UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"

PhD SCHOOL

PhD program in Business Administration and Management

Cycle: 34

Disciplinary Field (code): SECS-P/07

Social and Informational Inferences

Advisor: Joachim VOSGERAU

PhD Thesis by

Graham OVERTON

ID number: 3082228

Year 2023

Introduction

Consumers are constantly gathering, and exposed to, information that shapes the way they evaluate goods, form preferences, and infer preferences. Often, others, such as friends, partners, salespersons, managers, and marketers have a choice to present information to consumers, and in what form they would like to present the information. A marketer might decide whether they want to publicize the results of a blind “taste” test between their product and a competitor’s product, and if they do, whether they want to present the ratings of each product or the percentage of respondents preferring each product. A friend might be asked what their preference is among several options over which a consumer is debating purchasing one, and in their response, must decide how to express their opinion. A retailer might decide whether they want to inform consumers of the quantity remaining of a good, and if so, how they should highlight this scarcity information. In all these cases, consumers are likely to instinctually draw inferences from the information they learn in order to help them make better decisions. Across the three chapters below, I explore these inferences, if they are accurate, what drives them, and the consequences on consumer choice.

In Chapter 1, I explore consumer inferences about a ubiquitous form of social proof, the results of a poll. Specifically, I examine individuals’ inferences about how much better the majority preferred option, “A”, is than the minority preferred option, “B”, when an individual learns the results of a poll between the two (e.g. 70% prefer A over B). I find that individuals’ vastly overestimate how much better option A is than option B across a large range of consensus levels (the percentage of respondents preferring the majority preferred option A). This overestimation can cause participants to choose option A over option B when learning consensus information even though they would not if they saw the true average ratings for each option instead.

In Chapter 2, I investigate consumers inferences about how liked an option is when learning another's preferences over a set of options. Consumers often requests the opinions of others, friends, salespersons, dates, etc., when deciding which option they would like to consume. For example, when shopping for a new pair of shoes one might ask their friend for their opinion or when taking a date to a restaurant, one might ask their date what they think of the options. In all cases, the requestor is seeking to understand how much their friend likes each option. Friends, responding, to the requestor might sometimes have a preference for a single option, stating I like option A, or sometimes might like more than one option (expressing some indifference), stating I like both option A and option B. I find that requestors believe a friend likes option A less when their friend expresses indifference (they like both A and B) compared to if they only had a preference for option A. Interestingly, requestors tend to underestimate how liked option A is when positive cues accompany expressions of indifference. Consequently, even when options are known to be liked but indifference is expressed, requestors still desire to search more, and incur this cost, than if their friend had just liked only one option.

Finally, in Chapter 3, I examine how perceptions of a good's scarcity might depend on other good's scarcity. The classical definition of scarcity can generally be stated as the extent of a good's own (un)availability. The lower the number of units currently available of a good, the scarcer it is. While previous research has operationalized scarcity according to this definition, in this work I show that a good's scarcity depends on whether another good is less scarce. For example, a wine that has only 2 bottles left but initially there were three cases, 36 bottles, is perceived to be scarcer and valued more if one additionally learned that a different wine has 30 bottles left out of 36 bottles compared to if the additional wine had 2 bottles left. This occurs despite the target wine already having a reference point, its initial stock of 36 bottles, available and even though the actual, objective scarcity, its current availability of 2 bottles, has not changed. Thus, I document a psychological effect of product scarcity that occurs above and beyond actual product scarcity.

Chapter 1

People Believe if 90% Prefer A Over B, A Must be Much Better than B. They Are Wrong.

Abstract

We show that people confuse consensus information in polls—such as 90% prefer product A over product B—with utility differences—the extent to which poll respondents prefer A over B. Consequently, they interpret a 90% consensus in favor of A as the average utility of A being considerably higher than the average utility of B. We demonstrate empirically and with simulations that—while this can be true—it is more likely that the average utility of A is only slightly higher than that of B. People are not aware of this regularity, and believe that 90% consensus implies A being *much better* than B. Communicators can capitalize on these erroneous inferences and strategically display the same preference information in consensus or rating formats leading to dramatic shifts in choices. People' erroneous inferences can be corrected by educating them about the shape of the distribution of utility differences at a 50% consensus level. We discuss theoretical implications for social proof and preference cascades, and managerial implications for the understanding and usage of polls management and public policy.

1. Introduction

In the famous Pepsi Challenge of the 80s, the majority preferred Pepsi in blind taste tests over Coke. The Pepsi Challenge was very effective in changing people' preferences and caused a major decline in Coke's market share (History.com 2020). In subsequent years, Coke came back with their own taste poll, and Pepsi is still doing taste challenges today, most recently showing that 71% of the Republic of Ireland prefers Pepsi Max over the market leader in caffeinated carbonated sodas (Pepsimax.ie 2019).

Polls are a popular method to gauge and convey aggregate preferences in marketing (e.g., 71% prefer Pepsi Max, which restaurant is voted to have the best cheesesteak in Philadelphia; Phillymag.com 2018), sports (which NFL team is predicted to win the Superbowl; ESPN 2020), politics (75% of republicans approve of Donald Trump, Breitbart.com 2021), public policy (the majority of Americans support abortion rights; Forbes 2021), in short in any domain in which aggregate preferences are of interest. Polls are popular for two reasons. First, they make it easy for respondents to express their preferences since choosing is more natural and less cognitively demanding than evaluating options on rating scales (Fisher and Keil 2018; Fisher, Newman, and Dhar 2018; Huber, Ariely, and Fischer 2002; Peterson and Pitz 1988). Second, poll outcomes are easily communicated with a single number (71% prefer Pepsi Max), whereas preferences expressed as ratings or WTP require the comparison of two numbers, one for each option.

In this paper, we investigate how people interpret poll results. For example, in an autoguide.com (2017) poll where 3000 consumers answered, "Which is better, the Accord or the Camry?", 72% chose the Honda Accord over the Toyota Camry. We refer to the result of such a poll as "consensus information" as it indicates the proportion of respondents with the same preference ordering (72% like the Accord better than the Camry, and 28% like the Camry better than the Accord). With two choice options, 100% indicates complete consensus, and 50% no consensus.

According to rational choice theory (Keeney, Raiffa, and Meyer 1993; Luce 1977), choices made in a poll are based on a comparison of the absolute utility that each choice option affords. Whichever option is preferred (i.e., provides greater utility) is chosen. Hence, choice communicates nothing about the extent to which one option is preferred over the other, it only indicates the preference ordering of the two options. As we argue, however, people spontaneously infer from consensus information the extent to which one option is preferred over the other. From “72% prefer the Accord over the Camry,” they may infer that respondents liked the Accord much more than the Camry.

Surprisingly, this inference is not necessarily correct. The Accord may be preferred to a great extent over the Camry, but it is, in fact, more likely that the two cars are similarly liked. We will demonstrate this regularity empirically using real-world data and with simulations. Importantly, we show that people are not aware of this regularity, and believe that a large consensus implies large preference differences. Consequently, they tend to overestimate how much better the majority preferred option is than the minority preferred option, which is one reason why polls and consensus information are so persuasive.

2. Theoretical Development

2.1. Consensus Levels and Corresponding Utility Differences

Consider a poll with two options, A and B, where the utilities underlying respondents’ choices are expressed in integers and are uniformly distributed between 1 and 10 (where 1 = lowest utility level and 10 = highest utility level; we will later relax these assumptions). Table 1 below shows all utility differences between A and B that result from the 100 possible pairs. Any pair is as likely to occur as any other. In a sample of 100 respondents, we would expect that A is preferred over B by 45 respondents (those in the upper right grey area), and B is preferred over A by 45 respondents (those in the lower left grey area).

Table 1. All possible combinations of the utilities of two options A and B and their corresponding utility differences (utilities are uniformly distributed)

Utilities for B \ Utilities for A	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	1	2	3	4	5	6	7	8	9
2	−1	0	1	2	3	4	5	6	7	8
3	−2	−1	0	1	2	3	4	5	6	7
4	−3	−2	−1	0	1	2	3	4	5	6
5	−4	−3	−2	−1	0	1	2	3	4	5
6	−5	−4	−3	−2	−1	0	1	2	3	4
7	−6	−5	−4	−3	−2	−1	0	1	2	3
8	−7	−6	−5	−4	−3	−2	−1	0	1	2
9	−8	−7	−6	−5	−4	−3	−2	−1	0	1
10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0

Utility differences range from −9 (A = 1 and B = 10) to 9 (A = 10 and B = 1). Because polls usually require respondents to choose either A or B, we will only consider utility differences that are not 0 (i.e., indifference is precluded). There are nine ways for utility differences to be 1 or −1, eight ways for differences to be 2 or −2, seven ways for differences to be 3 or −3, and so on. So, when looking at all cases in Table 1 (i.e., consensus is 50%), small utility differences are more likely than large differences. This combinational logic holds for all consensus levels, whether 50% prefer A over B (all cases in Table 1) or 100% prefer A over B (the 45 cases in the upper right grey area of Table 1) or 100% prefer B over A (the 45 cases in the lower left grey area of Table 1). Small utility differences are always more likely than large utility differences.

