Consumers Undervalue Multi-Option Alternatives

Stephanie M. Smith* [a]
Stephen A. Spiller [b]

[a] Booth School of Business, University of Chicago, Chicago, IL 60637
[b] Anderson School of Management, UCLA, Los Angeles, CA 90095

*Stephanie M. Smith
5807 S Woodlawn Ave
Chicago, IL 60637
stephanie.smith@chicagobooth.edu

ABSTRACT

Initial decisions lead to subsequent decisions. Dominated options in such downstream choices ought to be ignorable in the initial choice for even minimally forward-looking people. Across eleven studies in two domains (consumer goods and risky gambles), we find that adding a less-valuable option (i.e., a less-preferred consumer good or a dominated gamble) decreases the choice share of an otherwise attractive alternative. This difference is moderated by the value difference between the more- and less-valuable options. Mouse-tracking reveals that participants who attend more to the dominated option are less likely to choose the multi-option alternative. This work contributes to our understanding of multi-option decision-making and how decision makers assess the overall value of choices.
Keywords: judgment and decision making, consumer behavior, mouse-tracking

Only rarely does a person's full decision process conclude at the moment of choice. Instead, each node in a decision tree typically leads to more decisions. Sometimes these are implicit: choosing a home implies choices among where to eat, where to shop, and who to visit. Other times these are explicit: choosing a restaurant implies choosing items from a menu and choosing to watch TV implies choosing a show.

Even this narrower case of explicit multi-option alternatives is ubiquitous. Multi-retailer gift cards constitute multi-option alternatives, since any part of the balance that is spent at a single retailer cannot be spent at another. Airline choices also constitute multi-option alternatives, as a given airline might have multiple routes available from origin A to destination B. Food and drink tickets, common at festivals and conferences, are typically multi-option alternatives, as are game tokens at arcades which can be used for a variety of games. Ultimately, many multi-attribute choices can be—and are—characterized as multi-option alternatives. For instance, purchasing a car includes multi-option alternative(s): a consumer might decide on a Ford Mustang, only to be faced with the choice of blue vs. yellow.

The examples above help to give the intuition for the structure of choices that we are interested in. Of course, with each of these examples, there might be many other considerations at play: whether other people are involved in the decision process, familiarity with the menu/options, past decision history, complementarity vs. substitutability of the sub-options, etc. In this paper, we remove these extra considerations to better understand how people integrate across options in a choice set when deciding among multi-option alternatives. We find they regularly sacrifice a chance at maximizing utility by integrating the value of less-attractive options.

Value-Maximizing Decision Rules in Multi-Option Choices

To a value-maximizing individual, a multi-option alternative is worth at least as much as its most-valuable option, because they can ignore the less-valuable option, knowing that they will choose the more-valuable option instead. Formally, we consider the case in which an individual decides between a single option S and a multi-option alternative M, where M is the choice set $\{M_H,M_L\}$ (i.e., the choice between M_H and M_L in which M_H, "multi-high," is preferred to M_L, "multi-low"). If there is no uncertainty regarding the values of M_H or M_L, then the value of M is $\max\{M_H,M_L\}$. Since M_H is preferred to M_L, $\max\{M_H,M_L\} = M_H$. Thus, the choice between S and M simplifies to the choice between S and M_H. S should be no more likely to be chosen when pitted against M than when pitted against M_H. If there are some states of the world in which M_L is preferred to M_H, S may be less likely to be chosen when pitted against M than when pitted against M_H. (For example, a consumer may generally prefer hot coffee (M_H) to iced coffee (M_L), but make an exception when the temperature exceeds 90° F.[1]) In other words, adding M_L as an option to M_H should not decrease the probability of choosing M.

Relatedly, though the analysis above suggests that decision makers should value a multi-option alternative equivalently to the expected maximum of its constituent pieces, some findings from the behavioral literature suggest that people might value a multi-option alternative more than the maximum value of its component options because the multi-option alternative enables choice (Brehm 1966; Bown, Read, and Summers 2003; Mochon 2013; Shin and Ariely 2004). Because people often value the ability to choose, the added value from the presence of choice

[1] Throughout, we consider cases in which the state of the world is exogenously determined: neither the choice set nor the choice affects the preference ordering among S, M_H, and M_L.

itself can lead to an overvaluation of multi-option alternatives.

Valuation of Sets

Alternatively, people might undervalue a multi-option alternative. If valuation processes for multi-option alternatives and bundles are similar, then people would undervalue multi-option alternatives because bundles of goods are valued according to a weighted average of their components' values (Brough and Chernev 2012; Chernev and Gal 2010; Gaeth, Levin, Chakraborty, and Levin 1991; Yadav 1994; Shenhav and Karmarkar 2019). There is related evidence that consumers estimate the value of a product to be a weighted average of its features (Weaver, Garcia, and Schwarz 2012; Troutman and Shanteau 1976), and adding a less-attractive bonus to a product decreases its value (Simonson, Carmon, and O'Curry 1994). These "weighted averages" are not mathematical averages of numerical properties, but rather central tendencies of subjective values. People can rapidly extract average economic value from a set of products (Yamanashi Leib et al. 2020) and sometimes use these assessments inappropriately (Frederick and Kahneman 2002; Kahneman 2003). These findings suggest that adding a less-desirable option to a multi-option alternative decreases its value in the same way that adding a lower-valued component to a bundle decreases the bundle's value: the worse the option, the greater the decrease.

Prior Research on Multi-Option Alternatives

Therefore, there are two hypotheses regarding how people might misvalue multi-option alternatives relative to the value-maximizing benchmark. Prior research on choice sets and assortments can help to inform the relative relevance of those prior literatures. Several papers have investigated various aspects of assortment choice (Kahn and Lehman 1991; Sood, Rottenstreich, and Brenner 2004; Karmarkar 2017; Friedman, Savary, and Dhar 2018; Bogard, Reiff, Caruso, and Hershfield, 2024). These papers focus on how assortments compare to one another and how assortment framing (as sets vs. alternatives) affects valuation. Therefore, this literature tends not to focus on how the value of a multi-option alternative compares to its highest-valued option. To our knowledge, only two prior papers have directly examined this latter phenomenon (Le Lec and Tarroux 2020; Spiller and Ariely 2020). However, these papers leave many questions unanswered.

Le Lec and Tarroux (2020) tested the phenomenon in a single study in a domain in which the relative values of products are subjective. In this study, participants reported their willingness-to-pay (WTP) for multi-option alternatives regarding what websites to spend time on at the end of the experiment. The authors found evidence for undervaluation of multi-option alternatives, and proposed two explanations: anticipation of future error and holistic evaluation. Spiller and Ariely (2020) focused entirely on a subset of multi-option alternatives: media of exchange (e.g., gift cards and promotional credit). Much like Le Lec and Tarroux (2020), this paper used a subjective domain and focused primarily on WTP judgments (though some supplemental studies examined choice). In both cases, the authors found the extent of undervaluation increases with the difference in value between options, consistent with a weighted averaging process.

How are these results reconciled with the findings regarding the inherent value of choice? An important feature of the options-increase-value literature cited above is that a consumer’s focus is primarily on the presence/absence of options rather than on the choice set as a singular entity. Thus, based on the most similar prior literature (e.g., Le Lec and Tarroux 2020; Spiller and Ariely 2020) and the arguments above, we propose and test three hypotheses:

H1: People are less likely to choose a multi-option alternative than they are to choose its higher-valued option over another fixed alternative.

H2: This reduction in choice share increases with the difference in value between the two component options in the multi-option alternative.

H3: This reduction in choice share is attributable to a reduction in the holistic value of the multi-option alternative (vs. the value of the more-attractive option).

This paper contributes to the literature in multiple ways. First, in contrast to Le Lec and Tarroux (2020) and the primary results of Spiller and Ariely (2020), we use choices to document the effect. Such choices are arguably more common in consumers' daily life than WTP ratings and sometimes exhibit qualitatively different effects (Grether and Plott 1979; Lichtenstein and Slovic 1971; Amir and Ariely 2007; O'Donnell and Evers 2019), perhaps because WTP ratings reflect market prices rather than personal preferences (Evangelidis, Jung, and Moon 2022). In the current case, this is important to examine as choices may encourage consumers to peer down the decision tree to possible outcomes whereas WTP judgments may encourage them to value a choice set holistically. Second, we document the robustness of the effect across eleven online lab studies (five in the main text, six in the supplement). These eleven studies span two domains, including one domain with objectively dominating/dominated component options, expanding the findings beyond preferential choices in which uncertainty over future preference states may play a larger role. Third, we rule out several possible explanations with additional studies and features (e.g., trinary choices). Fourth, we examine the potential for transparency of dominance relationships to moderate the effect. Fifth, we document the effect as the number of component options in a multi-option alternative increases. Sixth, we use process tracing to connect the information acquisition process to choices. Seventh, we provide evidence of mechanism: adding an inferior sub-option decreases the value of the holistic multi-option alternative, rather than just the superior sub-option. Finally, we use data on heterogeneous preferences to consider the consequences of undervaluation for aggregate effects with different levels of targeting precision.

Process Tracing

Our process tracing (mouse-tracking) substantively contributes to our understanding of the decision process and connects to other work in neuroeconomics. There are multiple reasons why people might under- or overvalue multi-option alternatives; a possible mechanism for these shifts in valuation (and choices) is attention. In binary choices, people tend to choose the option they have looked at longer (Armel, Beaumel, and Rangel 2008; Atalay, Bodur, and Rasolofoarison 2012; Chandon et al. 2009; Fiedler and Glöckner 2012; Fisher 2021; Krajbich, Armel, and Rangel 2010; Lim, O'Doherty, and Rangel 2011; Mormann et al. 2012; Newell and Le Pelley 2018; Pachur et al. 2018; Pärnamets et al. 2015; Pieters and Warlop 1999; Vaidya and Fellows 2015). Moreover, people tend to update their values and choose in line with the choice attribute that they focus on (Bhui and Jiao, 2023; Busemeyer and Townsend 1993; Busemeyer and Diederich 2002; Roe, Busemeyer, and Townsend 2001; Dai and Busemeyer 2014; Kim, Seligman, and Kable 2012; Fisher 2017; Cohen, Kang, and Liese 2017; Smith and Krajbich, 2018), including in situations with more than two alternatives (Krajbich et al. 2011; Gluth et al. 2020; Thomas, Molter, and Krajbich 2020). Use of mouse-tracking and information-search paradigms has successfully connected information acquisition and attentional patterns to choice (Payne 1976; Johnson et al. 1989; Payne, Bettman, and Johnson 1988; Konovalov and Krajbich 2020; Stillman, Krajbich, and Ferguson 2020; Reeck, Wall, and Johnson 2017; Stillman, Shen, and Ferguson 2018; Diehl 2005).

This body of literature suggests that the attention processes in this environment may inform the decision processes involved in the undervaluation of multi-option alternatives. Specifically, we propose and test the following hypothesis:

> **H4:** Undervaluation is negatively correlated with the difference between: (i) the relative time spent on the multi-option alternative, and (ii) the relative time spent on its highest-valued option in simple binary choice.

Measuring Undervaluation

Though the concept of undervaluation is straightforward, finding an efficient, effective way to measure it within-subjects is decidedly less so. Eliciting WTPs of multi-option alternatives would be simpler, but ratings are less-than-ideal for several reasons. First, they are not as common in the real world as choices; although they provide insight about value, we often care about that value because of how it informs choices. Second, there are well-documented cases in which ratings or WTP judgments deviate from choices (e.g., Grether and Plott 1979; Lichtenstein and Slovic 1971; Amir and Ariely 2007; O'Donnell and Evers 2019). Therefore, we measure undervaluation within-subject in a series of carefully constructed choices.

In each of our studies, we examine undervaluation by comparing choices in two types of decisions: test decisions and binary control decisions. In test decisions, participants choose between a single-option alternative (S) and a multi-option alternative (M). Within the multi-option alternative (M), there are two component options: M_H and M_L, between which the participant could make a future choice. In all cases, value(M_H) > value(M_L). In the binary control decisions, participants choose between the same single-option alternative (S) and the a priori determined higher-valued component option (M_H). Undervaluation occurs when the proportion of S choices is greater in test decisions than in control decisions. In other words, when participants choose M less often than they choose M_H, this indicates that participants undervalue M relative to its best component option (M_H).

Across five studies, we document the undervaluation phenomenon and investigate the underlying mechanism. In Study 1, we find evidence of undervaluation (H1) in a consumer domain with subjective values, and find stronger undervaluation as the difference between M_H and M_L increases (H2). In Study 2, we provide evidence for the mechanism underlying our effect (H3). In Study 3, we demonstrate the consistency of the effect as the number of sub-options within an alternative increases. In Study 4, we move into a more controlled domain (incentivized gambles) with objective dominance of M_H over M_L. In Study 5, we use mouse tracking in order to better understand the relationship between information acquisition and the undervaluation phenomenon (H4).

In the supplements, we describe six additional studies in which we (1) replicate the main results, (2) rule out possible alternative explanations (e.g., inattention), and (3) examine the relationships between undervaluation and several individual difference measures (e.g., risk aversion). For a comparison of the methods across all studies, see table S1.

STUDY 1

In Study 1, we tested for undervaluation of multi-option alternatives in a relevant consumer domain. We also investigated a potential moderator of the effect. Specifically, we investigated whether the subjective value difference between the best and worst component options is related to undervaluation, consistent with an averaging process.

Method

Participants. For this preregistered study (https://researchbox.org/124&PEER_REVIEW_passcode=WUXZVT), we collected responses from 305 Amazon Mechanical Turk (AMT) workers. They earned $1.25 for their participation.

Materials and Procedure. First, participants rated 50 well-known films on a scale of 0-10 (measured in increments of 0.1 via a slider) regarding how much they wanted to watch each of them. For each film, participants had the option to click a box labeled "I've never heard of this movie" (figure 1).

Next, participants were asked to imagine that they were planning to go see a movie and were asked (on each of 30 trials) to choose which of two hypothetical theaters they would go to (figure 1). Each of the theaters was described as having either one or two movies playing. Participants were told that if they chose a theater with two movies, they would get to choose one of the two movies to see. Participants completed comprehension questions before the choices to ensure that they understood that (1) the choices were hypothetical, (2) that they would get to choose one (and only one) movie to watch, even if they chose a theater with two movies, and (3) that their choices would not influence the number of choices that they would have to make. The instructions and comprehension questions provided to participants in all studies are available on ResearchBox (https://researchbox.org/124&PEER_REVIEW_passcode=WUXZVT).

