different good than the one they indicated a higher WTP for) was prompted to explain why they approached the two situations differently.

This approach allows us to tease apart the degree to which these preference reversal effects are caused by a different interpretation or understanding of the question that is posed to the participants in each elicitation procedure (Sher & McKenzie, 2006), or by the different psychological responses evinced by each procedure (Keys & Schwarz 2007). To tease apart whether the differences in preferences between choices and WTP occur because of an artifact of the design (e.g., participants thought WTP was asking only for the market price) as opposed to participants experiencing the two measures as meaningfully different psychological experiences,
we had participants complete both versions of the preference elicitation tasks, as well as both versions of three other “classic” framing effect problems. We then explained each inconsistency to participants, and allowed them to provide insight into their decision-making processes. We included the three classic framing-effects to provide a benchmark to interpret the proportion of participants who disagree with the explanation that the choice options are equivalent.

Participants and Procedure

Participants were 237 workers (47% female $M_{age} = 34.1$, $SD = 10.4$) recruited through Amazon’s Mechanical Turk Platform. Workers completed a number of framing effect problems in two blocks. The first block included the typical WTP task (prices indicated were for ice cream and trash bags, with MSRP information), the first half of the “Asian disease” problem, which we call an “unusual disease” in the study and the first halves of two sunk-cost problems (lost concert ticket and tennis elbow, Frisch, 1983; Kahneman & Tversky, 1984). The order of these tasks was randomized within the block. The second block included the choice task, the second half of the “Asian disease” problem, and the second halves of two sunk-cost problems, with the order randomized. Next, participants were shown explanations for why the options in each scenario are equivalent (e.g., in the “Asian disease” problem, it was explained that saving 200 out of 600 people means 400 of the 600 will die). See the materials for each problem, as well as the explanations for why the outcomes are equivalent.) Participants were then asked to indicate whether they agreed with each of these explanations, and to provide a rationale for their beliefs.

Results

Restricting our analysis to the 193 participants who did not indicate indifference in either choice or WTP, 31.1% (60/193) of participants preferred ice cream, while in the choice

---

4 We planned and preregistered to run 200 participants but due to an error with the completion code we ended up with more than planned.
condition, 76.7% (148/193) of participants preferred ice cream. This difference is significant, (McNemar’s $X^2 (1, N = 386) = 27.61, p < .001$. For the other framing effect problems, see Table 2.

For each of the three “classic” framing effects, we find that after providing an explanation as to why the typical framing effect pattern of results is inconsistent, participants largely agree with the explanation. For the “Asian disease” problem, 67.1% (159/237) of participants agreed with the normative explanation, for the concert ticket sunk-cost problem, 77.6% (184/237) of participants agreed with the normative explanation, and for the tennis elbow sunk-cost problem, 65% (154/237) of participants agree with the normative explanation. For the choice and WTP preference reversals, a significantly smaller proportion of participants, 39.7% (94/237), agreed with the normative explanation. Restricting the analysis only to those who demonstrated an inconsistent pattern of preferences in the WTP and choice conditions, only 31.6% (36/114) agreed with the explanation that the two elicitation procedures should yield identical preferences.

In addition to merely indicating whether or not they believe the two situations equivalent, we asked each participant to explain why they did or didn’t agree with the normative explanation that choice and WTP are equivalent. The responses are striking. A research assistant at a different university, who was blind to condition, flagged responses that indicate that WTP and choice are different because they “feel” different or as if they are tapping into different processes (e.g., wants vs. needs). This led to 53.85% (77/143) of responses being flagged as indicating a psychologically meaningful difference between choice and WTP. One participant wrote, “…ice cream has an immediate satisfaction factor and this influences the decision.” and still

---

5 Many of the other responses were difficult to interpret psychologically, such as “It just feels different” or “I don’t believe economists”.

another neatly captures the idea that WTP and choice are measuring different types of preferences, “I don't agree. I also think your willingness to pay for something vs. your choice is situational. I may need trash bags at the moment but my choice of what I want may be different.”

Table 2. “Proportion of inconsistent responses and agreements with equivalent explanations for each framing effect in Study 5.”