From Table 1, we can also calculate the expected utility difference for a given consensus level. To do so, we assume that instead of denoting preferences of single respondents, utilities in Table 1 represent polls, that is average utilities for a group of respondents with similar preferences. Let us take a look at the most extreme consensus level where 100% prefer A over B (the 45 cases in the upper right grey area with positive utility differences). Because the underlying utilities are uniformly distributed, the expected difference in utilities is simply the average of the 45 utility differences, that is 165/45 = 3.667. Thus, if 100% of respondents prefer A over B, the expected difference in utilities is 3.667. Conversely, if 100% preferred B over A, the expected difference in utilities would be −3.667.

Building on this, we can calculate the expected difference in utilities for different consensus levels by weighting utility differences by the relative size of the majority and the minority group (see Table 2). For example, when 90% prefer A over B and 10% prefer B over A, the expected difference in utilities is 3.667 x 90% − 3.667 x 10% = 2.933.

Table 2: Expected differences in utilities for two options A and B for different consensus levels.

Consensus	Expected utility difference for majority	Expected utility difference for minority	Overall expected difference in utilities
100% A (0% B)	3.667 x 100%	−3.667 x 0%	3.667
90% A (10% B)	3.667 x 90%	−3.667 x 10%	2.933
80% A (20% B)	3.667 x 80%	−3.667 x 20%	2.201
70% A (30% B)	3.667 x 70%	−3.667 x 30%	1.467
60% A (40% B)	3.667 x 60%	−3.667 x 40%	0.733
50% A (50% B)	3.667 x 50%	−3.667 x 50%	0

Table 2 shows that, as the level of consensus decreases, the expected difference in utilities also decreases. In fact, the two correlate with r = 1. This corresponds well with intuition.

The expected difference in utilities for options A and B is likely to be larger when 90% of respondents prefer A over B than when only 60% do so. At *any* given consensus level, however, smaller differences in utilities are more probable than larger differences due to the combinatorial logic outlined above. It is this key property that defies intuition. No matter whether 90% or 60% prefer A over B, it is more likely that both options provide similar utility levels than vastly different utility levels.

2.2. Are Small Differences in Utilities always more likely than Large Differences?

Before we outline the consequences of this counterintuitive regularity, let us briefly discuss the assumptions behind our modeling. In the above analyses, utilities for each option assumed only integer values, were uniformly distributed, and positively correlated[1]. Relaxing the first assumption by allowing for decimals changes the distributions of expected utility differences only minimally (see the simulations in Online Appendix A). Relaxing the second assumption by assuming that moderate levels of utilities are more likely than extreme levels (e.g., by switching from uniform to normal distributions truncated at the end points 1 and 10) causes smaller differences to become even more likely. Relaxing the second and third assumptions by specifying the sign of the correlation between utilities of option A and B causes shifts in opposite directions. When utilities of option A and B within a group are positively correlated (e.g., the more respondents like A the more they also like B), smaller differences

[1] Within groups, that is for those who prefer A over B and those who prefer B over A, average utilities for A and B correlate with 0.5 in our uniform simulations due to preference constraints. The intuition is as follows: imagine we graph utilities such that the x-axis displays utilities for A and the y-axis utilities for B. For those who prefer A over B, the graph will show a left-skewed distribution (because utilities for A must be greater than utilities for B). This creates a correlation of utilities for A and B of 0.5. Now imagine the same graph for those who prefer B over A. This graph will show a right-skewed distribution (because preferences for B must be greater than for A), resulting again in a correlation of 0.5 of the utilities. The correlation between A and B for each group can change when only a subset of all possible combinations are represented.

become more likely. When utilities of option A and B within groups are negatively correlated (the more respondents like A the less they like B), larger differences become more likely. But even in these cases, the effect of the combinatorial logic making small differences more likely than large differences still supersedes the effect of negative correlations. So, even when 90% prefer A over B and utilities for both options are correlated with −0.8, it is still the case that small differences in utilities are more likely than large differences (see simulations in Online Appendix A).

What about when each group has a love-hate relationship with their preferred and non-preferred options (i.e., when utility differences follow bimodal distributions)? In these cases, the combinatorial mechanism breaks down. For demonstration purposes, let's consider an extreme case in which utilities for A and B are maximally different and highly negatively correlated. Imagine the 90% majority prefers A over B, with mean utility levels for A = 9/10 and for B = 1/10, while the 10% minority prefers B over A with opposite utilities. We simulated this case with utilities for A and B within each group correlated at -.86 (see details in Online Appendix A), resulting in large utility differences being more likely than small utility differences.

2.3. Empirical Demonstration that Small Utility Differences are More Likely than Large Differences

To test our logically derived hypothesis empirically, we acquired three datasets of preference ratings in which respondents provided paired ratings of the same targets: jokes, beers, and movies. The jokes dataset contains more than 600,000 ratings of 100 jokes from 24,938 respondents, where the median respondent rated 24 jokes on a scale from -10 to 10 (Goldberg, Roeder, Gupta, and Perkins 2001). The beer dataset contains more than 1.5 million ratings of

56,000 beers from 33,388 respondents, where the median respondents rated 3 beers on a scale from 1 to 5 in .5 increments (https://www.kaggle.com/rdoume/beerreviews). The movie dataset contains more than 25 million ratings of 62,423 movies from 162,541 respondents, where the median respondent rated 71 movies on a scale from 0.5 to 5 in .5 increments (Harper and Konstan 2015; https://grouplens.org/datasets/movielens/25m/). For more details about the datasets please see the Online Appendix B.

In the following analyses, we treat ratings as measures of utility derived from the rated target objects (e.g., beer A and beer B). We selected ratings of target objects only from respondents who had rated both objects. This allowed us to calculate average differences in ratings (utilities) as well as consensus levels, the proportion of respondents preferring one option over the other according to their ratings. Like in our analyses above, we eliminated ratings resulting in a zero difference to mimic choices in polls.

Figure 1. Histograms of rating differences across consensus levels for jokes, beers, and movies, respectively.