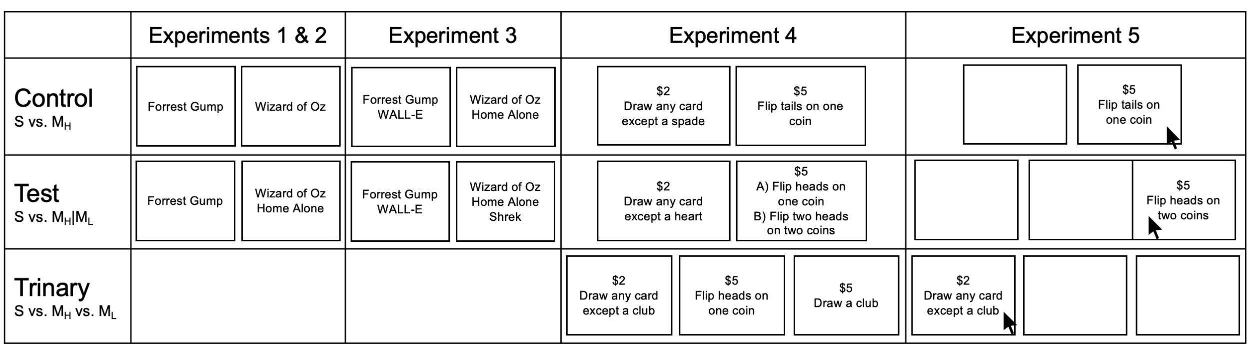
The 30 choices were randomly generated for each participant. They fell into 3 categories (with 10 choices in each category) but were presented in randomized order. The first category of choices were *test* choices. In each of these trials, participants chose between a theater with one available movie (i.e., movie S) and a theater with two available movies (i.e., movies M_H and M_L, where the participant's prior rating of M_H was greater than the rating of M_L). Each test choice trial comprised three unique movies, which did not overlap between trials. Thus, the 10 test choices consisted of 30 different movies. These movies were drawn randomly from the movies that the participant rated. If a participant rated fewer than 30 films, then we generated as many trials as possible from their rated films before drawing from the unrated films (i.e., the films for which they selected "I've never heard of this movie"). In this study, only one participant rated fewer than 30 films.

As an example, consider a participant who rated 42/50 films, and clicked "I've never heard of this movie" for the remaining 8/50. To generate the test trials for this participant, we drew (without replacement) 30 films from the 42/50 that were rated, and randomly sorted them into 10 groups comprising 3 films each. One of these groups may be the triad {Wizard of Oz, Home Alone, Forrest Gump}, which our example participant rated {6.5, 4.3, 8.2}, respectively. Within each triad, we randomly assigned one of the films to be S. In this case, Forrest Gump is randomly assigned to be S. The remaining films {Wizard of Oz, Home Alone} were designated

M_H and M_L. Since the example participant rated Wizard of Oz higher than Home Alone, Wizard of Oz is M_H and Home Alone is M_L.

The second category of choices were *control* binary choices. These trials consisted of a choice between two theaters, each with one available movie (e.g., Forrest Gump vs. Wizard of Oz). Importantly, each of the 10 control binary choices was matched to one of the test choices. Specifically, each control binary choice was a choice between the single-movie theater and the a priori higher-rated film from the two-movie theater from the matched test choice. Using the test choice example above (movie S vs. movies M_H or M_L; e.g., Forrest Gump vs. Wizard of Oz or Home Alone), the control binary choice would be movie S vs. movie M_H (where movie M_H's rating was higher than movie M_L's rating prior to any choices; e.g., Forrest Gump vs. Wizard of Oz).

Figure 1: Study Design.



Notes: In Studies 1-3, participants first rated their desire to watch each of 50 films. Participants then made 30 hypothetical choices between two theaters, each of which was showing one or two films. Participants were told to imagine that if they chose a two-film theater, they would get to choose one of the two films to watch. In Studies 4 and 5, participants chose between 2 (or 3) options.

Finally, the last category of 10 trials were *filler* choices, each of which comprised two single-movie theaters (e.g., Black Panther vs. Saving Private Ryan). The filler trial films were randomly selected from the rated films, and independently from the selection process for the test trials. These trials were used for our preregistered exclusion criteria and to add some variability to the choices being made by participants.

After rating the movies and before making any choices, participants demonstrated their understanding of the task by completing several comprehension check questions, including a question about the meaning of the test choices, i.e., that choosing "M_H or M_L" implied that they would get to choose which movie from the set $\{M_H, M_L\}$ they wanted to see. They were required to get all questions correct before moving on to the choices. In the supplements, we describe a replication of Study 1 and 1b (namely, Study 1c) that uses an open-ended comprehension check question to firmly rule out "misunderstanding the task" as the driving force behind the results that we describe below. Please see the supplements for additional detail on Study 1c.

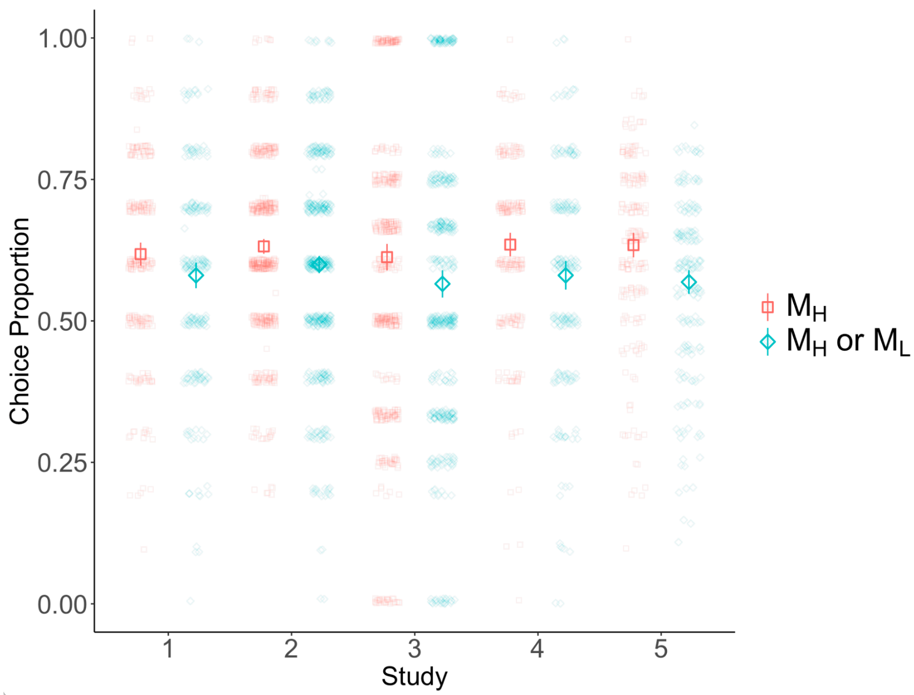
Exclusions and Data Preprocessing. As specified in our preregistration, we excluded anyone who failed to rate at least 20 films, as they would have fewer usable test and control choices and thus noisier estimates. We also excluded anyone whose filler choices were not directionally predicted by their ratings in a logistic regression of *ChooseLeft* on (*RatingLeft-RatingRight*) as that indicates lack of minimal attention during ratings, choice, or both. These criteria resulted in the exclusion of 37 participants, leaving us with a final sample size of 268. As specified in our preregistration, we excluded any test-control choice pairs that were generated from unrated films. For instance, if a participant only rated 25 films, then we were only able to generate eight valid test-control choice pairs. The other two test-control choice pairs were generated from unrated films, and thus excluded from analysis.

Results

To test for undervaluation, for every participant, we calculated S > {M_H, M_L} as (Chose S in test choices / number of valid test choices). We also calculated S > M_H as (Chose S in control choices / number of valid control choices). We tested (S > {M_H, M_L}) – (S > M_H) using a one-sample t-test and found evidence for undervaluation, $M = 0.04$, 95% CI = [0.02, 0.06], $t(267) = 3.73$, $p < .001$. In other words, choice of M was 4 percentage points lower than choice of M_H (see figure 2 for choice proportions of M and M_H and table 1 for means and effect sizes across all studies). 40% (106/268) of participants exhibited undervaluation, with 34% (91/268) exhibiting no difference and 26% (71/268) exhibiting an effect in the opposite direction. The 34% exhibiting no difference were not perfectly consistent, as in many cases it reflected multiple offsetting inconsistencies. This indicates that if there were more choices, the proportion of people exhibiting undervaluation would likely be greater.

Because M_H and M_L were randomly selected (and thus, the difference between ratings varied across choices and participants), we can examine the role of preference strength for M_H over M_L in undervaluation. We used matched choice sets as the unit of observation with the variable Choice Difference as our outcome measure, defined as (Chose S in test – Chose S in matched control), thereby taking a value of +1, 0, or -1. We regressed Choice Difference on the rating of S, the summed rating of M_H and M_L, and the difference in rating between M_H and M_L, clustering the standard errors at the subject level (eq. 1). Returning to our example participant from the methods above, the MH-ML preference strength for the example test choice (Forrest Gump vs. Wizard of Oz or Home Alone) is equal to 6.5-4.3 = 2.2. We computed the preference strength in this way for all triads of films.

Figure 2: Main Effect Across Studies.



Notes: Main effect of undervaluation across studies. Participants are less likely to choose {M_H, M_L} than they are to choose M_H when compared to the same single-option alternative S. Bars indicate s.e.m. across participants.

Table 1. Main Effect of Undervaluation Across All Studies.

Study	**M_H**	**M_{HL}**	**Main Effect**	**95% CI**	**t statistic[b]**	**p value**
1	0.618	0.580	0.04	[0.02, 0.06]	3.73	< .001
1b (supplements)	0.632	0.590	0.04	[0.03, 0.05]	6.75	< .001
1c (supplements)	0.667	0.645	0.02	[0.01, 0.04]	3.18	.002
2	0.632	0.600	0.03	[0.02, 0.04]	5.86	< .001
2b (supplements)	0.630	0.605	0.03	[0.01, 0.04]	2.94	.003
3[a]	0.587	0.544	0.05	[0.03, 0.07]	4.95	< .001
3b (supplements)[a]	0.614	0.534	0.08	[0.04, 0.12]	3.83	< .001
4	0.635	0.580	0.05	[0.03, 0.08]	5.04	< .001
4b (supplements)	0.572	0.508	0.06	[0.03, 0.09]	4.28	< .001
4c (supplements)	0.576	0.510	0.06	[0.02, 0.09]	3.37	< .001
5	0.634	0.569	0.06	[0.04, 0.08]	5.64	< .001

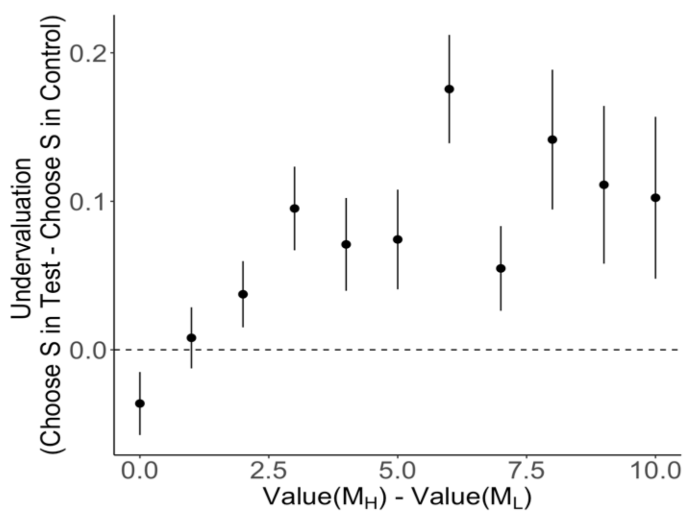
[a] Based on preregistered analysis of 1v1 and 1v2 choices where $Rating_{M_H} - Rating_{M_L} \geq 2$.
[b] Each of these tests are consistent with the results of a corresponding Wilcoxon signed-rank test.

$$Choice\ Difference_{it} = \beta_0 + \beta_1 S_{it} + \beta_2(M_H + M_L)_{it} + \beta_3(M_H - M_L)_{it} + \epsilon_{it} \quad (1)$$

We find significant evidence for an effect of M_H-M_L preference strength on undervaluation, $b_3 = 0.019$, $SE = 0.003$, $p < .001$ (figure 3), such that participants exhibited more undervaluation for multi-option alternatives with bigger rating differences. We replicated each of these results in a direct replication, Study 1b, detailed in the supplements. We interpret these results as indicating the effect grows with preference strength, although measurement error could plausibly contribute (such that smaller rating differences are more likely to be associated with mismeasured preferences). We further examine this possibility in Study 4.

Moreover, in the supplements, we demonstrate that the undervaluation effect persists, regardless of whether the rating of S is higher than M_H, lower than M_L, or falls in between the two. When S is rated higher than M_H, the choice share of M_H is below 0.5 (and the choice share of $\{M_H, M_L\}$ is even lower). This pattern indicates that adding a less-preferred option (M_L) to M_H in fact decreases choice share, rather than merely adding noise and pushing the choice share toward 50%.

Figure 3: Preference Strength and Undervaluation.



Notes: Relationship between (M_H-M_L) preference strength and degree of undervaluation in Study 1. Undervaluation increases as M_H gets progressively better than M_L. Analysis reported in text controls for rating of S and sum of ratings of M_H and M_L. Bars represent s.e.m. across participants.

STUDY 2

A key question is the exact method by which adding M_L to M_H to form the multi-option alternative $\{M_H, M_L\}$ decreases choice probability. One mechanism would be that adding M_L reduces the perceived value of M_H, thereby reducing the derived value of M. A second distinct mechanism would be that adding M_L reduces the perceived holistic value of M. To identify the process/mechanism, we designed the following study.

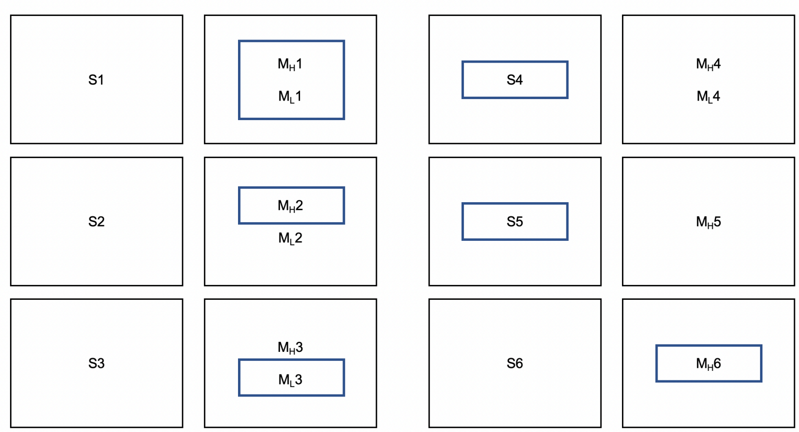
Method

*Participants***.** For this preregistered study (https://researchbox.org/124&PEER_REVIEW_passcode=WUXZVT), we collected responses from 601 Amazon Mechanical Turk (AMT) workers. They earned $1.50 for their participation.

Materials and Procedure. The procedure for this study was nearly identical to Study 1. The only difference is that participants were asked to rate 8 options or alternatives during the choice portion of the study. Following the respective choice, participants reported their rating (on the same 0-10 scale they had used for the ratings at the beginning of the study) of (1) S in a test choice, (2) M_H in a test choice, (3) M_L in a test choice, (4) M in a test choice, (5) S in a control choice, (6) M_H in a control choice, (7) the left option in a filler choice, and (8) the right option in a filler choice. The first 6 ratings are of primary interest and came from 6 different sets of matched trials (figure 4).