<table>
<thead>
<tr>
<th></th>
<th>Choice vs. WTP</th>
<th>Asian disease</th>
<th>Lost ticket</th>
<th>Tennis elbow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inconsistent</td>
<td>57.07%</td>
<td>32.91%</td>
<td>14.35%</td>
<td>32.07%</td>
</tr>
<tr>
<td>Agree with normative exp.</td>
<td>39.66%*</td>
<td>67.09%</td>
<td>78%</td>
<td>64.98%</td>
</tr>
<tr>
<td>Agree within inconsistent</td>
<td>31.58%</td>
<td>50.00%</td>
<td>44.12%</td>
<td>39.47%</td>
</tr>
</tbody>
</table>

*Note: The denominator here is the full sample 237 (that is including indifferent participants).

**Discussion**

In numerous problems where supposedly equivalent options yield different preferences, participants can respond in three different ways to an explanation of why these problems should be interpreted as equivalent. First, participants can acknowledge that the two problems are equivalent, and agree with the normative explanation. The second possibility is that participants can argue that the descriptions of the problems convey different information, even when objectively they do not literally do so. The third possibility is that participants can disagree, but argue that the explanations are not equivalent because the problems feel psychologically different. A review of the qualitative responses given by our participants reveals that the difference between choice and WTP clearly falls in the third category; participants indicate that
WTP and choices feel very different or explain how they focus on needs in WTP and wants in choices. If this is true, then we should be able to create similar preference-reversals merely by making the elicitation procedure feel more or less like an exchange of money. This is what we tested in Study 5a and Study 5b.

**STUDY 5A**

Study 4 provided initial evidence for the notion that the WTP and choice procedures differ in a psychologically meaningful way. A significant proportion of participants indicated that choice and WTP “feel different” or expressed a sense that when money is being considered, it should be used for more useful purposes. We developed Study 5a to test whether these differences can be manipulated such that the monetary exchange in a transaction is made more salient. We predicted that increasing the salience of monetary exchange in a choice condition would lead consumers to exhibit an increased preference for utilitarian items.

**Participants and Procedure**

Participants were 402 workers recruited from MTurk (57.36% female, \(M_{age} = 32.9, SD = 12.4\)). Participants were randomly assigned to one of two conditions (choice vs. monetized choice). All participants were told to imagine that they had been recruited to read a nonsensical excerpt of text and score the text based on three rules pertaining to the location of the letter “e” throughout the piece. Next, participants were asked to imagine that they would receive some compensation for completing the text-scoring task. Participants in both conditions could then choose one of ten randomly displayed items as compensation: a windup toy, a stress squeeze
brain, Toblerone chocolate, travel size Band-Aids, Avery star stickers, Chap Stick, Hefty quart storage bags, tennis balls, Puffs Facial Tissues, or Crest Toothpaste.

Participants in the monetized choice condition were told to imagine that they had been given a $2 bill (a picture was included), which they could use to buy one of the 10 gifts from the experimenter. Participants in the choice were shown a picture of a gift box and asked to imagine that they could choose one of the 10 items as a gift from the experimenter. Several other exploratory measures were included, but are not relevant to the primary goals of this study, and are described in the Web Appendix.

Results

Prior to collecting data, we believed (and pre-registered this belief) that the windup toy, stress squeeze brain, Toblerone chocolate, Avery star stickers, and tennis balls to be the more hedonic set of items while we believed the travel size Band-Aids, Chap Stick, Hefty quart storage bags, Puffs Facial Tissues and Crest Toothpaste to be the more utilitarian set of items. Accordingly, we coded anyone who selected one of the five items in the hedonic set as preferring a hedonic item and anyone who selected one of the five items in the utilitarian set as preferring a utilitarian item. We then conducted a chi-square analysis to determine if the preference for a hedonic item was dependent on the condition assignment. We found that participants assigned to the choice condition chose hedonic items at a significantly greater rate, 58% (116/200), than participants in the monetized choice condition, 44.6%, (112/202), \( X^2 (1, N = 402) = 7.27 \ p = .007 \).

STUDY 5B

We designed Study 5b as a complement to Study 5a. Whereas in Study 5a, we manipulated the design of the choice elicitation procedure to make the salience of monetary
exchange more prominent, thereby increasing preference for utilitarian items, we attempted to do the opposite in a WTP context. Specifically, we designed a WTP condition in which participants imagined exchanging tokens with a set monetary value for different goods, and compared this to the typical WTP scenario. In this case, we expected the tokens would be a level removed from money, and would lead more participants to indicate a higher WTP for the hedonic good than for the utilitarian good. In developing these studies we draw on literature that suggests non-cash transactions reduce the “pain of paying” means of payment such as credit cards are a degree separated from cash (Prelec & Loewenstein, 1998); other research has found that people treat tokens less like money when engaging in dishonest behavior (Mazar et al, 2008). We believe requiring participants to indicate WTP using tokens instead of money will reduce the salience of monetary exchange, leading to a greater preference for hedonic goods.