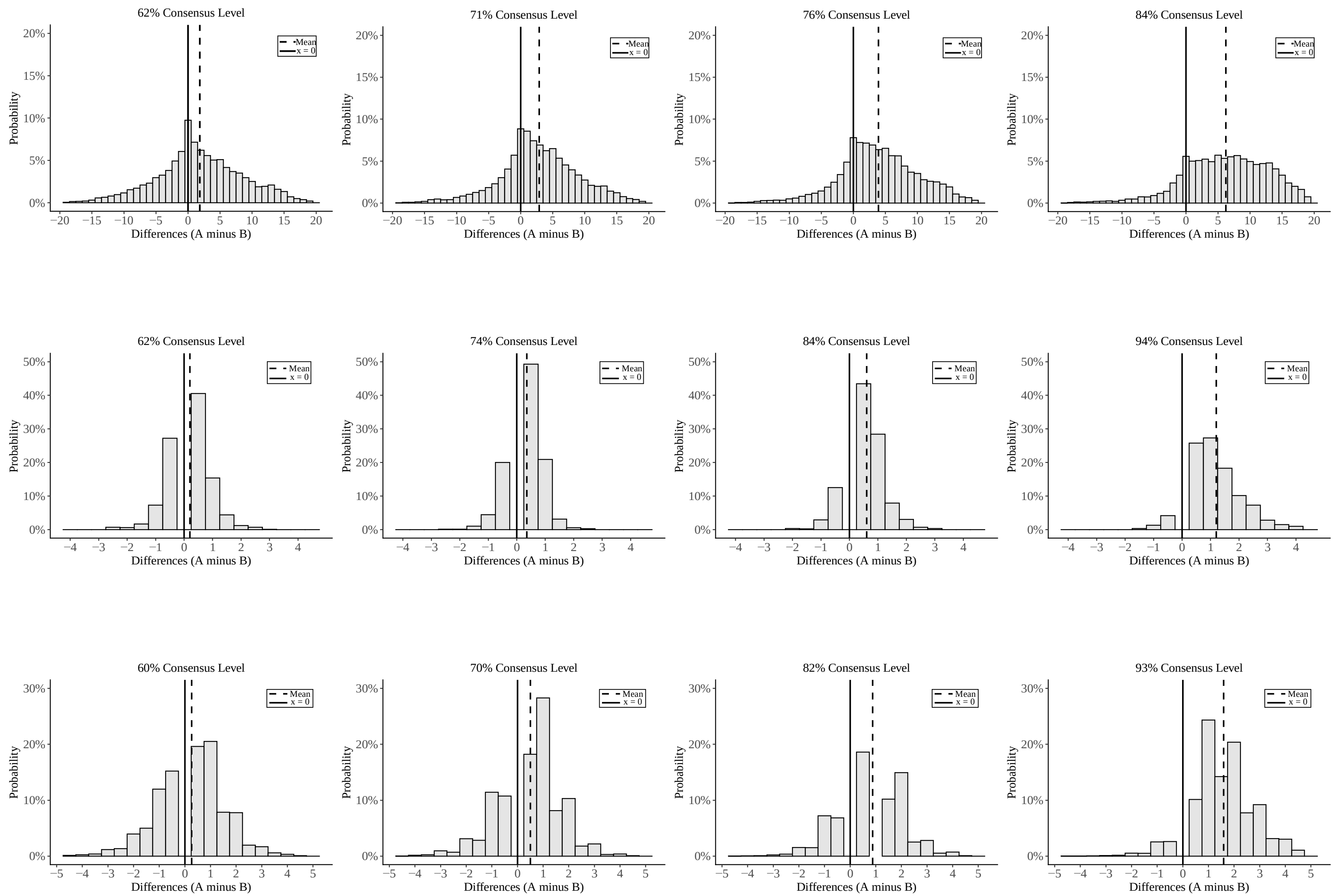


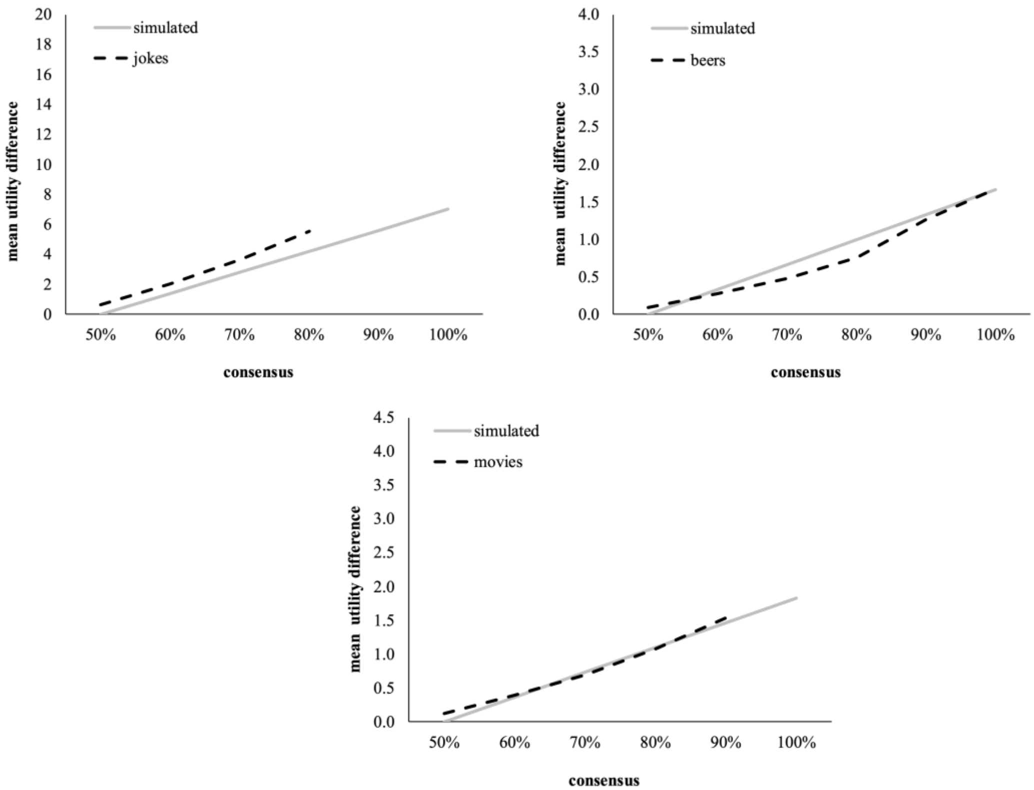
Figure 1 shows the distribution of rating differences for four different joke, beer, and movie pairs each, representing low to high consensus levels.

We obtain supporting evidence for the two key insights found in our analysis above and in our simulations. First, we observe that small utility differences are generally more likely than large utility differences [2]. Second, a comparison of average utility differences (vertical dashed lines) across graphs reveals that as consensus increases, so do average utility differences (i.e., they are highly positively correlated).

To further test the above observations, we examined thousands of possible joke combinations and more than 150,000 combinations of beers and of movies. For each combination, we calculated the consensus, mean, and modal value (see Online Appendix C). We then compared the mean and mode estimates at different consensus levels to those derived from our simulations. As can be seen in Figure 2, the average mean utility differences are small and extremely close to those derived in our simulations. The same holds for observed and simulated modal values.

[2] The rating differences for movies are particular in that respondents tend to rate movies in integers but not fractions, which explains why integer differences are more likely than fractional differences in movie ratings (cf., analysis in the Web Appendix B).

Figure 2. The grey solid line represents average utility differences according to our simulations, and the black dashed line are the actual observed mean utility differences in the joke, beer, and movie ratings data, respectively.



2.4. Given Consensus Information, People Overestimate Average Utility Differences

We are now in a position to investigate the consequences of these key insights for people's judgments and decisions. We surmise that when given consensus information such as 90% prefer A over B, people automatically infer from such information a corresponding average utility difference between A and B, that is by how much does the average respondent prefer A over B, or more generally, how much better is A than B. The vast amount of research on inferences and information processing suggests that people infer the target of a judgment (e.g., how much better is A than B?) from the reference information that is directly observable (e.g., consensus information) and the perceived correlation between the target and the reference (e.g., Broniarczyk and Alba 1994; Dick, Chakravarti, and Biehal 1990; Evangelidis and van Osselaer 2018, 2019; Ford and Smith 1987, Ganzach and Krantz 1990; Gunasti and Ross 2009; Jaccard and Wood 1988; Johnson and Levin 1985; Kardes et al. 2004a; Kardes, Posavac, and Cronley 2004b; Slovic and MacPhillamy 1974). And since the correlation between consensus levels and average utility differences is close to 1, people tend to perceive the two forms of preference information to be substitutes (Kahneman 2003; Kahneman and Frederick 2002). From 50% consensus, people are likely to infer an average utility difference close to zero, and as consensus levels increase, so, too, will the inferred average utility differences. However, because they are unaware of the combinatorial logic according to which small positive utility differences are—at any consensus level—more likely than large positive utility differences, they will infer distributions of utility differences that deviate from the truth, such that small utility differences are underrepresented and large utility differences are overrepresented. As a consequence,

Hypothesis 1*:* When given consensus information, people will tend to overestimate the mean and mode of utility differences.

When measures of underlying utilities are available, such as online ratings for example, information about others' preferences can be displayed as either consensus information or as average utility ratings of the two options. Given that people tend to overestimate utility differences inferred from consensus information (H1), we expect consensus information to exert a stronger impact on peoples' own preferences compared to information about the utilities of each option (e.g., ratings). Communicators, such as marketers, managers, public policy makers, politicians, etc. should hence be able to influence peoples' choices by displaying the same preference data in one way or the other. Particularly, we predict that:

Hypothesis 2*:* If the utility difference is at the mean implied by a given consensus level (cf., Figures 1 and 2 and S2 in the Online Appendix), people will be more likely to select the majority-preferred option over the minority-preferred option when they learn consensus information compared to learning information about the utilities of each option (e.g., mean ratings).

3. Experiments

3.1. Empirical Overview

We test our hypotheses in six experiments. Experiments 1a, 1b, 3 and 4 do so with real-world data, experiments 1c and 2 use our simulations as benchmarks. Specifically, in experiment 1a ESPN experts' predictions for Super Bowl games are given in consensus format to participants who then predict the experts' average point spreads. In experiment 1b, participants read a pair of jokes and are informed about the proportion of online-respondents preferring one over the other. They then estimate the average difference in ratings of the two jokes. Finally, in experiment 1c we test whether consensus information also sways statistics-savvy participants into (incorrectly) inferring that larger utility differences are more likely than smaller ones using a simple choice measure and a large monetary incentive.