Exclusions and Data Preprocessing. As specified in our preregistration, we excluded anyone who failed to rate at least 20 films. We also excluded anyone whose filler choices were not directionally predicted by their ratings in a logistic regression of *ChooseLeft* on (*RatingLeft-RatingRight*). These criteria resulted in the exclusion of 38 participants, leaving us with a final sample size of 563.[2] As specified in our preregistration, we excluded any test-control choice pairs that were generated from unrated films.

Figure 4: Primary Ratings in Study 2.



Notes: These 6 ratings followed choices from 6 different matched sets. When asked about an individual option (i.e., S, M_H, or M_L), participants were asked “What is your rating of this option?” When asked about the multi-option alternative (i.e., M), participants were asked “What is your rating of watching a movie of your choice from this option?”

[2] One participant’s choices, when regressed on the ratings of the films, were not directionally predicted (i.e., b = 0) in a linear regression. However, in a logistic regression of the same participant’s choices, the fitted coefficient was very small, albeit positive (i.e., 1e-16). We report the results without this participant. None of the results change if we include the participant.

Results

As in previous studies, we find evidence of undervaluation, $M = 0.03$, 95% CI = [0.02, 0.04], $t(562) = 5.86$, $p < .001$. Next, we make several comparisons between ratings. First, we look at whether the rating of S changes, depending on whether it is part of a test choice or a control choice, and do not find evidence for a significant difference between the rating of S following binary and the rating of S following test ($M = 0.20$, 95% CI = [-0.07, 0.48], $t(562) = 1.44$, $p = .15$). Next, we make within-subject comparisons among three values: (i) the rating of M_H following control choices, (ii) the rating of M_H following test choices and (iii) the rating of $\{M_H, M_L\}$ following test choices, using one-sample t-tests.

As specified in our preregistration, to reduce sampling variability in the selected movies and thereby increase power, we account for the initial ratings of the options in our analyses by subtracting the initial ratings. We find a significant difference between M_H following binary and $\{M_H, M_L\}$ following test ($M = 0.23$, 95% CI = [0.02, 0.44], $t(562) = 2.14$, $p = .03$), but we do not find a significant difference between M_H following binary and M_H following test ($M = –0.07$, 95% CI = [–0.27, 0.12], $t(562) = –0.72$, $p = .47$). The difference between M_H following binary and $\{M_H, M_L\}$ following test is significantly larger than the small difference between M_H following binary and M_H following test ($M = 0.30$, 95% CI = [0.09, 0.51], $t(562) = 2.83$, $p = .005$).[3] This test provides evidence that adding a lower-value option decreases the value of a set. Moreover, this effect is greater than any effect on the value of M_H alone.

[3] These results are consistent with those of another identical, albeit smaller, study, Study 2b. See the supplements for details.

STUDY 3

In the next study, we extended our results to a broader array of choice set sizes. Rather than just examining choices between two one-movie theaters (1v1) and choices between one one-movie theater and one two-movie theater (1v2), we also examined choices between one one-movie theater and one three-movie theater (1v3), two two-movie theaters (2v2), and one two-movie theater and one three-movie theater (2v3). This enabled us to consider the effect of adding a marginal option beyond the 1v1 set.

Method

Participants. For this preregistered study (https://researchbox.org/124&PEER_REVIEW_passcode=WUXZVT), we collected responses from 813 AMT workers. They earned $1.50 for their participation.

Materials and Procedure. First, as in Studies 1 and 2, participants rated 50 well-known films on a scale of 0-10 (measured in increments of 0.1 via a slider) regarding how much they wanted to watch each. For each film, participants could click a box labeled “I’ve never heard of this movie.”

Next, participants were asked to imagine that they were planning to go see a movie and were asked (on each of 42 trials) to choose which of two hypothetical theaters they would go to (figure 1). Each of the theaters was described as having either one, two, or three movies playing. Participants were told that if they chose a theater with more than one movie, they would get to choose one of the movies to see. Participants completed comprehension questions before the choices to ensure that they understood that (1) the choices were hypothetical, (2) that they would get to choose one (and only one) movie to watch, even if they chose a theater with more than one movie, and (3) that their choices would not influence the number of choices that they would have to make.

The 42 choices presented to participants were randomly generated for each participant. Of these 42 choices, 17 were filler choices, while the remaining 25 were used for our primary analyses. These 25 primary trials fell into five sets of trials, each of which included five matched trials. Each set of five matched trials included one trial of each of five types. The first type of trials was 1v1 choices. In each of these trials, participants chose between two theaters, each with one available movie (i.e., movie S_A and movie M_A); these correspond to binary control choices from Study 1. The second type of trials were 1v2 choices, where participants chose between a single-option theater (showing S_A) and a theater with two available movies (i.e., movies M_A and M_B); these correspond to test choices from Study 1. The third type of trials were 1v3 choices, where participants chose between a single-option theater (showing S_A) and a theater with three available movies (i.e., movies M_A, M_B, and M_C). The fourth type of trials were 2v2 choices, where participants chose between two two-option theaters (i.e., one theater showed movies S_A and S_B, and the other theater showed movies M_A and M_B). The fifth type of trials were 2v3 choices, where participants chose between a theater showing two movies (S_A and S_B) and a theater showing three movies (M_A, M_B, and M_C). In all cases, Rating(S_A) >= Rating(S_B), Rating(M_A) >= Rating(M_B), and Rating(M_A) >= Rating(M_C).

Each matched set of five trials thus included five different movies assigned to the roles of S_A, S_B, M_A, M_B, and M_C, and there were five such matched sets. Across all matched sets of trials,

S_A, S_B, M_A, M_B, and M_C were unique and non-overlapping. Thus, the 25 primary choices consisted of 25 different movies divided into 5 sets of 5 movies each. These movies were drawn randomly from the rated movies. If a participant rated fewer than 25 films, then we generated as many trials as possible from their rated films before drawing from the unrated films.

Among the 17 filler trials, 9 were simple binary trials, each of which comprised two single-movie theaters. The simple binary trial films were randomly selected from the rated films (independently from the selection process for the test trials, above). If a participant rated fewer than 18 films, then we sampled from the unrated films when necessary. These trials were used for our preregistered exclusion criteria and to add some variability to the choices being made by participants. The remaining 8 filler trials were designed to look like the trials from the other trial categories specified above. The films in each trial were randomly selected from the rated films (separately from the selection process mentioned above).

Trials were presented in a randomized-block order. Participants completed 3 or 4 filler questions before each block of 5 choices. Each block contained one question of each type (1v1, 1v2, 1v3, 2v2, 2v3), each from a different matched set. The display order of the sub-options (e.g., M_A-or-M_B vs. M_B-or-M_A) was randomly selected at the participant-by-trial level.

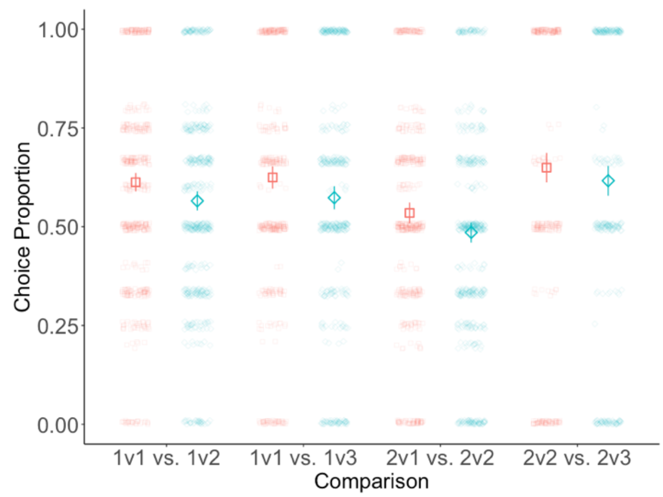
After rating the movies but before making any choices, participants demonstrated their understanding of the task by completing several comprehension questions, including a question about the meaning of the test choices (i.e., that choosing a theater with multiple options implied that they would get to choose which movie from the set they wanted to see). They were required to get all questions correct before moving on to the choices.

Exclusions and Data Preprocessing. As specified in our preregistration, we excluded anyone who failed to rate at least 18 films, as they would have fewer usable primary choices and thus noisier estimates. We also excluded anyone whose simple binary choices were not directionally predicted by their ratings in a logistic regression of *ChooseLeft* on (*RatingLeft-RatingRight*) as that indicates lack of minimal attention during ratings, choice, or both. These criteria resulted in the exclusion of 38 participants, leaving us with a final sample size of 775.

Also in line with our preregistration, we excluded any choice sets that were generated from unrated films. For instance, if a participant only rated 21 films, then we were only able to generate 4 valid 5-film sets. The remaining 5-film set would be generated from unrated films and thus excluded from analysis. Finally, for statistical power and as preregistered, our analyses only included sets in which the added sub-option had a rating at least 2 points lower than the existing sub-option(s).[4] We preregistered this criterion because (1) when you have more options, it is more likely that the direction of preference could be miscoded due to measurement error and (2) with a small enough difference, option value might outweigh the effects of undervaluation (see the x-intercept in figure 3).

[4] For the most basic comparison, between 1v1 and 1v2 trials, we required Rating(M_A) >= Rating(M_B)+2. For the comparison between 1v1 and 1v3 trials, we required Rating(M_A) >= Rating(M_B)+2 and Rating(M_A) >= Rating(M_C)+2. For the comparison between 1v2 and 2v2 trials, we required Rating(S_A) >= Rating(S_B)+2. Finally, for the comparison between 2v2 and 2v3 trials, we required Rating(M_B) >= Rating(M_C)+2.

Figure 5: Undervaluation in Study 3



Notes: Main effect of undervaluation in Study 3, using multi-option alternatives with different numbers of options. Participants are less likely to choose the alternative with an additional (inferior) sub-option. Bars indicate s.e.m. across participants.

Results

In this study, we can test for undervaluation of several theaters. First, analogous to analyses in Studies 1 and 2, we can compare choices in 1v1 and 1v2 trials to test for undervaluation of the 2-movie theater in 1v2 choices. To test for undervaluation in this case, for every participant, we calculated $S_A > \{M_A, M_B\}$ as (Chose S in 1v2 choices / number of valid 1v2 choices). We also calculated $S_A > M_A$ as (Chose S in 1v1 choices / number of valid 1v1 choices). We tested $(S_A > \{M_A, M_B\}) – (S_A > M_A)$ using a one-sample t-test and found evidence for undervaluation, $M = 0.05$, 95% CI = [0.03, 0.07], $t(655) = 4.95$, $p < .001$ (figure 5). In other words, choice of M_A was 5 percentage points higher than choice of $\{M_A, M_B\}$.

We used the analogous approach to compare 1v1 against 1v3, 1v2 against 2v2, and 2v2 against 2v3. In each case, we find undervaluation (1v1 vs. 1v3: $M = 0.05$, 95% CI = [0.03, 0.08],

$t(602) = 4.09$, $p < .001$; 1v2 vs. 2v2: $M = 0.05$, 95% CI = [0.03, 0.07], $t(656) = 4.72$, $p < .001$; 2v2 vs. 2v3: $M = 0.03$, 95% CI = [0.002, 0.06], $t(463) = 2.15$, $p = .03$[5]).

STUDY 4

In Studies 1-3, we established the presence of undervaluation and its moderation by M_H-M_L preference strength in a common consumer domain with familiar products in preferential choice. In the following two studies, we switch to the domain of risky choice, using card draws, die rolls, and coin flips. This new domain enables greater control, incentive-compatibility, and testing for cases of objective dominance, while maintaining a similar experimental design. These new domains also reduce the likelihood that participants drew inferences about theater quality or the context of the consumption experience based on the set of movies shown. Finally, the objective dominance in this domain further reduces the likelihood that participants undervalue a multi-option alternative due to uncertainty about their own future preferences.

Method

Participants. For this preregistered study (https://researchbox.org/124&PEER_REVIEW_passcode=WUXZVT), we collected responses from 304 AMT workers. They earned $1.75 (the first 20 participants) or $1.90 (the remaining 284 participants) for their participation. Five randomly-selected participants also received the

[5] All of these results are consistent with those of another identical study, Study 3b. See the supplements for more details.

outcome of one of their decisions, as detailed below.

Materials and Procedure. This study was similar to Study 1, with the following changes. First, choice options were gambles, as depicted in figure 1. These gambles were drawn from a large pool of potential gambles from various combinations of drawing cards of different suits, flipping coins or pairs of coins, or rolling numbers on a six-sided die. Binary control choices, test choices, and trinary choices (described below) were matched on winning dollar amounts and probabilities but could vary (non-systematically) in terms of the specific mechanism as shown in figure 1. Second, participants did not rate any options, as we could model each option's value in terms of its payout and probability. Third, in addition to test choices (e.g., choosing between S and a multi-option alternative $\{M_H, M_L\}$) and binary control choices (e.g., choosing between S and M_H), participants also made trinary choices (e.g., choosing among gambles S, M_H, and M_L). In total, participants made 34 incentivized choices (including three attention check questions and one M_H or M_L choice from a randomly-selected test choice).

At the end of data collection, we randomly selected five participants and then randomly selected one of their choices to play out. Participants were aware of this structure. We opted to reward one trial instead of all of the trials to avoid stockpiling/strategy variability across trials (Juechems et al. 2017). In this study, three of the five participants won money ($6, $7, and $10). See the supplements for additional information on the trial generation process.

As in Studies 1-3, participants completed several comprehension check questions, including a question about the meaning of the test choices, i.e., that choosing $\{M_H, M_L\}$ implied that they would get to choose one option from the set $\{M_H, M_L\}$. They were required to get all questions correct before moving on to the choices.

Exclusions and Data Preprocessing. As specified in our preregistration, we excluded anyone who picked an obviously dominated option in any of the three attention-check questions. This resulted in the exclusion of 63 participants, leaving us with a sample size of 241.

Results

Preregistered Results. To test for undervaluation, for every participant, we calculated $S >$ {M_H, M_L} as (Chose S in test choices / Number of test choices). We also calculated $S > M_H$ as (Chose S in binary choices / Number of binary choices). We tested ($S >$ {M_H, M_L}) – ($S > M_H$) using a one-sample t-test and found evidence for undervaluation, $M = 0.05$, 95% CI = [0.03, 0.08], $t(240) = 5.04$, $p < .001$ (figure 2). Choice of M_H was 5 percentage points higher than choice of M (table 1). Using that approach, we find that 46% (111/241) of participants exhibited undervaluation overall, with 34% exhibiting no effect and 20% exhibiting an effect in the opposite direction. We would again expect the number exhibiting undervaluation to increase as the number of trials increases.

Because the probabilities of M_H and M_L were randomly selected, we can examine the role of M_H-M_L dominance (i.e., how much better M_H is than M_L) in undervaluation. We used Choice Difference as our outcome measure, defined as (Chose S in test choices – Chose S in binary choices) at the set-level. We regressed Choice Difference on the expected value of S, the summed expected values of M_H and M_L, and the difference in expected values between M_H and M_L, clustering the standard errors at the subject level (eq. 2).

$$Choice\ Difference_{it} = \beta_0 + \beta_1 EV_{S_{it}} + \beta_2 \left(EV_{M_H} + EV_{M_L}\right)_{it} + \beta_3 \left(EV_{M_H} - EV_{M_L}\right)_{it} + \epsilon_{it} \qquad (2)$$

The coefficient on the difference in expected values was in the expected direction, though it was not statistically significant, $b_3 = 0.011$, $SE = 0.008$, $p = .16$ (figure 6a).