Participants and Procedure

Participants were 403 workers recruited from MTurk (46.9% female $M_{age} = 36.97, SD = 24.01$). Participants were randomly assigned to one of two WTP conditions (money vs. tokens). All participants were told to imagine that they had been recruited to complete the same boring psychology task as in Study 5a. Participants were told that the researcher would allow them to exchange some currency for a small prize as compensation for completing the study. Half the participants were told to imagine that they were given 40 tokens that they could trade for a gift, while the other half were told to imagine that they had received $4.00 in cash that they could use to purchase a good. All participants completed the BDM procedure, which included a practice round, and then bid on both a tube of Crest toothpaste (utilitarian) and a Toblerone chocolate bar (hedonic). As expected, participants in the “token” condition were more likely to indicate a
higher WTP for the hedonic item (44.64%, 75/168) than were participants in the money condition (26.78%, 49/183) $\chi^2 (1, N = 351) = 12.24, p < .001$.

Discussion

Study 5a and Study 5b together show that, even within a given preference elicitation procedure, it is possible to manipulate the psychological frame associated with that procedure, and thus manipulate participants’ preferences. Specifically, while we have repeatedly found that a choice condition is associated with a relatively greater preference for hedonic goods than a WTP condition, even the superficial manipulation of the salience of monetary exchange can alter the degree to which participants exhibit these relative patterns of preferring hedonic and utilitarian goods. We believe the manipulations we used in these studies are relatively weak forms of creating a different psychological context, as they do not fundamentally change the elicitation procedure, but rather merely change the labels used. Yet even these subtle manipulations appear to alter the salience of monetary exchange and change the items our participants prefer. Taken together with the results of Study 4, which show that participants believe that choice and WTP represent non-equivalent psychological contexts, the results of Study 5a and Study 5b show that the context of a preference elicitation procedure is both malleable and relevant in determining how preferences are expressed.

GENERAL DISCUSSION

In a set of 16 experiments, we repeatedly 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 preference, such as WTP appears to lead to a relatively greater preference for a utilitarian, than a hedonic good, in comparison to the use of a
more affective measure of preference such as choices. We document this phenomenon in online scenario experiments (Study 1) and in fully incentive-compatible field sample studies (Studies 2a-2b). The results of Study 2b are perhaps more striking because the bundled options that include both product and cash in the choice condition make the tradeoff of money for products and goods more salient in the choice condition, yet the effect persists. We find that deliberation attenuates this effect (Study 3), indicating that giving more attentive thought to the decision under consideration leads preferences in a choice condition to more closely resemble the WTP condition. We also find that this effect is likely multiply determined, as participants believe, contrary to the precepts of procedure invariance, that WTP and choice are not equivalent, but appear to be psychologically different (Study 4). Finally, building off the text responses we analyzed in Study 4, we designed Study 5a and Study 5b to show that an elicitation procedure can be framed in such a way as to increase preferences for a class of goods framing the procedure to match the underlying hedonic or utilitarian nature of the goods under consideration.

One can wonder to which 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 in this manuscript; either market value is incorporated more strongly in preferences expressed through willingness to pay as compared to choices, or our participants must have 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, 2b, 4, 5a, and 5b. Specifically, when participants paid the goods with their own money and when they were confronted with the outcomes of their answers, we still found evidence for preference reversals. Moreover, in Study 4 we asked participants why they believed preferences through WTP and choices could differ from each other and no participant indicated misunderstanding the WTP
question as a question about market value. A way in which perceived market value could play a role is if participants believe it is more appropriate to incorporate market value in WTP than it is in choices. This description 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 preference. Market value, as well as other attributes such as quantity, perceived quality, and opportunity costs, are 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. Taken together, we see a robust effect that violates the tenets of rational choice theory, and which are unique from extant instances of preference reversals.