In experiment 2, we test the downstream consequences of the overestimation bias for choice (H2). Participants see either consensus information or the corresponding average ratings for two hotel rooms. We test for which information display participants are more likely to choose the majority-preferred but more expensive room over the minority-preferred but cheaper room.

Experiments 3 and 4 test whether participants overestimate how much better A is than B because they infer utility difference distributions that are too spread out, such that small utility differences are underrepresented and large utility differences are overrepresented. Specifically, in experiment 3 we elicit participants' beliefs about how differences in real beer ratings are distributed, and in experiment 4 we test whether the beliefs observed in experiment 3 can be debiased by showing participants the correct distribution of differences in beer ratings at the 50% consensus level.

All six experiments are preregistered. All stimuli, experimental materials, code, data, and preregistrations are accessible at https://researchbox.org/446&PEER_REVIEW_passcode=JEPIYR. Correct answers were monetarily incentivized in experiments 1a, 2, 4, and 5.

3.2. Experiment 1a: Inferring ESPN Experts' Point Spread Predictions for Super Bowl Games

Experiment 1a was designed as a test of H1 using real world data. Every year, ESPN football commentators and analysts predict the outcome of the Super Bowl. Interestingly, ESPN.com provides readers with the percentage of experts predicting each team to win, that is consensus information. This consensus information is directly calculated from the experts' score predictions for each team, which are provided on the same webpage right below the consensus information. Thus, for each expert we know how much better they think the winning team is than the losing team (i.e., we know each expert's utility difference or point spread).

From the four Super Bowl games for which data are available on ESPN.com (2020), we chose Super Bowl LI and LIII. For Super Bowl LIII, 62% of experts predicted Team A to win by, on average, 0.97 points. The Vegas spread was 2.5 points, and Team A actually won the game by 10 points. For Super Bowl LI, 72% of experts predicted Team A to win by, on average, 1.87 points. The Vegas spread was 3 points, and Team A actually won the game by 6 points. In our experiment, we presented participants with the ESPN experts' prediction consensus of either game (where we replaced the actual names of the teams with Team A and Team B), and asked them to guess the average spread predicted by the experts (estimates were monetarily incentivized for accuracy).

3.2.1. Method

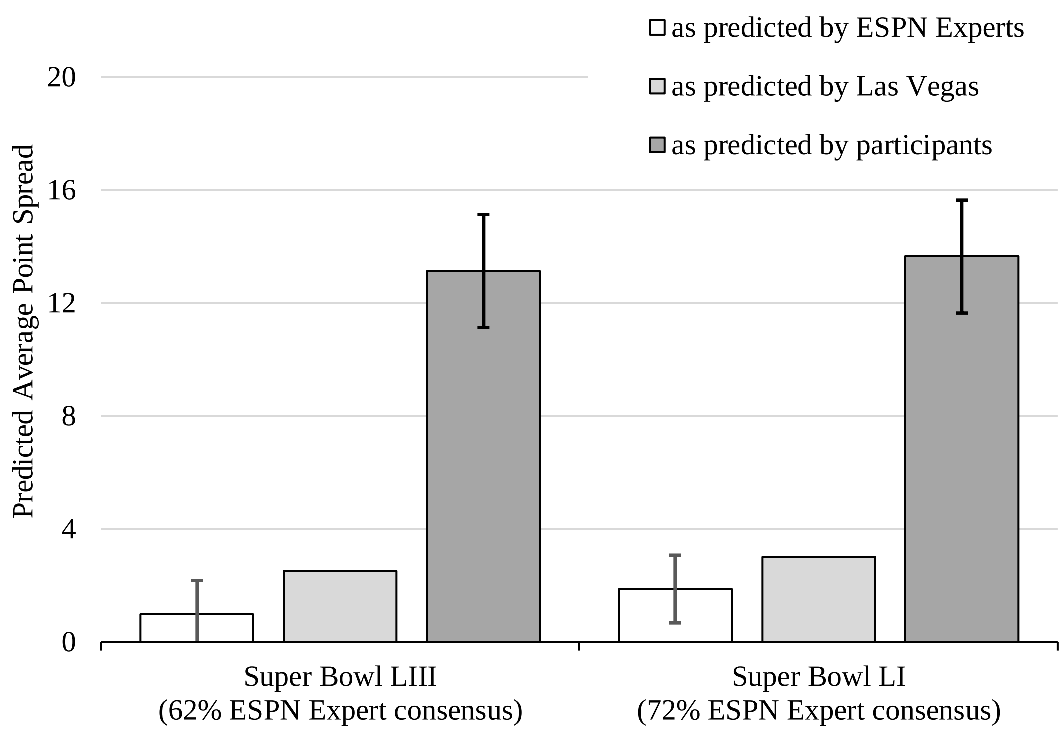
We recruited 400 participants via Amazon Mechanical Turk to complete our study in exchange for $.20 and a potential bonus of $.10 for answers within 10% of the actual point spreads. In our MTurk listing, we requested football fans. To ensure participants were football fans, MTurkers entering our survey were asked two screening questions (number of players on the field, and identification of a referee signal). Participants who failed to answer both questions correctly were (as preregistered) not allowed to continue with our study. Of those who successfully passed this screening, we dropped anyone who took our survey more than once leaving us with 368 participants (79.8% female, M_{age} = 37.2, *SD* = 11).

Participants were randomly assigned to either the Super Bowl LI (72% consensus) or the Super Bowl LIII (62% consensus) condition. In each condition, participants learned that "ESPN experts regularly predict the outcomes of professional football games." We then presented them with information specific to the Super Bowl game to which they had been assigned to. In the

62% [72%] condition, they read "For one such game (we cannot disclose the name of the teams, so we will call them Team A and Team B), 97 [100] ESPN football experts predicted the final scores for each team.[…] 62% [72%] of the ESPN experts predicted Team A would win. 38% [28%] of the ESPN experts predicted Team B would win."

Participants were then asked, "What do you think is the average point spread predicted by the ESPN experts (i.e., the average predicted number of points by which one team wins over the other)?" Participants were informed that if their answer was within 10% of the average point spread of the experts, they would get a $0.10 bonus.

Figure 3. Predicted average point spreads for Super Bowl games LI and LIII. Error bars show 95% CIs.



3.2.2. Results and Discussion

In order to control for outlandish guesses, we winsorized (as preregistered) the data at the 5% and 95% percentiles. As predicted, participants overestimated experts' predicted point spreads, both for the 62% consensus ($M_{\text{predicted}}$ = 13.13, *SD* = 14.14; M_{actual} = 0.97, *SD* = 5.67; $t(273) = 8.12$, $p < .001$) and the 72% consensus ($M_{\text{predicted}}$ = 13.64, *SD* = 13.85; M_{actual} = 1.87, *SD* = 5.86; $t(288) = 8.12$, $p < .001$, see Figure 3). These results are robust across stricter winsorization cutoffs and persist when compared to the actual Vegas spreads.

Experiment 1a demonstrates that football fans, who were told the proportion of ESPN experts predicting the winner of two Super Bowl games, overestimate the average point spread predicted by the sports experts. Participants' estimates were on average four times as high as the Vegas spreads, and at least 7.5 times larger than the ESPN experts' point spreads. The results of Experiment 1a thus provide support for H1.

One limitation of Experiment 1a, however, may be that any given consensus level is compatible with many different levels of average point spreads (cf., Figures 2 and S2 in the Online Appendix). By picking out two Super Bowl games, we may have simply been lucky in having selected two games for which point spreads were unusually low. Next, we further test H1 using a different real-world benchmark and stimuli.

3.3. Experiment 1b: Inferring Funniness Ratings and Testing the Role of Priors

Experiment 1b is a conceptual replication of experiment 1a with joke ratings. We provided participants with consensus information about two jokes of the joke dataset and asked them to estimate how much funnier they thought one joke was rated than the other. Half of the participants were given the opportunity to read the two jokes before providing their ratings. We did so to test whether forming prior beliefs about the funniness of the two jokes would influence participants' estimates of funniness ratings.