Exploratory Results. One possible reason for the non-significant coefficient on expected value is that participants were less attuned to the expected values of the options and more attuned to the probabilities themselves, perhaps due to how the gambles were displayed. To investigate this possibility, we turned to the binary choices. Using logistic regression, we regressed choice of S on the expected values of S and M_H (EV), the monetary amounts associated with S and M_H (*M*), and the probabilities of S and M_H (*P*), clustering the standard errors at the subject level (eq. 3).

$$Choice_{it} = \beta_0 + \beta_1 EV_{S_{it}} + \beta_2 EV_{M_{H_{it}}} + \beta_3 M_{S_{it}} + \beta_4 M_{M_{H_{it}}} + \beta_5 P_{S_{it}} + \beta_6 P_{M_{H_{it}}} + \epsilon_{it} \qquad (3)$$

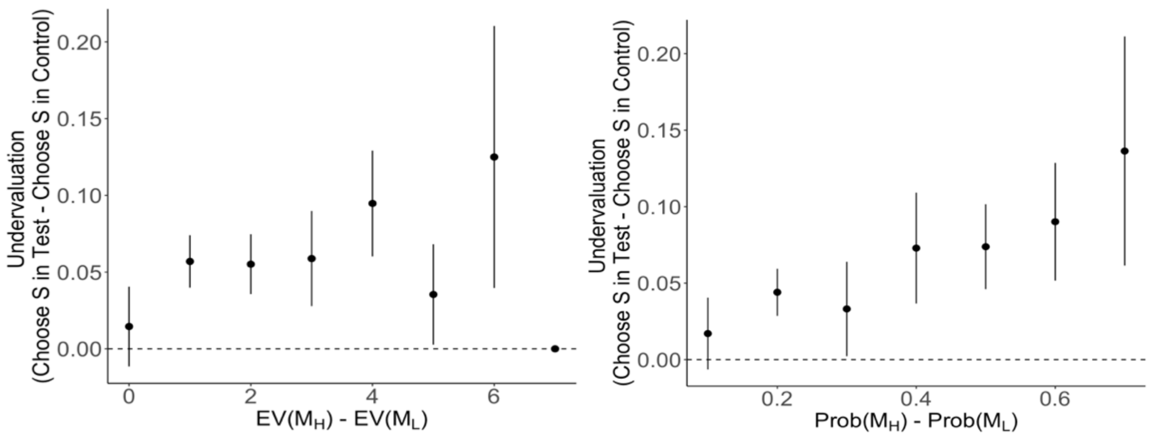
After accounting for the effects of expected value (S: $b_1 = 0.42$, $SE = 0.12$, $p < .001$; M_H: $b_2 = -0.24$, $SE = 0.15$, $p = .11$), we find significant effects of both monetary amounts (representing the simple effects of monetary amount when probabilities are equal to 0; S: $b_3 = 0.19$, $SE = 0.06$, $p = .002$; M_H: $b_4 = -0.28$, $SE = 0.10$, $p = .003$) and both probabilities (representing the simple effects of probabilities when monetary amounts are equal to 0; S: $b_5 = 2.96$, $SE = 0.77$, $p < .001$; M_H: $b_6 = -2.83$, $SE = 0.91$, $p = .002$).

These results suggest that payout and probabilities contributed to evaluation of the gambles above and beyond their contribution to expected value. We thus adapted our preregistered analysis. Instead of regressing Choice Difference on expected values, we regressed it on probabilities.[6] Specifically, we regressed Choice Difference on the probability of S, the summed probabilities of M_H and M_L, and the difference in probabilities between M_H and M_L, with clustered SEs at the subject level (eq. 4).

[6] We are unable to conduct this analysis using monetary amounts because M_{M_H} was always the same as M_{M_L}.

$$Choice\ Difference_{it} = \beta_0 + \beta_1 P_{S_{it}} + \beta_2 (P_{M_H} + P_{M_L})_{it} + \beta_3 (P_{M_H} - P_{M_L})_{it} \quad (4)$$

Figure 6: Dominance Strength and Undervaluation.



Notes: Relationship between (M_H>M_L) dominance strength and degree of undervaluation. (a) Dominance strength is defined using expected value. (b) Dominance strength is defined using probabilities. Across both definitions of dominance, undervaluation increases as M_H gets progressively better than M_L, though this relationship is not significant for expected value (p = .16) and marginally significant for probability (p = .07). Bars represent s.e.m. across participants. See the supplements for plots with data overlaid (figure S3).

Consistent with our findings in Study 1, we find a marginally significant relationship between M_H-M_L dominance and Choice Difference (b_3 = 0.11, *SE* = 0.06, p = .07, figure 6b). This speaks in favor of the preference strength explanation (vs. measurement error), as probabilities here are known rather than measured with a noisy instrument.

We also compared choices in the test choices (S vs. {M_H, M_L}) to choices in the trinary (S vs. M_H vs. M_L) choices. Using a one-sample t-test of (Choose {M_H, M_L} in test choices) – (Choose M_H or M_L in trinary choices), we find that participants choose the multi-option alternative (M_H or M_L) in test choices significantly less than they choose the same two options (M_H; M_L) in the trinary choices, M = –0.08, 95% CI = [–0.10, –0.06], $t(240)$ = –7.95, p < .001 (see figure S4). There was not a significant difference between choice for the multi-option alternative ({M_H, M_L}) in test choices and choice for the higher-valued option (M_H) in trinary choice, $M = 0.001$, 95% CI = [–0.02, 0.03], $t(240) = 0.10$, $p = .92$. Participants chose M_H significantly more in binary choice than they did in trinary choice ($M = 0.06$, 95% CI = [0.04, 0.07], $t(240) = 5.79$, $p < .001$).

Replications

We ran two additional, similar studies (4b: N = 298, 194 after exclusions; 4c: N = 298, 176 after exclusions). Participants in both studies exhibited undervaluation (4b: $M = 0.06$, 95% CI = [0.03, 0.09], $t(193) = 4.28$, $p < .001$; 4c: $M = 0.06$, 95% CI = [0.02, 0.09], $t(175) = 3.37$, $p < .001$). In these studies, we examine the relationship between undervaluation and individual difference measures (risk aversion, analytic vs. holistic thinking, and elaboration on potential outcomes); we do not find any meaningful relationships between these individual difference measures and undervaluation. In these studies, we also provide evidence against a few alternative explanations for the main effect of undervaluation. First, we rule out strong avoidance of multi-option alternatives by showing that people are more likely to pick “M_H or M_L” (vs. S) than they are to pick M_L (vs. S). Second, we provide further evidence against the notion that noisy/careless responding and/or misunderstanding could explain our effect; when we limit our analysis to participants who (a) got every comprehension question correct on the first try, (b) never selected a dominated alternative, or (c) the combination of these criteria, our effect does not go away. If anything, the effect size increases. Finally, in Study 4c, participants made choices between M_H and M_L. This rules out the possibility that the results are driven by the fact that participants never encounter the second-stage choice. This is in addition to the fact that in every study, participants answered a comprehension check question that they understood the number of choices they faced would not depend on the options they chose. Additional details are available in the supplements.

STUDY 5

In Study 5, we introduce a process-tracing measure to our design: mouse-tracking. We measure the information acquisition process during all choices and connect it to our main effect. We also manipulated whether the M_H vs. M_L dominance relationship was high-transparency (easy to identify) or low-transparency (difficult to identify). The monetary outcomes and the probabilities of winning were held constant between high-transparency pairings and low-transparency pairings, but in the high-transparency pairings, M_H always contained the event M_L and in low-transparency pairings, M_H never contained the event M_L. An example of a high-transparency pairing is M_H = "$2 if you roll a 1 or 2 or 3" and M_L = "$2 if you roll a 1." An example of a low-transparency pairing is M_H = "$2 if you draw a black card" and M_L = "$2 if you roll a 4." We hypothesized that a more (vs. less) transparent dominance relationship would lead to less undervaluation as the implications for the second-stage choice of M_H vs. M_L are clearer for more-transparent pairings.

Method

Participants. For this preregistered study (https://researchbox.org/124&PEER_REVIEW_passcode=WUXZVT) we collected responses from 302 AMT workers. They earned $2.50 (the first 20 participants) or $2.75 (the remaining

282 participants) for their participation. Five randomly-selected participants received the outcome of one of their decisions, as detailed below.

Materials and Procedure. This study was similar to Study 4, with the following changes. First, the trials were not randomly generated at the subject-level and instead came from a predetermined set. Second, the M_H vs. M_L dominance relationship was manipulated to be either high- or low-transparency. Third, the information about a gamble (i.e., the details of S, M_H, and/or M_L) was not visible unless the participant hovered their cursor over the gamble. Fourth, M_H and M_L were presented horizontally instead of vertically. Fifth, we tracked participants' mouse movements. Specifically, while participants made their choices, we recorded the order in which participants viewed each piece of information (e.g., S, M_H, and M_L) and the durations of these information acquisitions in a MouseLab-like paradigm (Johnson et al. 1989) (figure 1). Participants made a total of 65 incentivized choices. These choices included: (1) high-transparency test choices: S vs. M_H or M_L (HT), (2) low-transparency test choices: S vs. M_H or M_L (LT), (3) high-transparency control choices: S vs. M_H (HT), (4) low-transparency control choices: S vs. M_H (LT), (5) trinary choices: S vs. M_H vs. M_L, and (6) binary choices between M_H and M_L. At the end of the survey, participants rank-ordered 10 {S, M_H, M_L} triplets. See the supplements for more specific trial information. In this study, all five randomly-selected participants won money ($3, $5, $4, $2 and $4).

As in the previous studies, participants completed several comprehension check questions, including a question about the meaning of the test choices, i.e., that choosing {M_H, M_L} implied that they would get to choose an alternative from the set {M_H, M_L}. They were required to get all questions correct before moving on to the choices.

Exclusions and Data Preprocessing. As specified in our preregistration, we excluded anyone who picked the obviously dominated option in any of the three attention-check questions and we excluded anyone who did not mouse-over the boxes in the instructions as instructed. This resulted in the exclusion of 93 participants, leaving us with a sample size of 209.

For the mouse-tracking data, we converted the hover-times (i.e., the times that participants spent hovering over the available information) into proportions from 0 to 1 at the trial level.

Results

Choice Results. We tested for undervaluation using the method in Study 2, and we find evidence for undervaluation, $M = 0.06$, 95% CI = [0.04, 0.08], $t(208) = 5.64$, $p < .001$ (figure 2). Choice share of M was 6 percentage points lower than choice share of M_H. We tested for moderation by transparency by calculating the within-subject differences in undervaluation on high-transparency trials and undervaluation on low-transparency trials. Using a one-sample t-test we find significant evidence of the moderation ($M = –0.03$, 95% CI = [–0.06, –0.01], $t(208) = –2.76$, $p = .006$). There is less undervaluation when the M_H-M_L dominance is more transparent (M(HT) = 0.04, 95% CI = [0.02, 0.06], $t(208) = 3.70$, $p < .001$; M(LT) = 0.07, 95% CI = [0.05, 0.10], $t(208) = 5.95$, $p < .001$).[7] This moderation indicated attenuation but not elimination. We

[7] Studies 4b and 4c reported in the supplements also included a transparency manipulation without process tracing. We did not observe moderation by transparency in those two studies. As described in the General Discussion, we speculate that the focusing enforced by the process tracing increased the moderating effect of transparency on choice.

still find evidence of undervaluation even when the dominant relationship is transparent.

We find that 61% (127/209) of participants exhibited undervaluation overall (46% (97/209) exhibited undervaluation on high-transparency trials and 59% (123/209) exhibited undervaluation on low-transparency trials). Of the remaining 39%, 10% (20/209) exhibited no effect and 29% (62/209) exhibited an effect in the opposite direction. As in prior studies, we expect the percentage exhibiting undervaluation would increase with more choices. We observe undervaluation for 9/10 of the choice sets overall, 8/10 of the high-transparency choice sets, and 9/10 of the low-transparency choice sets.

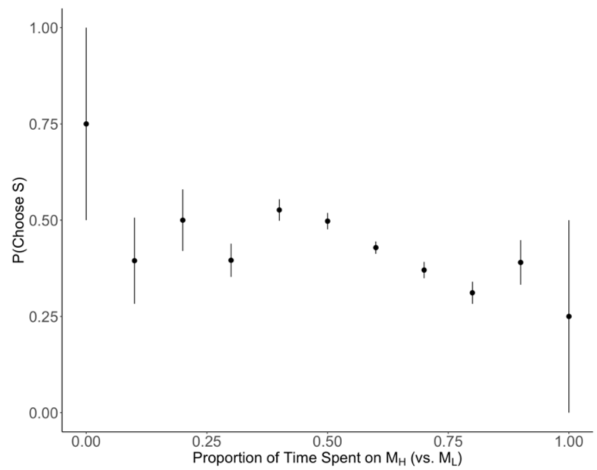
Mouse-Tracking Results. We tested for a relationship between aggregate information acquisition (i.e., mouse movements) and undervaluation. For each subject, we computed the difference in average proportion of time spent on S between test and binary control trials (i.e., average proportion spent on S in S vs. {M_H, M_L} choices – average proportion spent on S in S vs. M_H choices). We find a positive correlation between-subject between this average proportion mouse difference and degree of undervaluation, $r = 0.27$, $t(207) = 4.04$, $p < .001$. Participants who spent relatively more time inspecting S when paired with M than when paired with M_H undervalued M more, which supports our process-tracing hypothesis (H4).[8]

In a trial-level analysis, we regressed choice for S (in S vs. {M_H, M_L} choices only) on the proportion of time spent on M_H relative to M_L (i.e., time spent on M_H / total time spent on M_H or M_L). We find a significant negative relationship ($b = –1.70$, $SE = 0.24$, $p < .001$), which

[8] We also tested for trial-level associations between mouse movements and choice. Across all binary, test, and trinary choices, we used logistic regression to regress choice (of S) on the proportion of mouse-hover time spent on S, with random intercepts and slopes at the subject level. We find a significant relationship between information acquisition and choice within-subject, $b = 4.79$, $SE = 0.18$, $p < .001$.

implies that the more time participants spent looking at M_H (i.e., the better outcome out of the M_H or M_L option), the less likely they were to choose S (i.e., the more likely they were to choose M_H or M_L) (figure 7).

Figure 7: Mouse-tracking and Choices



Notes: Relationship between mouse-tracking and choices in test choices. As participants spent more time on M_H (relative to M_L), they were less likely to choose the single-option alternative (S), $b = -1.70$, $SE = 0.24$, $p < .001$. 60% of all proportion of time spent on M_H lie in the interval [0.5, 0.75]. Within that interval, the coefficient is even more sharply negative (-3.05). Bars represent s.e.m. across participants.