We believe our work is novel because we have documented a set of circumstances under which participants’ expressed preferences cannot be reconciled with a single, well-defined and rank-ordered set of preferences. We do not claim to be the first authors to demonstrate preference reversals, and it is unlikely to be a controversial position to argue that preferences are labile. Where we believe our work differs, however, is that unlike many other situations in which researchers find evidence for constructed preferences or preference reversals, such as typical framing effects or joint and separate evaluations (Hsee, 1996; Hsee & LeClerc, 1998; Shaffer & Arkes, 2009), we did not design our studies specifically to demonstrate preference reversals or create situations that participants and consumers are unlikely to encounter in their day-to-day experiences. For this reason, we believe our results are so surprising: we are merely comparing two commonly-used, gold standard, methods of preference elicitation and use products that most consumers are very familiar with and find large differences in expressed preferences.
Implications for Understanding Preferences

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. 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 – 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.

It is also possible to question 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 (like 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.
Relation to Other Instances of Preference Reversals and Constructed Preferences

While we (to our best knowledge) are the first to show preference reversals for consumer goods based on choices vs. WTP, there is of course a large body of work on preference reversals in different contexts. First, our findings clearly relate to, and are consistent with, the work by Lichtenstein and Slovic (1971) on elicitation effects for decisions under uncertainty. More specifically, Lichtenstein and Slovic demonstrated a preference for gambles with low probabilities of winning high payouts ($ gambles) when participants indicate WTP, but a preference for higher probability, lower payout gambles ($ gambles) when choosing between them. Tversky et al (1990) developed a clear and intuitive explanation for these findings, scale compatibility, in which participants over-price the high-payoff low-probability gambles because they anchor on the higher payouts of these gambles and indicate higher values for WTP. Critically, our findings fall outside the scope of these explanations. 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 should not account for the observed preference reversals. However, out of an abundance of caution, we designed a study to test whether scale issues cause the effects we find. We found that they do not, and report this study in the Web Appendix. Another crucial difference between the studies presented in this paper, 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, neither outcome is subject to any uncertainty.
There are also some similarities between our studies, in which participants either price or choose between products, and the work on joint vs. separate judgments (Hsee, 1996). Specifically, in work on joint vs. separate evaluations participants either price or evaluate one of two different products or both products at the same time. Preference reversals are claimed to 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 process underlying joint vs separate preference reversals cannot explain the effects documented in this manuscript. 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, MSRP (the easiest to compare attribute) should weigh more heavily in choices as compared to WTP. Instead, we find a preference for the hedonic good in choices regardless of whether its MSRP is displayed or not.

Finally, Nowlis & Simonson (1997) have documented a different type of preference reversal that at first glance seems to conflict with the effects documented here. They find that more 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.
Two differences between our studies and those by Nowlis & Simonson could explain why our results appear to conflict with theirs. First, Nowlis & Simonson do not ask participants to merely make a choice between two goods, but instead ask participants if they had to buy one of the two goods which one they would prefer buying, meaning that participants did not just make a trade-off between the two goods, but also had to take into account the costs of the different goods. Additionally, in our studies we allowed participants to express indifference whereas Nowlis & Simonson used a forced choice in the choice conditions and dropped participants who expressed indifference in the rating conditions. This creates 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 contradicting findings. Therefore we attempted to replicate Nowlis & Simonson’s findings, both using the original materials, and adding a new condition where participants could indicate indifference in the choice condition (see the web appendix for full write-up). Instead of finding a preference for the non-enriched off-brand toaster in the choice condition, we find the opposite; the choice condition is associated with a marginally higher preference for the brand-name item (30.52% vs. 26.25%), $X^2 (1, N = 1269) = 2.85, p = .09$.

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 (c.f. Moon and Nelson (2017) who have found a dissociation between WTP and anticipated enjoyment in decision-making under uncertainty. We think it worth reiterating the advice offered by Frederick & 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 as we continued we realized that a crucial, but not the sole 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 time-course 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 choosing between different options, where consumers consider various attributed 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), in other cases they clearly make a choice like 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.

In closing, we turn to the analogy of the baseball umpire that has frequently been used to describe preferences. While the analogy leads to three dispositions about how preferences are formed, (“I call them as I see them”, “I call them as they are”, or “They ain’t nothing ‘till I call them”), we suggest that, at least, when it comes to different elicitation methods, the evidence
points more to an umpire officiating different games, rather than one game with multiple perspectives. We posit that choice and willingness-to-pay are more like baseball and cricket, and less like different perspectives on the same ballgame.
REFERENCES


Moon, Alice and Nelson, Leif D., The Uncertain Value of Uncertainty: When Consumers are Unwilling to Pay for What They Like (March 1, 2017). Available at SSRN: https://ssrn.com/abstract=2676699 or http://dx.doi.org/10.2139/ssrn.2676699