3.3.1. Method

We recruited 600 participants via Prolific to complete our study online in exchange for £0.38. Following our preregistration, we dropped anyone who took our survey more than once or gave an answer beyond the range restricted by our Qualtrics coding (i.e., hackers), leaving us with 599 participants (70.4% female, M_{age} = 34.7, *SD* = 12.2). To select two jokes from the joke dataset, we looked at pairs where each joke had more than 10,000 ratings and eliminated jokes that might be too offensive (e.g., sexual jokes and black humor). We finally selected a joke pair that had been rated by 15,096 respondents. Of the 15,096 respondents, 71% had rated joke A about a group of managers trying to measure the height of a flagpole as funnier than joke B about a dog sending a telegram (the jokes can be found in the pdf and qsf survey files on Researchbox).

Participants first read that "you will learn the opinion of 15,096 people who went on a joke recommendation site called Jester. On that website, people rate jokes from "Not Funny" to "Very Funny" on the scale below. While no numerical rating is displayed to participants, their ratings are saved as a corresponding number from 1 (leftmost value) to 21 (rightmost value)." Along with this text, they saw an image of the scale. After two comprehension check questions that participants had to answer correctly to advance with the survey, they were randomly assigned to one of three conditions and informed that they would be provided with a link to the joke website at the end of the study.

In the 'control' condition, participants learned that "71% rated Joke A higher (funnier) than Joke B. 29% rated Joke B higher (funnier) than Joke A."

In the ‘learn’ condition, participants were told that they would see the jokes that the 15,096 raters saw. They then read both jokes (presented in random order) and were given the same consensus information as in the control condition. This allowed us to test whether forming prior beliefs would influence estimates of the average difference in funniness ratings. After reading the jokes, participants learned the same consensus information as in the other conditions.

In a third ‘rate and learn’ condition, participants saw both jokes and additionally rated how funny they thought each joke was from 1 to 21 with up to two decimal places.

Participants were then asked “What do you think is the most likely difference in Joke A's rating minus Joke B's rating produced by the 15,096 participants?” Participants could enter a value between −20 and 20 with at most two decimals, and were given two examples, one illustrating joke B being funnier than joke A resulting in −20 and the other illustrating joke A being funnier by 20 points. Finally, they answered demographic questions, were provided a link to the joke website if they wished to view it and completed the survey.

We predicted that the majority of participants in the control condition would enter a value greater than the observed median of 2.57 (cf., first row of Figure 1 at the 71% consensus level). Using the median as a benchmark is a very conservative test as it is always substantially higher than the actual most likely difference. We did not make formal predictions about the learn and the learn and rate conditions.

3.3.2. Results and Discussion

As predicted, the majority of participants overestimated the most likely difference in ratings, entering a value higher than the observed median (94.4%, Pearson $\chi^2(1) = 154.46$, $p <$ .001). A slightly smaller majority did so in the ‘learn’ (89.1%) and the ‘learn and rate’ (84.1%)

conditions (89.1% vs. 94.4%, Pearson $\chi^2(1) = 3.64$, $p = .056$; 84.1% vs. 94.4%, Pearson $\chi^2(1) = 10.91$, $p = .001$). The percent overestimating did not differ significantly between the 'learn' and 'learn and rate' conditions (89.1% vs. 84.1%, Pearson $\chi^2(1) = 2.19$, $p = .138$).

We also explored whether the degree of overestimation differed by condition. Compared to the control, the 'learn' ($M_{\text{learn}} = 9.00$, $SD = 7.82$ vs. $M_{\text{control}} = 10.54$, $SD = 6.72$, $t(396) = 2.10$, $p = .036$) and 'learn and rate' ($M_{\text{learn\_and\_rate}} = 8.50$, $SD = 7.81$ vs. $M_{\text{control}} = 10.54$, $SD = 6.72$, $t(395) = 2.79$, $p = .006$) conditions both significantly reduced the degree of overestimation. They did not significantly differ from each other ($M_{\text{learn}} = 9.00$, $SD = 7.82$ vs. $M_{\text{learn\_and\_rate}} = 8.50$, $SD = 7.81$, $t(401) = 0.64$, $p = .52$). All estimates were well above the median of 2.57 (all $p < .001$).

The results of experiment 1b replicate the findings of experiment 1a. Like the football fans in experiment 1a overestimated point spreads predicted by ESPN experts when learning about consensus of the ESPN experts, participants in experiment 1b overestimated the difference in funniness ratings for two jokes when learning that 71% preferred joke A over joke B. And as in experiment 1a, the degree of overestimation is also substantial in experiment 1b and is barely influenced by participants' own beliefs about how funny the jokes are.

3.4. Experiment 1c: Overestimation of Average Preference Differences with Simulated Data

Experiment 1c tested H1 with simulated data. Participants were asked to predict the most likely difference in average wine ratings for two wines, A and B, where either 60% or 90% of wine tasters had given Wine A a higher rating than Wine B. To make experiment 1c a conservative test of our hypotheses, we recruited a group of analytically trained students from a master's program at a European business school who were taking a data analytics course, and incentivized accuracy by offering 25 euros each to two students who gave the correct answer.

3.4.1. Method

We aimed to recruit 100 participants and ended up with 86 (67.4% female, M_{age} = 23.3, *SD* = 1.7) from the data analytics class on the one day the study was run. Participants were randomly assigned to either a low consensus (60%) or a high consensus (90%) condition. Specifically, participants were informed that "500 people rated both Chardonnays below (Wine A and Wine B) on a scale from 1 to 6 (1 = poor quality and 6 = very high quality)." In the low [high] consensus condition participants learned that "60% [90%] rated Wine A as higher quality than Wine B. 40% [10%] rated Wine B as higher quality than Wine A." A comprehension check ("What rating scale was each wine rated on?") had to be answered correctly to advance with the study. Students were then asked which difference between the ratings of Wine A and Wine B they thought was the most likely. The answer options were:

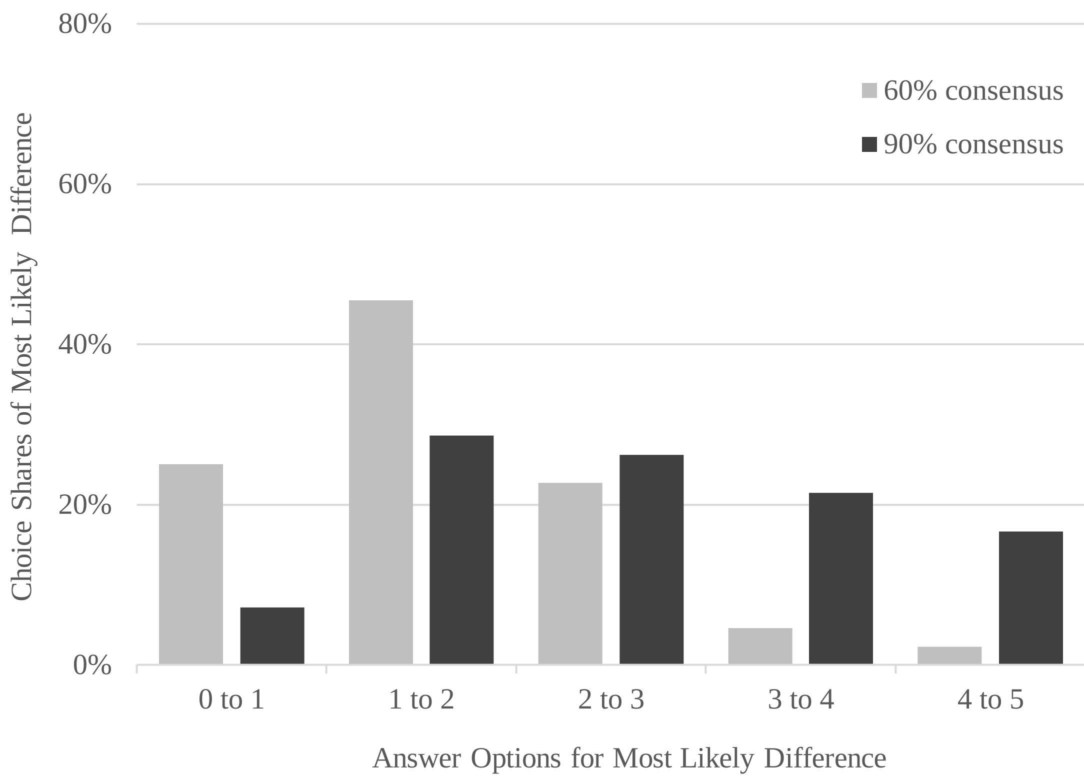
a) The average rating of Wine A is about 0-1 point higher than that of Wine B,
b) The average rating of Wine A is about 1-2 points higher than that of Wine B,
c) The average rating of Wine A is about 2-3 points higher than that of Wine B,
d) The average rating of Wine A is about 3-4 points higher than that of Wine B,
e) The average rating of Wine A is about 4-5 points higher than that of Wine B."