Finally, we examined mouse-tracking differences between trinary and multi-option alternative choices. First, we examined the difference between average proportions of time spent on S, M_H, and M_L in each trial type (i.e., trinary vs. test) and compared these differences to undervaluation. No correlations were significant (all *p*s > 0.1). Additionally, we looked at the difference in average proportion of transitions between M_H and M_L in the trinary trials compared to the test trials (i.e., number of transitions between M_H and M_L / total number of transitions). We did not find any significant correlations there either (all *p*s > 0.1).

Exploratory Results. An advantage of mouse-tracking (and other process tracing measures) is the ability to look further into the decision process than simple choice outcomes allow. The significant transparency moderation lends itself particularly well to investigation through the lens of process tracing. Thus, we compared the proportion of time spent on M_H (relative to M_L) in the high-transparency trials to the proportion of time spent on M_H (relative to M_L) in the low-transparency trials. We found a small but significant difference, such that subjects spent relatively more time on M_H when the dominance was more transparent, $M = 0.01$, 95% CI = [0.002, 0.02], $t(207) = 2.46$, $p = 0.01$. Moreover, this subject-level difference in hover-times correlates with the subject-level difference in undervaluation between high- and low-transparency trials, $r = -0.13$, 95% CI = [–0.26, 0.008], $t(206) = -1.86$, $p = .06$. In other words, as subjects spent more relative time on M_H in high-transparency trials (vs. low-transparency trials), they showed a smaller degree of undervaluation in high-transparency trials (vs. low-transparency trials).

As in Study 4, we also compared choices in the test choices (S vs. $\{M_H, M_L\}$) to choices in the trinary (S vs. M_H vs. M_L) choices. Participants choose the multi-option alternative ($\{M_H, M_L\}$) in test choices significantly less than they choose the same two options (M_H; M_L) in the trinary choices, $M = –0.07$, 95% CI = [–0.09, –0.05], $t(208) = –6.77$, $p < .001$ (figure S6). In this study, participants choose the multi-option alternative (M_H or M_L) in test choices significantly less than they choose the higher-valued of the two (M_H) in trinary choice, $M = –0.02$, 95% CI = [–0.05, –0.001], $t(208) = –2.09$, $p = .04$. Also as in Study 4, participants choose M_H significantly more in binary choice than in trinary choice, $M = 0.04$, 95% CI = [0.02, 0.05], $t(208) = 3.62$, $p < .001$.

Although our main effect of interest (undervaluation) is statistically significant, and the presence of undervaluation is widespread across participants, there is variability in the size of the effect across participants ($M = 0.06$, $SD = 0.15$). Some of this variability can be attributed to individual differences. Specifically, the degree of undervaluation in low-transparency trials is correlated with the degree of undervaluation in high-transparency trials, $r = 0.50$, 95% CI = [-0.39, 0.59], $t(207) = 8.30$, $p < .001$. In other words, if a participant displayed a high degree of undervaluation on one half of the trials, they also displayed a high degree of undervaluation on the other half. This is consistent with individual differences in the tendency to exhibit undervaluation, rather than a uniform extent of undervaluation combined with pure noise in the choice data.

Unlike Studies 1-4, there was a limited number of unique choices and little variability in the relative dominance of M_H over M_L in Study 5 across choice sets, so that test had very low power and we do not discuss it here.

GENERAL DISCUSSION

Across eleven controlled online lab studies in two domains, we document a consistent effect: decision-makers undervalue multi-option alternatives. More specifically, they are less likely to choose a multi-option alternative $\{M_H, M_L\}$ than they are to choose a single-option alternative M_H. This effect also holds when the multi-option alternative has more than two options. Moreover, we find that the strength of undervaluation is related to the difference in values between M_H and M_L: as the difference in values increases, undervaluation increases as well. However, when the dominance of M_H over M_L is more transparent, there is less undervaluation. In replication studies (see supplements), we provide evidence against multiple alternative explanations (noisy responding; strong delayed choice aversion). We also find process evidence which suggests that undervaluation is strongly associated with information acquisition patterns and is driven by devaluating the holistic option M and not just M_H.

Related Work and Alternative Explanations

These findings connect to work on agenda effects (Plott and Levine, 1978; Tversky and Sattath 1979; Hauser 1986), in which the order of decisions influences the option that is ultimately chosen. However, this literature does not offer explanations for the present results. Instead, this undervaluation phenomenon seems to be an additional instance in which agenda effects matter. Moreover, in contrast to work on agenda effects, we find that undervaluation persists even in the simplest case: a single option pitted against a two-option alternative. Our current findings also have certain similarities with the uncertainty effect, indicating that in certain cases, consumers value a prospect as less valuable than its worst outcome (Gneezy, List, & Wu 2006; Simonsohn 2009; cf. Mislavsky & Simonsohn 2018; Yang, Vosgerau, & Loewenstein 2013). However, two key distinctions are that (a) the uncertainty effect addresses a risky outcome whereas here, the decision maker chooses an option from the multi-option alternative, and (b) the uncertainty effect is assessed relative to the worse outcome whereas undervaluation is assessed relative to the better outcome.

There are several literatures that address similar phenomena, including assortment choice and multi-alternative (i.e., more than 2 alternatives) choice. However, these literatures ultimately do not address the question at hand. Assortment choice research focuses on the evaluation of assortments but does not typically enable comparisons to a value-maximizing benchmark, nor does it compare evaluations of assortments to the evaluations of the constituent parts of the assortment in a choice context. Regarding multi-alternative choice research, our test choices (i.e., choices between a single-option alternative S and a multi-option alternative $\{M_H,M_L\}$) are formally equivalent to trinary choices (i.e., choices between S, M_H, and M_L). However, since participants are significantly less likely to choose $\{M_H,M_L\}$ (in a test choice) than they are to choose M_H or M_L (in a trinary choice), these two types of choices are neither psychologically nor practically equivalent.

Within the literature on assortment and multi-option choice, a natural comparison is Sood et al. (2004)'s finding that lone options are more likely to be chosen than grouped items when grouped items are initially compared to one another. This is because the intra-group comparisons lead to focusing on downsides of grouped options more than single options. Yet in the current examination in the risky domain, there are no tradeoffs among the grouped options, suggesting such comparative loss aversion does not explain these results. Another point of comparison is the model proposed by Kahn and Lehman (1991). This model suggests that the value of an assortment will decrease as the number of unacceptable items in the assortment increases. However, our results do not depend on this characterization. Indeed, M_L is not preferred to M_H, but it is preferable to receiving nothing, so it is not an unacceptable item. Our findings suggest that model may extend beyond unacceptable items and into less-desirable items.

We have addressed several potential alternative explanations with our data. First, we find evidence in Study 2 that the value of M in test choices is significantly lower than that of M_H in binary choices. Second, we rule out the possibility that our results are due to noisy or poor responding by showing that the effect is robust to multiple attention checks and comprehension questions; if anything, the effect is larger among participants who show signs of being more attentive. We rule out delayed-choice aversion in two ways: (1) participants understand that they will make a fixed number of decisions, and (2) we find that participants choose $\{M_H, M_L\}$ more often than they choose M_L when pitted against S, so they do not have an absolute aversion to delayed choices. In supplemental analyses, using a subset of choices for which M_H is chosen less than 50% of the time, we see that adding a less-attractive option further decreases choice share. This indicates undervaluation is not merely a regression to 50%.

Marketing Implications

The present research has a number of implications for marketing managers. First, it reinforces the importance of understanding the choice environment as construed by consumers. The same (e.g., trinary) choice may be construed as a two-stage decision by some consumers and as a single-stage decision by others. The current findings shed light on how dominated options differentially influence choice share depending on whether they are considered as options within multi-option alternatives or isolated alternatives. Thus, understanding how consumers structure the choice environment will permit more accurate predictions of consumer choice.

In addition to improving choice predictions, the current work also suggests ways to structure the choice environment with formal choice structures or informal choice cues to shape how the choice is framed. If marketers can establish a priori what a consumer would likely prefer within a given choice set, they may strategically structure a set to enhance options by excluding less-attractive options (or reduce choice share by including less-attractive options). Of course, the extent to which such practices are likely to be effective will depend on the precision with which marketers are able to predict consumers' choice shares: better prediction models will enable greater tailoring (see the next section on heterogeneity). This structuring is at odds with what one might predict from research on the attraction effect (Huber, Payne, and Puto, 1982). Whether the asymmetrically dominated option appears as a third option or as an additional component option within one alternative may reverse its effect on relative choice shares. In a trinary choice setting, the attraction effect benefits M_H because M_H dominates M_L and so M_H's choice share increases. In contrast, when M_H and M_L are framed as component options of M (as in all of our studies), choice share for M decreases relative to the simple binary choice between S and M_H. However, since we do not observe the attraction effect in our studies, this proposition is yet to be tested.

Finally, these findings reinforce the potential benefit for decision aids to encourage peering down the branches of the decision tree (Shafir 1994). As noted above, one approach to improve choice outcomes is to simply exclude options from consideration. Alternatively, decision aids may prompt consumers to consider second-stage choices when choosing a first-stage option (see Spiller and Ariely 2020). While our research thus far has focused on consumer decisions, it is plausible that similar effects manifest among managers as well.

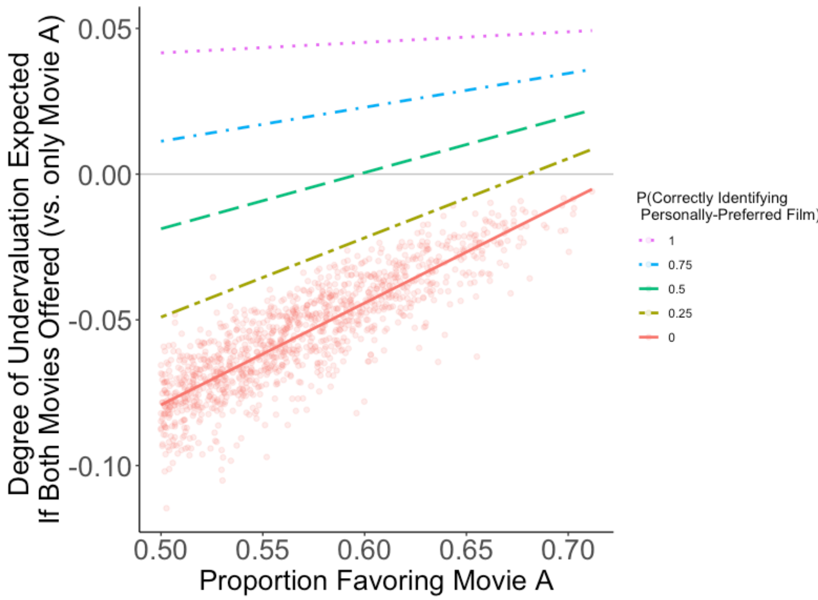
Impacts of Heterogeneity and Targeting

A key factor in assessing the presence of undervaluation is the heterogeneity of consumer preferences. With perfect information about each individual's preferences, offering only each consumer's most preferred option would lead to the highest choice share, regardless of the underlying heterogeneity in preferences. However, marketers regularly have incomplete information. Using the data from Studies 1 and 1b, we simulated several scenarios with varying amounts of heterogeneity and ability to target consumers based on their individual preferences;

details are given in the supplement. Consider the subset of film pairs (e.g., {The Shawshank Redemption, Legally Blonde}) where one film is favored by 60% to 70% of consumers (i.e., the right half of figure 8). If a marketer has no information about the preferences of the consumers (i.e., has no ability to target consumers; the solid red line in figure 8), then there would be a benefit to offering both films (vs. only one): they could expect roughly a 3 percentage-point increase in take-up by the consumers if both films are available as options. If a marketer has more information (i.e., can correctly target half of the consumers; the dashed green line in figure 8), then it is better (by roughly 1 percentage-point) to provide consumers with only the predicted-to-be-preferred film (vs. both). If a marketer has even more complete information (i.e., can correctly target 75% of the consumers; the blue dash-dot line in figure 8), then it is even better (by roughly 3 percentage-points) to provide only the predicted-to-be-preferred film to each consumer.

The benefit to providing additional options depends on the extent to which the additional options are preferred by part of the population. But the effects of adding options to an alternative will depend not only on the heterogeneity of the population, but also the targeting capabilities of the marketer. In other words, the degree of benefit depends on preference heterogeneity in the population and the ability to target different segments. Indeed, in this simulation, a 75% majority is equivalent to 75% absolute accuracy in predicting preferences. Heterogeneity of preferences and targeting ability are critical moderators to consider in assessing the extent of undervaluation when perfect personalization is not possible.

Figure 8: Heterogeneity and Undervaluation.



Notes: Relationship between heterogeneity and expected undervaluation of a multi-option alternative, as a function of knowledge or targeting ability. Expected undervaluation increases with a decrease in heterogeneity (i.e., as preferences become more homogenous) and as knowledge or targeting ability increases. Lines for {0.25, 0.5, 0.75, 1} represent best-fit lines to our simulations. See the supplements for additional details.

Future Directions

Several unanswered questions about this phenomenon remain. Although our studies find process evidence for the effect regarding attention to options, these results are correlational. In order to establish a causal path from information acquisition to decisions in this domain, future research would need to manipulate the information acquisition. (There is evidence from other domains that attention is causally related to choices, but we do not present causal evidence for that here; see Gwinn et al. 2019; Armel et al. 2008; Mormann et al. 2012; Pärnamets et al. 2015). Furthermore, though incentive-compatible, the stakes in Studies 4 and 5 were admittedly modest.

Finally, an interesting puzzle in the present paper is the moderation by transparency. This effect was strong and precisely estimated in Study 5, but was not significant in Studies 4b and 4c reported in the supplement. There were several key differences between the supplemental studies (4b-c) and the main text study (5). Most notably, information about the options was only visible when the participants moused over each box in Study 5, but not in the others. The more-directed information acquisition method may have better enabled consumers to edit out the transparently dominated option.

A promising avenue for future research would include consideration of which consumers are more likely to exhibit stronger effects. We began to explore this in studies reported in the supplements, but our initial findings did not yield diagnostic results. We hypothesized that consumers who elaborated more extensively on potential outcomes (Nenkov et al. 2009) or who use more analytic rather than holistic approaches to thinking through decisions (Choi et al. 2007) would exhibit less undervaluation. However, neither of these hypotheses held. Given the finding of systematic heterogeneity in Study 5, identifying the source of such heterogeneity would be informative.

Conclusion

The present research identifies and investigates a curious pattern of choice results: decision makers often choose against their best interests when confronted with a multi-option alternative. This finding has important implications for consumers and practitioners. Offering a multi-option alternative may appear like an attractive approach to increase choice for one's product(s) as it enables appeal to multiple segments of consumers with heterogeneous preferences. But if the options within the alternative are discrepant in value, our results suggest that choice likelihood for that alternative may be lower than it would be otherwise for any given consumer. It is not clear from the present research whether knowledge of this tendency would reduce or reverse the undervaluation effect, but it is important for marketing managers and choice architects to be aware of this way in which consumers' choices defy value-maximizing expectations.

REFERENCES

Amir, O., and Ariely, D. (2007). Decisions by rules: The case of unwillingness to pay for beneficial delays. *Journal of Marketing Research*, *44*, 142–152.

Armel, K. C., Beaumel, A., and Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 396–403.

Atalay, A. S., Bodur, H. O., and Rasolofoarison, D. (2012). Shining in the center: Central gaze cascade effect on product choice. *Journal of Consumer Research*, *39*(4), 848–866.

Bhui, R., and Jiao, P. (2023). Attention constraints and learning in categories. *Management Science, 0*(0).

Bown, N. J., Read, D., and Summers, B. (2003). The lure of choice. *Journal of Behavioral Decision Making*, *16*(4), 297–308.

Brehm, J. W. (1966). *A theory of psychological reactance.* (pp. x, 135). Academic Press.

Brough, A. R., and Chernev, A. (2012). When Opposites Detract: Categorical Reasoning and Subtractive Valuations of Product Combinations. *Journal of Consumer Research*, *39*(2), 399–414.

Busemeyer, J., and Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, *43*(3), 345–370.

Busemeyer, J., and Townsend, J. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459.

Chandon, P., Hutchinson, J. W., Bradlow, E. T., and Young, S. H. (2020). Does In-Store Marketing Work? Effects of the Number and Position of Shelf Facings on Brand Attention and Evaluation at the Point of Purchase. *Journal of Marketing*, *73*(6), 1–17.

Chernev, A., and Gal, D. (2010). Categorization Effects in Value Judgments: Averaging Bias in

Evaluating Combinations of Vices and Virtues. *Journal of Marketing Research*, *47*(4), 738–747.

Choi, I., Koo, M., and Jong An Choi. (2007). Individual Differences in Analytic Versus Holistic Thinking. *Personality and Social Psychology Bulletin*, *33*(5), 691–705.

Cohen, A. L., Kang, N., and Leise, T. L. (2017). Multi-attribute, multi-alternative models of choice: Choice, reaction time, and process tracing. *Cognitive Psychology*, *98*(Supplement C), 45–72.

Dai, J., and Busemeyer, J. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, *143*(4), 1489–1514.

Diehl, K. (2005). When Two Rights Make a Wrong: Searching Too Much in Ordered Environments. *Journal of Marketing Research*, *42*(3), 313–322.

Evangelidis, I., Jung, M., and Moon, A. (2022). When willingness-to-pay seems irrational: The role of perceived market price. [Conference presentation]. ACR 2022, Denver, CO, United States.

Fiedler, S., and Glöckner, A. (2012). The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology*, *3*.

Fisher, G. (2017). An attentional drift diffusion model over binary-attribute choice. *Cognition*, *168*, 34.

Fisher, G. (2021). Intertemporal Choices Are Causally Influenced by Fluctuations in Visual Attention. *Management Science*, mnsc.2020.3732.

Gaeth, G. J., Levin, I. P., Chakraborty, G., and Levin, A. M. (1991). Consumer evaluation of multi-product bundles: An information integration analysis. *Marketing Letters*, *2*(1), 47–57.

Gluth, S., Kern, N., Kortmann, M., and Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour*.

Gneezy, U., List, J. A., & Wu, G. (2006). The uncertainty effect: When a risky prospect is valued less than its worst possible outcome. *The Quarterly Journal of Economics*, *121*(4), 1283-1309.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, *69*(4), 623-638.

Gwinn, R., Leber, A. B., and Krajbich, I. (2019). The spillover effects of attentional learning on value-based choice. *Cognition*, *182*, 294–306.

Hauser, J. R. (1986). Agendas and Consumer Choice. *Journal of Marketing Research*, *23*(3), 199–212.

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

Johnson, E. J., Payne, J. W., Bettman, J. R., and Schkade, D. A. (1989). *Monitoring Information Processing and Decisions: The Mouselab System:* Defense Technical Information Center.

Juechems, K., Balaguer, J., Ruz, M., and Summerfield, C. (2017). Ventromedial Prefrontal Cortex Encodes a Latent Estimate of Cumulative Reward. *Neuron*, *93*(3), 705-714.e4.

Kahn, B. E., and Lehmann, D. R. (1991). Modeling Choice Among Assortments. *Journal of Retailing*, *67*(3), 274–299.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697.

Kim, B. E., Seligman, D., and Kable, J. W. (2012). Preference reversals in decision making under risk are accompanied by changes in attention to different attributes. *Frontiers in*

Neuroscience, *6*.

Konovalov, A., and Krajbich, I. (2020). Mouse tracking reveals structure knowledge in the absence of model-based choice. *Nature Communications*, *11*(1).

Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., and Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Le Lec, F., and Tarroux, B. (2019). On Attitudes to Choice: Some Experimental Evidence on Choice Aversion. *Journal of the European Economic Association*.

Lichtenstein, S., and Slovic, P. (1971). Reversals of Preference Between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology*, *89*(1), 46–55.

Lim, S.-L., O'Doherty, J. P., and Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *The Journal of Neuroscience*, *31*(37), 13214–13223.

Mislavsky, R., & Simonsohn, U. (2018). When risk is weird: Unexplained transaction features lower valuations. *Management Science*, *64*(11), 5395-5404.

Mochon, D. (2013). Single-Option Aversion. *Journal of Consumer Research*, *40*(3), 555–566.

Mormann, M. M., Navalpakkam, V., Koch, C., and Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, *22*(1).

Newell, B. R., and Le Pelley, M. E. (2018). Perceptual but not complex moral judgments can be biased by exploiting the dynamics of eye-gaze. *Journal of Experimental Psychology:*

General, *147*(3), 409–417.

O'Donnell, M., and Evers, E. R. K. (2019). Preference Reversals in Willingness to Pay and Choice. *Journal of Consumer Research*, *45*(6), 1315–1330.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., and Hertwig, R. (2018). Prospect Theory Reflects Selective Allocation of Attention. *Journal of Experimental Psychology: General*, *147*(2), 147–169.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., and Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, *112*(13), 4170–4175.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–387.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552.

Pieters, R., and Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing*, *16*(1), 1–16.

Plott, C. R., and Levine, M. E. (1978). A Model of Agenda Influence on Committee Decisions. *The American Economic Review*, *68*(1), 146–160.

Reeck, C., Wall, D., and Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences*, *114*(45), 11890–11895.

Roe, R. M., Busemeyer, J., and Townsend, J. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370–392.

Shenhav, A., and Karmarkar, U. R. (2019). Dissociable components of the reward circuit are involved in appraisal versus choice. *Scientific Reports*, *9*(1), 1958.

Shin, J., and Ariely, D. (2004). Keeping Doors Open: The Effect of Unavailability on Incentives to Keep Options Viable. *Management Science*, *50*(5), 575–586.

Simonsohn, U. (2009). Direct risk aversion: Evidence from risky prospects valued below their worst outcome. *Psychological Science*, *20*(6), 686-692.

Simonson, I., Carmon, Z., and O'Curry, S. (1994). Experimental Evidence on the Negative Effect of Product Features and Sales Promotions on Brand Choice. *Marketing Science*, *13*(1), 23–40.

Smith, S. M., and Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*, *147*(12), 1810–1826.

Sood, S., Rottenstreich, Y., and Brenner, L. (2004). On Decisions That Lead to Decisions: Direct and Derived Evaluations of Preference. *Journal of Consumer Research*, *31*(1), 17–25.

Spiller, S. A., and Ariely, D. (2020). How does the perceived value of a medium of exchange depend on its set of possible uses? *Organizational Behavior and Human Decision Processes*, *161*, 188–200.

Stillman, P. E., Krajbich, I., and Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences. *Proceedings of the National Academy of Sciences*, *117*(50), 31738.

Stillman, P. E., Shen, X., and Ferguson, M. J. (2018). How Mouse-tracking Can Advance Social Cognitive Theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Thomas, A., Molter, F., and Krajbich, I. (2020). *Uncovering the Computational Mechanisms Underlying Many-Alternative Choice* [Preprint]. PsyArXiv.

Troutman, C. M., and Shanteau, J. (1976). Do Consumers Evaluate Products by Adding or Averaging Attribute Information? *Journal of Consumer Research*, *3*(2), 101.

Tversky, A., and Sattath, S. (1979). Preference trees. *Psychological Review*, *86*(6), 542–573.

Vaidya, A. R., and Fellows, L. K. (2015). Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature Communications*, *6*, 10120.

Weaver, K., Garcia, S. M., and Schwarz, N. (2012). The Presenter's Paradox. *Journal of Consumer Research*, *39*(3), 445–460.

Yadav, M. S. (1994). How Buyers Evaluate Product Bundles: A Model of Anchoring and Adjustment. *Journal of Consumer Research*, *21*, 342–353.

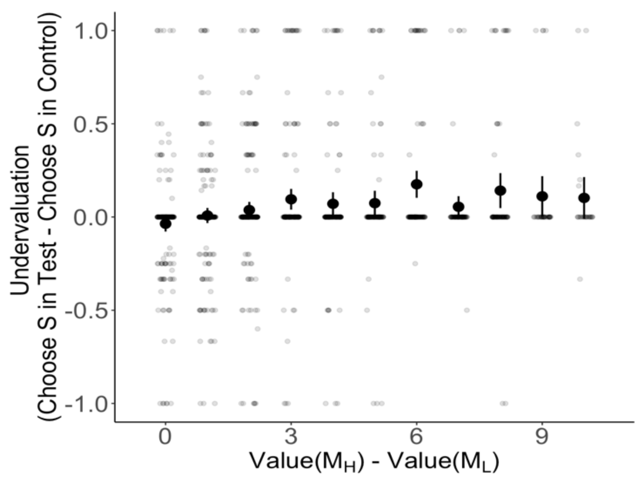
Yang, Y., Vosgerau, J., & Loewenstein, G. (2013). Framing influences willingness to pay but not willingness to accept. *Journal of Marketing Research*, *50*(6), 725-738.

Supplementary Material for

Consumers Undervalue Multi-Option Alternatives in Two-Stage Choice

Experiment 1

Figure S1: Preference Strength and Undervaluation.



Notes: Relationship between (M_H>M_L) preference strength and degree of undervaluation. Undervaluation increases as M_H gets progressively better than M_L. Bars represent s.e.m. across participants.

Experiment 1b

Method

Participants

For this preregistered experiment (https://aspredicted.org/blind.php?x=wi4xk7), we collected responses from 604 Amazon Mechanical Turk workers; each earned $1.25 for participating.

Materials and Procedure

The materials and procedure were identical to Experiment 1.

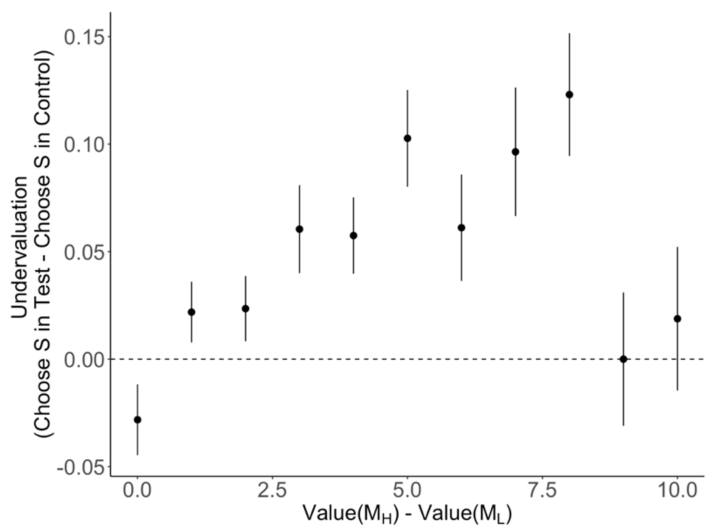
Exclusions and Data Preprocessing

We followed the same exclusion rule used in Experiment 1. These criteria resulted in the exclusion of 80 participants, leaving us with a sample size of 524.

Results (Preregistered)

Using the same approach as in Experiment 1, we found evidence for undervaluation, M = 0.04, 95% CI = [0.03, 0.05], $t(523) = 6.75$, $p < .001$. Using that approach, we find that 41% of participants exhibited undervaluation, with 37% exhibiting no difference and 22% exhibiting an effect in the opposite direction. As in Experiment 1, we find significant evidence for an effect of M_H-M_L preference strength on undervaluation, $b = 0.011$, $SE = 0.002$, $p < .001$ (Fig. S2).[9]

Figure S2: Preference Strength and Undervaluation



Notes: Relationship between (M_H-M_L) preference strength and degree of undervaluation. Undervaluation increases as M_H gets progressively better than M_L. Analysis reported in text

[9] Due to a since-addressed coding error in an early analysis of Experiment 1, we collected this dataset with additional preregistered exclusions. We specified that we would remove all choices where the M_H-M_L difference was less than 5. With that exclusion, we find a larger effect, M = 0.09, 95% CI = [0.06, 0.12], $t(335) = 6.64$, $p < .001$.

controls for rating of S and sum of ratings of M_H and M_L. Bars represent s.e.m. across participants.

Experiment 2b

Method

Participants

For this preregistered experiment (https://aspredicted.org/ZCK_M34), we collected responses from 201 Amazon Mechanical Turk (AMT) workers. They earned $1.50 for their participation.

Materials and Procedure

The procedure for this experiment was identical to Experiment 2.

Exclusions and Data Preprocessing

As specified in our preregistration, we excluded anyone who failed to rate at least 20 films. We also excluded anyone whose filler choices were not directionally predicted by their ratings in a logistic regression of *ChooseLeft* on (*RatingLeft-RatingRight*). These criteria resulted in the exclusion of 24 participants, leaving us with a final sample size of 177. As specified in our preregistration, we excluded any test-control choice pairs that were generated from unrated films.

Results

As in previous experiments, we find evidence of undervaluation, $M = 0.03$, 95% CI = [0.01, 0.04], $t(176) = 2.94$, $p = .003$. Next, we make several comparisons between ratings. First, we look at whether the rating of S changes, depending on whether it is part of a test choice or a control choice, and find that it was rated marginally significantly lower following binary than test ($M = -0.45$, 95% CI = [-0.93, 0.04], $t(175) = 0.069$). Next, we make within-subject comparisons among three values: (i) the rating of M_H following control choices, (ii) the rating of M_H following test choices and (iii) the rating of $\{M_H,M_L\}$ following test choices, using one-sample t-tests.

To reduce sampling variability in the selected movies we controlled for differences in how those movies were rated by those participants in their initial ratings. We again find that $\{M_H,M_L\}$ is rated lower following test than is M_H following binary ($M = -0.67$, 95% CI = [-1.00, -0.34], $t(174) = 4.01$, $p < .001$). But in this more powerful analysis, M_H is rated marginally significantly lower following test than following binary ($M = -0.28$, 95% CI = [-0.57, .005], $t(174) = 1.94$, $p = .054$) and $\{M_H,M_L\}$ is rated significantly lower following test than M_H is following test ($M = -0.37$, 95% CI = [-0.73, -0.01], $t(174) = 2.02$, $p = .044$).