Note that each choice option comprises a range of values rather than asking about a specific value. We did this because we thought it would be easier to compare likelihoods of ranges of values than likelihoods of specific values. We predicted that: 1) in both the 60% and 90% consensus conditions, a minority of participants would choose the correct answer option (a) as observed in our simulations, 2) in the 90% consensus condition compared to the 60% consensus condition, more participants would choose larger difference answer options, (e) and (d), than relatively smaller difference answer options, (b) and (c). Participants were informed that there was only one correct answer. If they guessed it, they would be entered into a draw where two winners would each win 25 euros.

3.4.2. Results and Discussion

Results are summarized in Figure 4. Participants in both conditions erroneously believed that larger differences in ratings were more likely than smaller ones. Importantly, as predicted, in the 60% consensus condition, the majority of participants (75%) erroneously believed a larger difference was more likely than the incentivized, correct answer of “about 0-1” (25%; Pearson $\chi^2(1) = 11.00$, $p = .001$). Likewise, in the 90% consensus condition, as predicted the majority of participants (92.9%) erroneously chose larger differences to be more likely than the incentivized, correct answer of “about 0-1” (7.1%, Pearson $\chi^2(1) = 30.86$, $p < .001$). As a conservative test, we also tested whether the majority of participants in the 90% consensus condition chose an answer greater than “about 1-2”, they did (69% vs. 31%, Pearson $\chi^2(1) = 5.77$, $p = .016$). Finally, as predicted more participants in the 90% than the 60% consensus condition believed that the largest differences of “about 3-4” and “about 4-5” were more likely than the smaller differences of “about 1-2” and “about 2-3” (69.6% vs. 9.1%, Pearson $\chi^2(1) = 9.38$, $p = .002$).

Figure 4. Proportion of respondents choosing each answer option for the most likely difference in ratings in Experiment 1c.



As in experiments 1a and 1b, participants in experiment 1c overestimated the mode of utility differences as they predominantly failed to choose the correct answer of a 0 to 1 rating point difference despite being incentivized to do so. This occurred even with master level students that had taken several classes in data analytics. Having demonstrated how people intuitively conflate consensus information with larger utility differences, in the next experiment we test a behavioral consequence of such overestimation (H2).

3.5. Experiment 2: Influence of Consensus versus Utility Difference Information on Consumer Choice

According to social proof (Cialdini 2007) and choice mimicry (Young et al. 2014), people are particularly likely to follow others' preferences when they are uncertain about their own preferences. Others' preferences can be communicated as consensus information, as an average utility difference, or as the average utilities of the choice options. Since people overestimate utility differences inferred from consensus information, they should be more likely to select a majority-preferred option over a minority-preferred option when they learn consensus information compared to the underlying average utilities (H2). Specifically, we asked participants to choose between a more expensive hotel A and a cheaper hotel B. Some participants were told about the proportion of other people preferring hotel A over B (consensus information), while others were told the average ratings for each hotel.

3.5.1. Method

We invited 450 participants via Amazon Mechanical Turk to participate in our study in exchange for $.20. Four hundred and fifty-one MTurkers completed the study (53.4% female, M_{age} = 37.9, *SD* = 12.4). Participants were asked to choose between two hotels A and B—Hotel A cost $125 a night and Hotel B $98 a night—and were further informed that 500 people had rated the two hotels (Hotel A and Hotel B) out of 100 points (1 = low quality, 50 = average quality, 100 = high quality). Participants were then randomly assigned to one of three conditions:

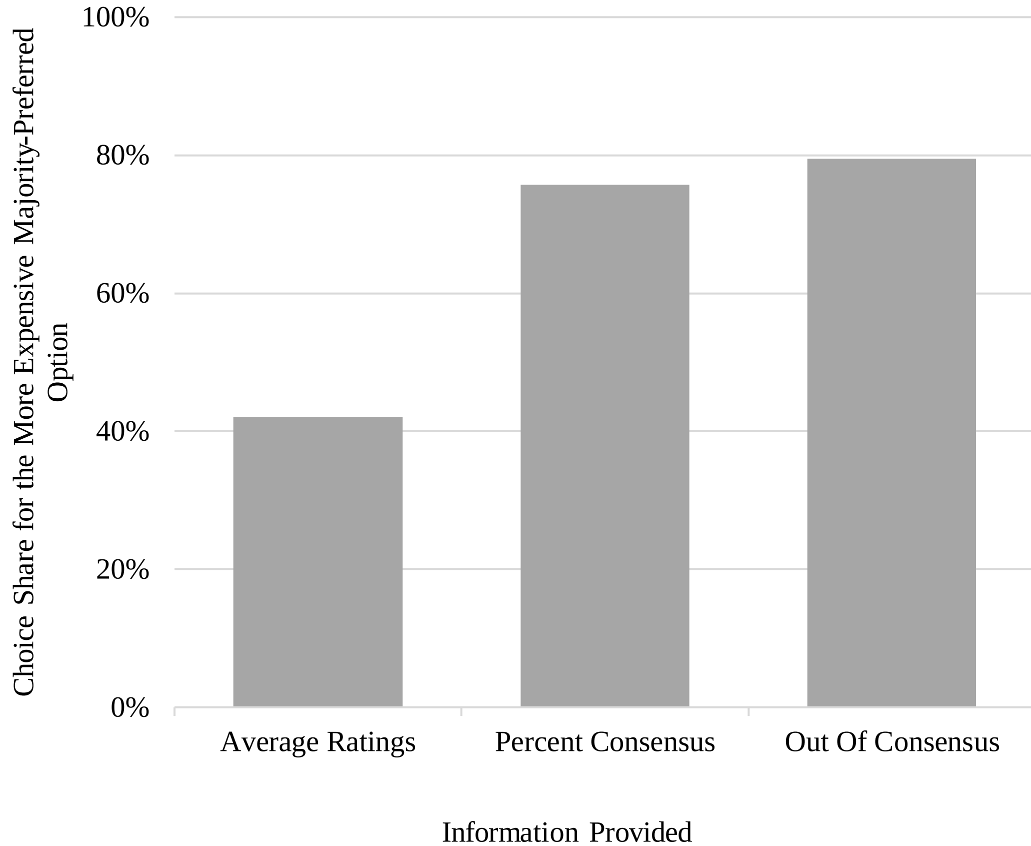
In the 'percent consensus' condition, participants learned that "70% of people rated Hotel A higher and 30% of people rated Hotel B higher." In the 'average ratings' condition, participants learned that the average rating of Hotel A is 87 out of 100 and the average rating of

Hotel B is 75 out of 100. In a third, ‘out of consensus’ condition, participants learned that “7 out of 10 people rated Hotel A higher and 3 out of 10 people rated Hotel B higher.”

Because any given consensus level is consistent with many utility differences, we referred to our simulations described in Online Appendix A to derive a corresponding utility difference. Specifically, using simulations with independent and uniformly distributed utilities for two options ranging from 1 to 100, a 70% consensus level is associated with an average utility difference of 12. Given the skewed nature of the distribution of utility differences, this average utility difference covers more than half of all utility differences (see Figure S2).

Comparing choice shares of the ‘percent consensus’ and the ‘average ratings’ conditions allows us to test H2. We included the third ‘out of consensus’ condition to test a potential alternative explanation for the predicted result that a greater proportion of participants in the ‘percent consensus’ than the ‘average ratings’ condition will choose the more expensive Hotel A. According to the alternative explanation, participants may anchor on the difference between the numbers provided, rather than infer utility differences from consensus information. That difference is 40% in the ‘percent consensus’ condition compared to 12 rating points in the ‘average ratings’ condition, which should produce a larger contrast in favor of the more expensive Hotel A in the former condition (e.g., Fernberger 1920; Heintz 1950; Sherif, Taub, and Hovland 1958; Wever and Zener 1928). To rule out this alternative account, the third ‘out of consensus’ condition frames consensus information as “7 out of 10 rated Hotel A higher and 3 out of 10 rated Hotel B higher.” The difference in numbers provided here is 4, compared to the 12 rating point difference in the ‘average ratings’ condition. According to the alternative account, a greater proportion of participants should choose the more expensive Hotel A in the ‘average ratings’ condition than in the ‘out of consensus’ condition because numerical differences are larger in the former condition. In contrast, we predict that a greater proportion of participants would choose Hotel A in the 'out of consensus' condition than in the 'average ratings' condition (consistent with H2).

Figure 6. Proportion of respondents choosing the more expensive majority-preferred hotel A in Experiment 3.



3.5.2. Results and Discussion

As predicted, participants were more likely to choose the more expensive Hotel A in the 'percent consensus' than the 'average ratings' condition (75.7% vs. 42.1%, Pearson $\chi^2(1) = 34.85$, $p < .001$; see Figure 6). Similarly, participants were more likely to choose the more expensive Hotel A in the 'out of consensus' than the 'average ratings' condition (79.5% vs.

42.1%, Pearson $\chi^2(1) = 44.34$, $p < .001$). Choice shares did not differ between 'percent consensus' and 'out of consensus' conditions (75.7% vs. 79.5%, Pearson $\chi^2(1) = 0.62$, $p = .431$).

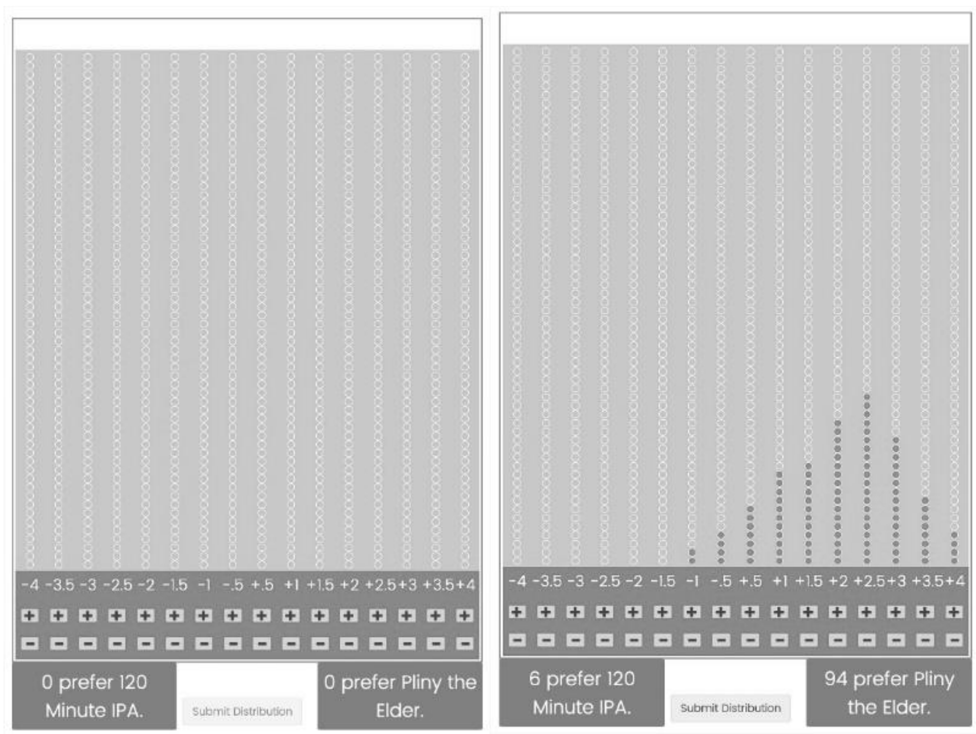
Supporting H2, a larger proportion chose the more expensive Hotel A when informed about the proportion of others preferring it rather than the average ratings in favor of it. Notice that the increase in the choice share of the expensive Hotel A is quite large. Describing the same preference information in terms of consensus rather than average utilities (i.e., mean ratings), from which the utility difference is seen, increased choice shares by more than 30%! This difference was not caused by participants anchoring on the larger numerical difference of "40%" in the 'percent consensus' condition, because we observed the same difference when consensus was described as 7 (vs. 3) out of 10 (and hence the numerical difference was only "4"). Importantly, the results of experiment 2 suggest that communicators can sway peoples' preferences dramatically by choosing how to display information about other peoples' preferences.

3.6. Experiment 3: Eliciting Distributional Beliefs about Utility Differences

In experiment 3, we examined the source of peoples' overestimation by eliciting their beliefs about how utility differences are distributed. To this end, we showed participants the results of two polls from our real-world beer data and elicited participants' beliefs about each poll responder's utility difference. Each participant thus revealed their belief about over- (or under) estimation of mean differences by constructing the entire distribution of preference differences. Preference distributions were elicited with the distribution builder, see Figure 7 (Sharpe, Goldstein, and Blythe 2000; Goldstein and Rothschild 2014; quentinandre.net 2022).

As before, we predicted that participants would overestimate how much better option A is than option B. Additionally, we expected that overestimation would cause participants to shift their distributions further right (“higher”) than they should be as they underestimate the likelihood of small utility differences.

Figure 7. Examples of the distribution builder tool used in experiments 3 and 4.



3.6.1. Method

We recruited 200 participants via Prolific to complete our online study in exchange for £0.90 (48% female, M_{age} = 27.63, *SD* = 9.67). All participants viewed two consensus levels, 74% and 94%, one at a time, and were randomly assigned to the order in which they viewed the consensus levels.

Our stimuli were derived from the beer dataset mentioned in our introduction (https://www.kaggle.com/rdoume/beerreviews). The dataset contains ratings of thousands of beers allowing for many possible pairs of beers that we could use for stimuli. To select ratings of four beers to form two beer-pairs with different consensus levels, we looked at all beer-pairs where each beer had more than 500 ratings (in order to derive more precise estimates of rating distributions). From that list of beer-pairs, we chose Pliny the Elder and 120 Minute IPA for which 601 users rated both resulting in a 94% consensus favoring Pliny the Elder, and Old Rasputin and Fat Tire Amber for which 638 users rated both resulting in a 74% consensus level favoring Old Rasputin.

Participants saw either the beer pair with 94% or with 74% consensus level first. For each consensus level beer-pair, participants read that "100 users on www.beeradvocate.com rated the two beers shown below on a scale from 1 to 5 with 0.5 increments (1 = worse rating and 5 = best rating)," and were provided with images of the two beer bottles. For example, when viewing the 94% consensus level beer pair, participants learned that "94% gave Pliny the Elder a higher rating than the 120 Minute IPA" and "6% gave the 120 Minute IPA a higher rating than Pliny the Elder." Below this they were asked: "What do you think the distribution of ratings looks like in this case?" Participants were then instructed to use the distribution builder to allocate each of the 100 users to a difference in ratings, which in the 94% case read "Pliny the Elder's rating minus 120 Minute IPA's rating." They were informed that a positive rating [negative] means that a user prefers Pliny the Elder [120 Minute IPA] and the greater this positive [negative] value, the more they prefer it. Participants learned they can add and subtract users with the buttons below the x-axis, and that their allocation must match the consensus level in order for them to proceed. Participants were finally told that if their distribution matched the actual distribution, they would receive a 20 cent bonus. After building preference distributions for both consensus level beer-pairs, participants provided their demographic information.

To make the task less demanding, we asked participants to allocate 100 users from each consensus level rather than the 601 and 638 raters upon which the 94% and 74% consensus level ratings were based, respectively. For each observed participant-distribution we computed the mean, mode, probability of the mode, probability of the smallest positive difference, and probability of the maximum difference. Compared to the actual distributions of rating differences, we predicted participants would: overestimate the mean, overestimate the mode, underestimate the probability of the mode, underestimate the probability of the smallest positive difference (i.e. 0.5), and overestimate the probability of the maximum difference. We also predicted that a majority of participants would commit each of these errors.

When building distributions, one consequence of asking participants to allocate 100 users instead of the total number of users is that a participant cannot easily replicate the true, full distribution for a given consensus level. As a solution, we created benchmarks as follows: we randomly drew 100 users (rating-differences) from the true distributions subject to the corresponding consensus level constraint, and repeated this 10,000 times for each consensus level. This resulted in two datasets, one for each consensus level. From these datasets, we computed the parameters of interest with their corresponding 95% confidence intervals. We classified a participant's parameter estimates as an overprediction when they exceeded the 95^{th} percentile, and as an underprediction when they fell below the 5^{th} percentile.

3.6.2. Results and Discussion

For ease of exposition, we present our results and analysis in Table 3 below. Like in our previous studies, participants overestimated mean preference differences for both consensus levels. Furthermore, as predicted, participants overestimated the mode, and underestimated the mode's probability and the probability of the smallest positive difference (i.e. 0.5), and overestimated the probability of the maximum difference. Particularly striking is that for both consensus levels, participants severely overpredicted the probability of maximum differences by more than 6 times their actual probability. A majority of participants committed each of these errors. No order effects were observed for any of the estimates.

Table 2. Results and analyses of experiment 3. In the 74% and 94% consensus columns, parenthesis contain standard deviations and the percentage erring are tested with a binomial test.