In our preregistration for this study (which was collected before Experiment 2), we did not specify whether we would control for the initial ratings of the films. When we do not control for them, we find that the rating of $\{M_H,M_L\}$ following test choices is significantly lower than the rating of M_H following binary choices ($M = -0.60$, 95% CI = [-1.06, -0.14], $t(175) = 2.59$, $p = .010$) and that the rating of M_H following test choices is not significantly lower than the rating of M_H following binary choices ($M = -0.29$, 95% CI = [-0.73, 0.14], $t(175) = -1.33$, $p = .186$). The rating of $\{M_H,M_L\}$ following test choices is not significantly lower than the rating of M_H following test choices ($M = -0.31$, 95% CI = [-0.74, 0.13], $t(175) = -1.39$, $p = 0.166$).

In sum, it is clear that $\{M_H,M_L\}$ is rated lower following test than M_H is following binary. Given the most powerful test (accounting for sampling variability of differences in movies), $\{M_H,M_L\}$ is rated lower following test than M_H is following test, suggesting that part of the undervaluation effect is indeed due to undervaluation of the holistic option, not just the more-preferred option in the presence of the other option.

Experiment 3b

Method

Participants

For this preregistered experiment (https://aspredicted.org/R6L_9LV), we collected responses from 517 AMT workers. They earned $1.50 for their participation.

Materials and Procedure

The procedure was identical to Experiment 3.

Exclusions and Data Preprocessing

As specified in our preregistration, we excluded anyone who failed to rate at least 18 films, as they would have fewer usable primary choices and thus noisier estimates. We also excluded anyone whose simple binary choices were not directionally predicted by their ratings in a logistic regression of *ChooseLeft* on (*RatingLeft-RatingRight*) as that indicates lack of minimal attention during ratings, choice, or both. These criteria resulted in the exclusion of 158 participants, leaving us with a final sample size of 359.

Also in line with our preregistration, we excluded any choice sets that were generated from unrated films.

Results

In this study, we can test for undervaluation of several theaters. First, analogous to analyses in Experiments 1, 1b, 2, and 2b, we can compare choices in 1v1 and 1v2 trials to test for undervaluation of the 2-movie theater in 1v2 choices. To test for undervaluation in this case, for every participant, we calculated $S_A > \{M_A, M_B\}$ as (Chose S in 1v2 choices / number of valid 1v2 choices). We also calculated $S_A > M_A$ as (Chose S in 1v1 choices / number of valid 1v1 choices). We tested $(S_A > \{M_A, M_B\}) – (S_A > M_A)$ using a one-sample t-test and found evidence for undervaluation, $M = 0.05$, 95% CI = [0.03, 0.07], $t(655) = 4.95$, $p < .001$. In other words, choice of M_A was 5 percentage points higher than choice of $\{M_A, M_B\}$.

We used the analogous approach to compare 1v1 against 1v3, 1v2 against 2v2, and 2v2 against 2v3. In each case, we find undervaluation (1v1 vs. 1v2 trials: $M = 0.08$, 95% CI = [0.04, 0.12], $t(283) = 3.83$, $p < .001$; 1v1 vs. 1v3 trials: $M = 0.08$, 95% CI = [0.03, 0.12], $t(224) = 3.23$, $p = .001$; 1v2 vs. 2v2 trials: $M = 0.07$, 95% CI = [0.03, 0.12], $t(270) = 3.02$, $p = .003$; 2v2 vs. 2v3 trials: $M = 0.06$, 95% CI = [0.01, 0.12], $t(180) = 2.16$, $p = .03$.).

For this experiment (which was collected prior to Experiment 3), we did not preregister the requirement that the added sub-option be rated at two points lower than the existing sub-option(s), as we did in Experiment 3. Without that requirement, only the first two comparisons are significant: 1v1 vs. 1v2 trials: $M = 0.04$, 95% CI = [0.01, 0.06], $t(358) = 2.95$, $p = .003$; 1v1 vs. 1v3 trials: $M = 0.02$, 95% CI = [–0.002, 0.05], $t(358) = 1.80$, $p = .07$; 1v2 vs. 2v2 trials: $M = 0.004$, 95% CI = [–0.02, 0.03], $t(358) = 0.37$, $p = .72$; 2v2 vs. 2v3 trials: $M = –0.03$, 95% CI = [–0.06, 0.01], $t(345) = –1.62$, $p = .11$.

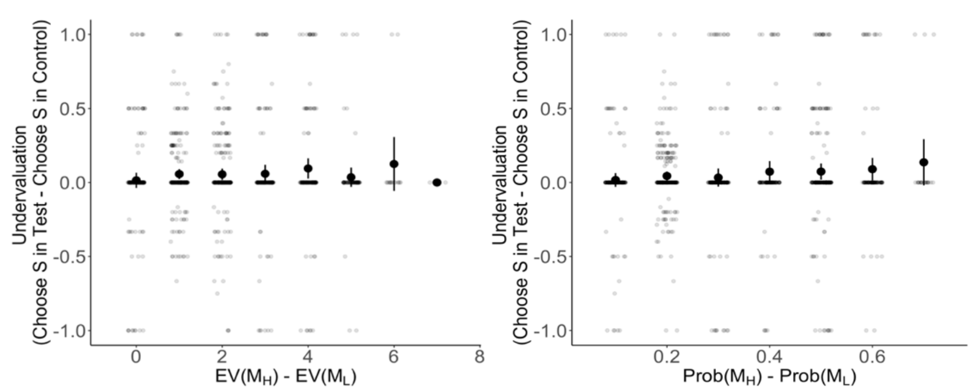
Experiment 4

Trial Generation Process

To generate the 34 trials, we started with the set of all possible events, separated by domain (i.e., cards, dice, and coins) and probability. In the card domain, we only used suits/colors (e.g., drawing a black card, drawing a spade, drawing a red card or a club). In the dice domain, we used every combination of numbers (e.g., rolling a 1, rolling a 1 or 2 or 4 or 6, rolling a 2 or 3 or 5). In the coin domain, we used every outcome of flipping one or two coins (e.g., flipping heads on one coin, flipping tails on two coins, flipping at least one head on two coins). We then randomly selected, for each participant and with replacement, three domains and three probabilities within those domains to fill the roles of S, M_H, and M_L. The only constraint we placed on the assignment of domains/probabilities to the three roles (S, M_H, and M_L) was to require that the probability of M_H be greater than the probability of M_L. If this was not possible (i.e., if the sampled probabilities were equal), then we resampled all three roles. Once the domains and probabilities were established, to populate binary control, test, and trinary choices, we randomly sampled with replacement three events for S, three events for M_H, and two events for M_L. For instance, if the domain for S was dice and the probability for S was 0.5, then we sampled three events (with replacement) from the entire list of possible dice events with probability 0.5 (e.g., rolling a 1 or 2 or 3; rolling a 3 or 5 or 6; rolling a 1 or 3 or 4). To determine the monetary amounts for each option, we randomly selected (with replacement) two dollar amounts ranging from \$2 to \$10; the first was assigned to the S events and the second was assigned to M_H, M_L, and M_H or M_L.

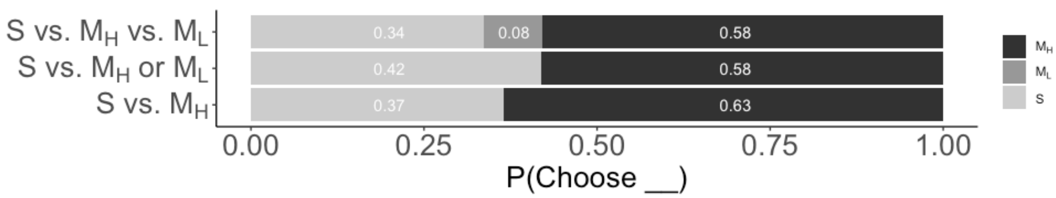
At the end of the survey, participants also completed a corresponding M_H vs. M_L trial from one of the 10 trial sets that they completed. Three attention check trials were used as detailed in Experiment 4b.

Figure S3: Dominance Strength and Undervaluation



Notes: Relationship between (M_H>M_L) dominance strength and undervaluation. Across both definitions of dominance ((a) expected value, (b) probabilities), undervaluation increases as M_H gets progressively better than M_L, though this relationship is not significant for expected value (p = .16) and marginally significant for probability (p = .07). Bars represent s.e.m. across participants.

Figure S4: Proportion of Choices



Notes: Proportion of choices for each option in binary, test, and trinary choices for Experiment 4.

Experiment 4b

Method

Participants

For this preregistered experiment (https://aspredicted.org/blind.php?x=rd495k), we collected responses from 298 Amazon Mechanical Turk workers. They earned $2.50 (the first 20 participants) or $1.50 (the remaining 278 participants) for their participation. Five randomly-selected participants received the outcome of one of their decisions, as detailed below.

Materials and Procedure

This study was very similar to Experiment 4, except for the following changes. First, the choices were not randomly generated as in Experiment 4. See below for trial generation information. The other important difference is that the M_H vs. M_L dominance relationship was designed to be either high-transparency or low-transparency as in Experiment 5. Participants made a total of 39 or 40 choices. At the end of the study, we randomly selected five participants, but none of them won anything from their randomly-selected decision.

Trial Generation

There are 10 trial sets (i.e., t = {1:10}). Each trial set *t* consists of three option types: S^t, M_H^t, and M_L^t each with an associated monetary winning (*m*, ranging from $2 to $10) and probability of winning (*p*, ranging from 0.167 to 0.75). Within a trial set, $m(M_H) = m(M_L)$ and $p(M_H) > p(M_L)$. Within each trial set, there are three trial varieties (*v*) that share *p* and *m* for each of {S^{tv}, M_H^{tv}, and M_L^{tv}}. However, the exact mechanisms may vary across varieties. For instance, if S^{t1} were "$5; Rolling a 2 or a 3 or a 4" then S^{t2} could be "$5; Flipping heads on one coin" and S^{t3} could be "$5; Drawing a black card." The dollar amounts and probabilities for each of the option types are consistent across varieties, but the exact mechanisms (e.g., coin flips vs. card draws vs. die rolls) can vary.

Within each trial variety, there are two levels of M transparency (i.e., high and low). For the high transparency level, M_{Hh} and M_{Lh} share a domain (e.g., coin flips vs. card draws vs. die rolls) and the set of probabilistic events that constitute M_{Lh} is contained in the set of probabilistic events that constitute M_{Hh}. For instance, if M_{Hh} were "$6; Rolling a 1 or 2 or 5 or 6" then M_{Lh} could be "$6; Rolling a 1 or 2 or 5." On the other hand, for the low transparency level, M_{Hl} and M_{Ll} did not share a domain and thus, M_{Ll} was not contained in M_{Hl}, but $p(M_{Hl})$ remained larger than $p(M_{Ll})$. For instance, if M_{Hl} were "$6; Rolling a 1 or 2 or 5 or 6" then M_{Ll} could be "$6; Drawing a black card." Therefore, there are 10 (trial sets) * 3 (trial varieties) * 2 (transparency levels) = 60 varieties of each of S, M_H, and M_L. A full list of possible trials is available in our Research Box.

Choices

For each of 10 trial sets t, a given participant completed the following trials: (1) low transparency: S^{t1} vs. $\mathrm{M}_{\mathrm{H}}^{\mathrm{t1}}$ or $\mathrm{M}_{\mathrm{L}}^{\mathrm{t1}}$ (L), where $\mathrm{M}_{\mathrm{H}}^{\mathrm{t1}}$ and $\mathrm{M}_{\mathrm{L}}^{\mathrm{t1}}$ use different mechanisms; (2) high transparency: S^{t2} vs. $\mathrm{M}_{\mathrm{H}}^{\mathrm{t2}}$ or $\mathrm{M}_{\mathrm{L}}^{\mathrm{t2}}$ (H), where $\mathrm{M}_{\mathrm{H}}^{\mathrm{t2}}$and $\mathrm{M}_{\mathrm{L}}^{\mathrm{t2}}$ use the same mechanism; and (3) control 1: binary choice between S^{t3} and $\mathrm{M}_{\mathrm{H}}^{\mathrm{t3}}$ (L), where the mechanism (card draw, die roll, or coin flip) for $\mathrm{M}_{\mathrm{H}}^{\mathrm{t3}}$ is the same as the mechanism for $\mathrm{M}_{\mathrm{H}}^{\mathrm{t1}}$ (L). If the mechanisms for $\mathrm{M}_{\mathrm{H}}^{\mathrm{t1}}$ (L) and $\mathrm{M}_{\mathrm{H}}^{\mathrm{t2}}$ (H) were different, then the participant completed a 4th trial for that set: (4) control 2: S^{t3} vs. $\mathrm{M}_{\mathrm{H}}^{\mathrm{t3}}$ (H), where the mechanism for $\mathrm{M}_{\mathrm{H}}^{\mathrm{t3}}$ (H) is the same as the mechanism for $\mathrm{M}_{\mathrm{H}}^{\mathrm{t2}}$ (H). Across participants, the assignment of varieties within each trial set (i.e., whether the i in S^{ti} is 1 or 2 or 3 for each of the four types of choices) was randomized. Additionally, participants were assigned to one of two between-subjects presentation conditions; the odd (even) numbered sets presented the "M_{H} or M_{L}" option as "M_{H} or M_{L}" and the even (odd) numbered sets presented the "M_{H} or M_{L}" option as "M_{L} or M_{H}" to control for presentation-order effects. In one case where S^{t3} was identical to $\mathrm{M}_{\mathrm{H}}^{\mathrm{t3}}$, the appropriate control trial was excluded (as we could use its noiseless expected value of 0.5 instead).

In addition to the main trials of interest, participants also completed three attention-check questions, in which one option was designed to be (trivially obviously) dominant over the other according to monetary winnings, probability, or both. These three trials were spaced so that they occurred roughly one-fourth, one-half, and three-fourths of the way through the survey. All of the main trials of interest were presented in a fully-randomized order.

At the end of the survey, participants also completed the two corresponding M_{H} vs. M_{L} trials (i.e., $\mathrm{M}_{\mathrm{H}}^{\mathrm{t1}}$ vs. $\mathrm{M}_{\mathrm{L}}^{\mathrm{t1}}$ and $\mathrm{M}_{\mathrm{H}}^{\mathrm{t2}}$ vs. $\mathrm{M}_{\mathrm{L}}^{\mathrm{t2}}$) from one of the 10 trial sets.

Exclusions and Data Preprocessing

As specified in our preregistration, we excluded anyone who picked a dominated option in any of the three attention-check questions. This resulted in the exclusion of 104 participants, leaving us with a sample size of 194.

Results

Results (Preregistered)

We analyzed the results as in Experiment 5 and found undervaluation, $M = 0.06$, 95% CI = [0.03, 0.09], $t(193) = 4.28$, $p < .001$. We tested for an effect of transparency as in Experiment 5 and find none, $M = 0.002$, 95% CI = [–0.02, 0.02], $t(193) = 0.12$, $p = 0.91$.

Using that approach, we find that 49% of participants exhibited undervaluation, 20% showed no difference, and 31% showed an effect in the opposite direction. On the high-transparency trials, 48% exhibited undervaluation, 21% exhibited no difference, and 30% showed an effect in the opposite direction. On the low-transparency trials, 51% exhibited undervaluation, 19% showed no difference, and 30% showed an effect in the opposite direction. At the choice set level, we find that all 10 of the choice sets exhibit undervaluation, 9/10 exhibit undervaluation on high-transparency trials, and all 10 exhibit undervaluation on low-transparency trials.

Exploratory Results

We tested for within-subject consistency in undervaluation by correlating the undervaluation in high-transparency trials with the undervaluation in low-transparency trials. We found a significant positive relationship, $r = 0.67$, $t(192) = 12.46$, $p < .001$. Participants who displayed more undervaluation in high-transparency trials also displayed more in low-transparency trials.

Experiment 4c

Introduction

In Experiment 4c, we investigate possible connections between undervaluation and various individual difference measures: risk aversion, analytic-holistic thinking (Choi et al. 2007), and elaboration on potential outcomes (Nenkov et al. 2008). We predicted that greater undervaluation would be associated with more holistic thinking (i.e., treating the "M_H or M_L" option as a holistic unit, rather than as its component parts) and less elaboration on potential outcomes.

In addition to investigating individual differences, we also sought to rule out some alternative explanations for the effect. First, it is possible that people choose S more when it is paired with "M_H or M_L" because people have a strong aversion to multi-option alternatives. If this were the case, we would expect them to be more likely to choose M_L (over S) than "M_H or M_L."

Method

Participants

For this preregistered experiment (https://aspredicted.org/blind.php?x=z6mx9f), we collected responses from 298 Amazon Mechanical Turk workers. They earned $2.50 for their participation. Five randomly-selected participants received the outcome of one of their decisions, as detailed below.

Materials and Procedure

This experiment was very similar to Experiment 4b, with the following changes. Participants completed S vs. M_L choices and trinary choices in addition to binary and multi-option choices. Moreover, at the end of the experiment, participants completed 5 incentivized risk aversion trials (Holt and Laury 2002). They also completed the analytic-holistic thinking scale (Choi et al. 2007) and the elaboration on potential outcomes scale (Nenkov et al. 2008). In this study, two of the five randomly-selected participants won money ($2 and $0.10).

Additional Trial Details

In this experiment, participants made 43 incentivized choices (including 3 attention check questions and 5 risk aversion measures); some were control binary choices (e.g., S vs. M_H), some were trinary choices (e.g., S vs. M_H vs. M_L) and some were test choices (e.g., S vs. M_H or M_L). For each of 10 trial sets *t*, a given participant completed the following trials: (1) low transparency: S^{t1} vs. M_H^{t1} or M_L^{t1} (L), where M_H^{t1} and M_L^{t1} use different mechanisms; (2) high transparency: S^{t2} vs. M_H^{t2} or M_L^{t2} (H), where M_H^{t2} and M_L^{t2} use the same mechanism; (3) trinary: S^{t3} vs. M_H^{t3} vs. M_L^{t3}; (4) SM_L: S^{t1} vs. M_L^{t1} (or S^{t2} vs. M_L^{t2}); (5) M_HM_L: M_H^{t2} vs. M_L^{t2} (or M_H^{t1} vs. M_L^{t1}); (6) control 1: binary choice between S^{t3} and M_H^{t3} (L), where the mechanism (card draw, die roll, or coin flip) for M_H^{t3} (L) is the same as the mechanism for M_H^{t1} (L). If the mechanisms for M_H^{t1} (L) and M_H^{t2} (H) were different, then the participant completed a seventh trial for that set: (7) control 2: S^{t3} vs. M_H^{t3} (H), where the mechanism for M_H^{t3} (H) is the same as the mechanism for M_H^{t2} (H).

Participants were assigned to either complete trial sets 1-5 or trial sets 6-10. For trial types 3 and 4 (i.e., trinary and SM_L), participants were randomly assigned in each set to complete the first or second versions of the trials, as detailed above. The version for M_HM_L was set to be the opposite of the version participants completed for SM_L. The full list of possible trials is available in our Research Box.

At the end of the experiment, participants completed the two corresponding M_H vs. M_L trials from one of the five trial sets that they completed. They also completed 5 risk aversion trials (Holt and Laury 2002). Finally, they completed the analytic-holistic thinking scale (Choi et al. 2007) and the elaboration on potential outcomes scale (Nenkov et al. 2008). As in Experiment 4, at the end of data collection, we randomly selected five participants and then randomly selected one of their choices to play out. In this experiment, two of the five participants won money ($2 and $0.10).

Exclusions and Data Preprocessing

As specified in our preregistration, we excluded anyone who picked the dominated option in any of the three attention-check questions. This resulted in the exclusion of 122 participants, leaving us with a sample size of 176.

Results

Preregistered Results

We tested for undervaluation using the approach in Experiment 4b, and find evidence for undervaluation, $M = 0.06$, 95% CI = [0.02, 0.09], $t(175) = 3.37$, $p < .001$. We tested for moderation by transparency as in Experiment 2b and do not find evidence of the moderation, $M = 0.002$, 95% CI = [–0.04, 0.04], $t(175) = 0.11$, $p = 0.91$. 55% of participants exhibited undervaluation, 19% exhibited no difference, and 27% exhibited an effect in the opposite direction. 41% exhibited undervaluation on high-transparency trials, and 49% exhibited undervaluation on low-transparency trials. At the choice set level, 8/10 of the choice sets exhibit undervaluation overall, 8/10 exhibit undervaluation on high-transparency trials, and 8/10 exhibit undervaluation on low-transparency trials.

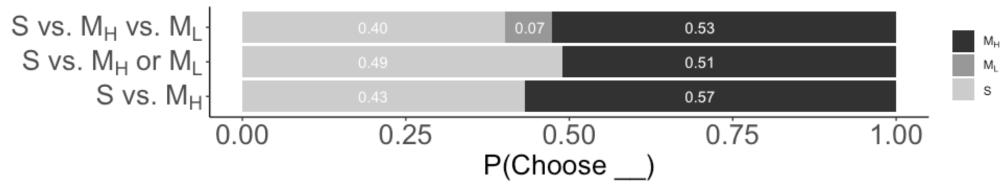
We did not find any evidence of a relationship between undervaluation and risk aversion ($r = 0.003$, $t(174) = 0.03$, $p = 0.97$) or elaboration on potential outcomes ($r = 0.007$, $t(174) = 0.10$, $p = 0.92$). We found a small negative correlation between undervaluation and analytic/holistic thinking (r = –0.17, $t(174)$ = –2.28, p = 0.02), which suggests that participants who reported being more holistic thinkers displayed less undervaluation. However, as this relationship is small, in the opposite direction as anticipated, and was one of several possible measured individual differences (increasing the probability of false positives), we do not strain to interpret it here.

With this data, we were also able to address two possible alternative explanations for our results. First, one possible explanation is that the undervaluation effect is due to blind avoidance of "M_H or M_L" options. Comparing choice proportions in the multi-option alternative choices to choice proportions in the S vs. M_L choices allows us to address this possibility. People choose "M_H or M_L" more often than they choose M_L, which indicates that they do not blindly avoid the "M_H or M_L" options, M = 0.18, 95% CI = [0.15, 0.22], $t(175)$ = 10.93, p < .001.

Exploratory Results

Another alternative explanation is that our results are simply due to noisy and/or careless responding. However, when we limit our dataset to participants who never chose a dominated option (i.e., chose M_H in all of the M_H vs. M_L choices), the effect is still there and is even larger, M = 0.11, $t(91)$ = 5.57, p < .001. Similarly, when we further restrict our dataset to participants who also got all of the comprehension questions correct on the first try, we find a similar (stronger) effect, M = 0.10, $t(67)$ = 4.61, p < .001. Thus, when we limit our dataset to more attentive participants, the effect does not get smaller and, if anything, it gets larger.

Figure S5: Proportion of Choices.



Notes: Proportion of choices for each option in binary, test, and trinary choices for Experiment 4c.

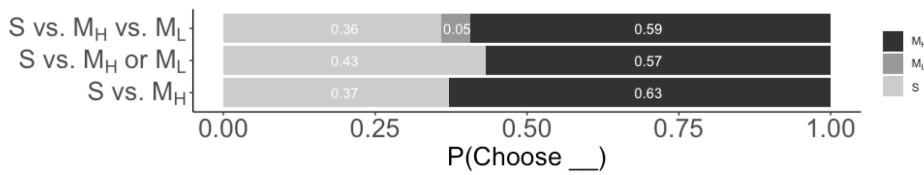
Experiment 5

Trial Generation Process

In this experiment, participants made 65 incentivized choices. Some were control binary choices (e.g., S vs. M_H), some were trinary choices (e.g., S vs. M_H vs. M_L) and some were test choices (e.g., S vs. {M_H, M_L}). For each of 10 trial sets *t*, a given participant completed the following trials: (1) low transparency: S^{t1} vs. M_H^{t1} or M_L^{t1} (L), where M_H^{t1} and M_L^{t1} use different mechanisms; (2) high transparency: S^{t2} vs. M_H^{t2} or M_L^{t2} (H), where $M_H^{t2}$and M_L^{t2} use the same mechanism; (3) trinary: S^{t3} vs. M_H^{t3} vs. M_L^{t3}; (4) control 1: binary choice between S^{t3} and M_H^{t3}, where the mechanism (card draw, die roll, or coin flip) for M_H^{t3} is the same as the mechanism for M_H^{t1}. If the mechanisms for M_H^{t1} and M_H^{t2} were different, then the participant completed a fifth trial for that set: (5) control 2: S^{t3} vs. M_H^{t3}, where the mechanism for M_H^{t3} is the same as the mechanism for M_H^{t2}.

Participants completed all 10 trial sets. At the end of the survey, participants completed a rank-ordering choice (of S, M_H, and M_L) within each set with the same mechanisms as (1), above. They also completed the two corresponding M_H vs. M_L trials (i.e., M_H^{t1} vs. M_L^{t1} and M_H^{t2} vs. M_L^{t2}) from one of the trial sets that they completed. As in Experiment 4, at the end of data collection, we randomly selected five participants and then randomly selected one of their choices to play out. In this study, all five participants won money ($3, $5, $4, $2 and $4).

Figure S6: Proportion of Choices.



Notes: Proportion of choices for each option in binary, test, and trinary choices for Experiment 5.

Heterogeneity Simulation

To simulate behavior in multi-option alternative decisions, we used the data from Experiments 1 and 1b. First, using all control binary choices from the data, we estimated the probability of choosing a film in a binary choice, based on the rating of the film and the rating of its competitor (i.e., using logistic regression we regressed ChooseFilmM_H on RatingS and RatingM_H). Then, using all test choices from the data, we estimated the probability of choosing a multi-option alternative (i.e., M_HM_L), based on the ratings of the single-option alternative and both options of the multi-option alternative (i.e., using logistic regression we regressed ChooseM_HM_L on RatingS, RatingM_H, and RatingM_L). We use these results to estimate the probabilities of choosing single-option and multi-option alternatives as detailed below.

Next, for each unique pair of films (50*49/2 = 1225 pairs), we identified the more-popular film (i.e., the film that was rated higher by more participants, which we will call "Film A") and the less-popular film (i.e., whichever film is not Film A, which we will call "Film B"). For each unique pair of films for each participant, there is also a personally-preferred film (which we call "Film M_H") and a personally-unpreferred film (which we call "Film M_L").

We also identified (using participant-level ratings of S, M_H, and M_L and the estimates from the regressions identified above) at the participant level: (1) the probability that Film A would be chosen, (2) the probability that Film B would be chosen, (3) the probability that Film M_H would be chosen, (4) the probability that Film M_L would be chosen, and (5) the probability that a multi-option alternative comprising both films (Films A and B; a.k.a. Films M_H and M_L) would be chosen. We then averaged each of the five measures above across participants to get average probabilities for each pair of films.

Next, we considered the results assuming different levels of targeting capabilities. First, we consider the situation where we have no ability to target consumers based on their preferences. In this case including both films as options maximizes probability of choice. In other words, we expect "overvaluation" of the multi-option alternative, relative to the more-popular option (Figure 8). Second, we consider the situation where we are able to identify the personally-preferred film for a given consumer with some probability (*w*). In this case, we expect to observe probability (3) above in *w**100% of consumers and observe probability (1) above in (1-*w*)*100% of consumers on average. To represent this in our simulations, we randomly selected which consumers' preferences would be accurately identified (using a binomial distribution with probability *w*). We repeated this random selection 1000 times, averaging at the movie-pair level. As *w* increases, and as offerings are tailored to individuals, expected undervaluation of the multi-option alternative increases as well.

Table S1: Summary of Choice Trials Across All Experiments.

Exp.	Total Number of Choices	S vs. M_H or M_L	S vs. M_H	S vs. M_H vs. M_L	S vs. M_L	M_H vs. M_L	Attention Check Questions	Other
1 1b	30	10	10	0	0	0	0	10 filler binary questions
2 2b	30	10	10	0	0	0	0	10 filler binary questions; post-choice ratings: S, M, M_H, M_L
3 3b	42	5	5	0	0	0	0	17 filler binary questions, 5 1v3, 5 2v2, 5 2v3
4	34	10	10	10	0	1	3	
4b	39-40	20	14-15	0	0	2	3	
4c	43	10	10	5	5	5	3	risk aversion, analytic / holistic thinking, elaboration on potential outcomes
5	65	20	20	10	0	2	3	10 rank-order questions

Table S2: Full Regression Results.

Exp.	Equation Number	Equation	Coefficients	SE	t value	p-value
1	1	Choice Difference ~ b0 + b1*S + b2*(MH+ML) + b3*(MH-ML)	b0 = –0.05 b1 = 0.01 b2 = –0.0004 b3 = 0.02	0.02 0.002 0.001 0.003	–2.64 3.40 –0.27 5.70	.008 <.001 .79 <.001
4	2	Choice Difference ~ b0 + b1*EV(S) + b2*(EV(MH)+EV(ML)) + b3*(EV(MH)-EV(ML))	b0 = 0.04 b1 = –0.005 b2 = 0.001 b3 = 0.01	0.03 0.004 0.003 0.008	1.64 –1.22 0.38 1.39	.10 .22 .71 .16
4	3	Choose S ~ b0 + b1*EV(S) + b2*EV(MH) + b3*M(S) + b4*M(MH) + b5*P(S) + b6*P(MH)	b0 = –0.25 b1 = 0.42 b2 = –0.24 b3 = 0.19 b4 = –0.28 b5 = 2.96 b6 = –2.83	0.72 0.12 0.15 0.06 0.10 0.77 0.91	–0.35 3.49 –1.61 3.09 –2.98 3.83 –3.12	.73 <.001 .11 .002 .003 <.001 .002
4	4	Choice Difference ~ b0 + b1*P(S) + b2*(P(MH)+P(ML)) + b3*(P(MH)-P(ML))	b0 = 0.02 b1 = –0.01 b2 = 0.004 b3 = 0.11	0.04 0.04 0.03 0.06	0.57 –0.40 0.14 1.80	.57 .69 .89 .07