Parameter from Distribution of A minus B	Actual Parameter Value 74% Consensus	Actual Parameter Value 94% Consensus	Statistic	Participants' Estimate: 74% Consensus	Participants' Estimate: 94% Consensus	Test: 74% Consensus vs Actual	Test: 94% Consensus vs Actual
Mean	0.55	1.32	Average Estimate	1.22 (0.48)	2.36 (0.64)	$t(199) = 19.67$, $p < .001$	$t(199) = 22.98$, $p < .001$
			Percentage Erring	90.5% Test Against 50% $p < .001$	94% Test Against 50% $p < .001$		
Mode	1.0	1.0	Average Estimate	2.47 (1.40)	2.82 (1.18)	$t(199) = 14.75$, $p < .001$	$t(199) = 21.69$, $p < .001$
			Percentage Erring	75.5% Test Against 50% $p < .001$	85.5% Test Against 50% $p < .001$		
Mode Probability	0.28	0.20	Average Estimate	0.06* (0.10)	0.09 (0.12)	$t(199) = -31.36$,* $p < .001$	$t(199) = -13.06$, $p < .001$
			Percentage Erring	98%* Test Against 50% $p < .001$	90% Test Against 50% $p < .001$		
Smallest Positive Difference Probability	0.28	0.19	Average Estimate	0.06* (0.10)	0.05 (0.09)	$t(199) = -31.36$,* $p < .001$	$t(199) = -21.74$, $p < .001$
			Percentage Erring	98%* Test Against 50% $p < .001$	93% Test Against 50% $p < .001$		
Maximum Difference Probability	0.01	0.03	Average Estimate	0.10 (0.10)	0.20 (0.18)	$t(199) = 13.83$, $p < .001$	$t(199) = 13.78$, $p < .001$
			Percentage Erring	82% Test Against 50% $p < .001$	81.5% Test Against 50% $p < .001$		

* For the 74% consensus level, the actual modal difference (not the bootstrapped 95% of the modal value reported in the table) is 0.5, which is also the smallest positive difference. Because of this, these parameters have the same results in Table 2.

Using another real-world dataset, experiment 3 provides further evidence that participants infer from consensus information preference differences that are too large. Rather than asking participants directly to estimate mean preference differences, in experiment 3 participants' beliefs were revealed by the distributions of preference differences that they built. The elicited distributions show that it is the underestimation of the likelihood of small differences that causes participants not only to overestimate mean preference differences, but also the likelihood of maximum differences.

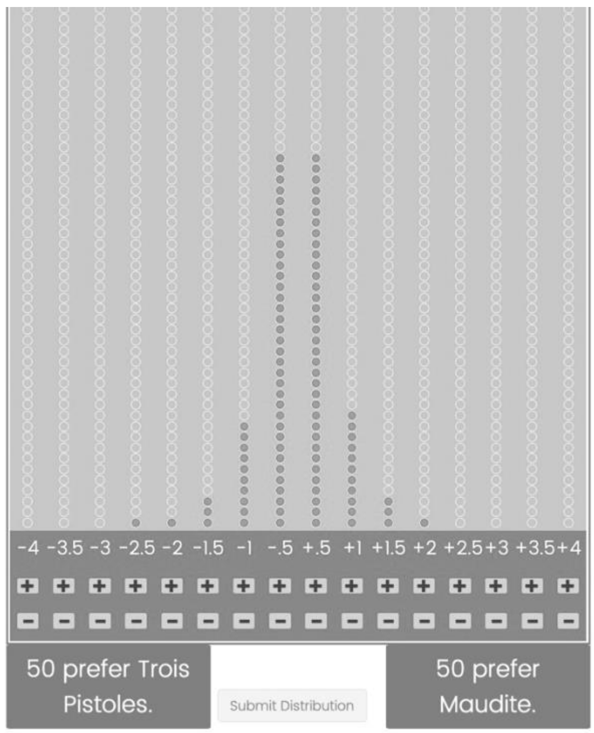
Note that we do not believe that people encountering polls spontaneously conjure a distribution and draw their inferences about preference differences from it. Instead, we use the distribution builder to capture the visceral feeling when seeing that 94% prefer Pliny the Elder: damn, that must be a much tastier brew!

Lastly, it should be noted that distributions for the 74% and 94% consensus levels did not depend on which distribution participants were asked to build first. The absence of an order effect suggests that participants are not naively anchoring on the consensus level they had seen before, but instead seem to spontaneously infer magnitude of preference differences when being confronted with consensus information.

3.7. Experiment 4: Debiasing with a Reference

In experiment 4, we test whether showing participants the preference distribution for a 50% consensus case will make them realize that, in general, small preference differences are more likely than large differences, and thus help reduce overestimation of mean preference differences for consensus levels other than 50%. As in experiment 3, we ask participants to build the distribution of preference differences for the 94% consensus beer-pairing. Half of the participants, however, are first shown an actual distribution of preference differences for a beer-pair displaying a 50% consensus level. This 50% consensus level distribution is nearly symmetrical with descending staircases away from zero: one into the positive numbers for those who prefer beer A and one into the negative numbers for those who prefer beer B (see Figure 8).

Figure 8. The 50% consensus distribution shown to half of the participants in experiment 4.



3.7.1. Method

We recruited 300 participants via Prolific to complete our online study in exchange for £0.90 (48% female, M_{age} = 27.63, *SD* = 9.67). All participants were asked to build the distribution of preference differences for the 94% consensus beer-pair as in experiment 3. Half of the participants were first told that they would see an example of the task they were about to complete for a different beer-pair—Trois Pistoles and Maudite—rated by users on beeradvoactes.com. Participants further learned that the percentage of users liking each beer is 50%, and were provided with a picture of the beer bottles. They were then asked, “What do you think the distribution of ratings looks like in this case?” Participants read the same instructions for how to use the distribution builder as in experiment 3, and were told, “Below we have already filled in the correct answer for this pair of beers, once you advance you will complete this task for a different pair of beers with different information.” Upon advancing they were then shown the beer-pair with 94% consensus and asked to complete the distribution builder. As in experiment 3, participants were incentivized with a 20-cent bonus if their distribution matched the true distribution.

Like before, we predicted participants in the control condition would: overestimate the mean, overestimate the mode, underestimate the probability of the mode, underestimate the probability of the smallest positive difference (i.e. 0.5), and overestimate the probability of the maximum difference. We predicted that a majority of participants in the control condition would commit each of these errors, but that the proportion doing so in the debias condition would be lower. Furthermore, we predicted that those in the debias condition would show less of over- and underestimation of the parameters in question.

3.7.2. Results and Discussion

We display all our results and analyses in Table 3 below. We replicate all of the findings from experiment 3 in the control condition. In contrast, the percentage of participants erring in the debias condition was reduced by at least 30%, and the degree to which they erred was significantly lessened. For the probability estimate of the smallest positive difference, the debiasing treatment removed overestimation entirely.

These results suggest that participants aren’t merely changing their estimate of the mode and shifting their faulty shaped distribution leftward, but are adjusting the shape of the distribution itself. Most importantly, participants are correctly updating their beliefs about the probability of small differences occurring in high consensus levels like 94%.

Table 3. Results and Analyses of experiment 4. In the Control and Debias Condition columns, parenthesis contain standard deviations and the percentage erring are tested with a binomial test.

Parameter from Distribution of A minus B	Actual Parameter Value	Statistic	Participants' Estimate: Control Condition	Participants' Estimate: Debias Condition	Test: Control vs Actual	Test: Debias vs Actual	Test: Control vs Debias
Mean	1.32	Average Estimate	2.34 (0.65)	1.71 (0.78)	$t(149) = 19.17$, $p < .001$	$t(149) = 6.03$, $p < .001$	$t(298) = 7.65$, $p < .001$
		Percentage Erring	94.7% Test Against 50% $p < .001$	60.7% Test Against 50% $p = .011$			$\chi^2(1) = 48.04$, $p < .001$
Mode	1.0	Average Estimate	2.72 (1.28)	1.66 (1.34)	$t(149) = 16.50$, $p < .001$	$t(149) = 6.02$, $p < .001$	$t(298) = 7.02$, $p < .001$
		Percentage Erring	77.3% Test Against 50% $p < .001$	44% Test Against 50% $p = .165$			$\chi^2(1) = 33.54$, $p < .001$
Mode Probability	0.20	Average Estimate	0.10 (0.12)	0.16 (0.13)	$t(149) = -11.01$, $p < .001$	$t(149) = -3.37$, $p < .001$	$t(298) = -4.77$, $p < .001$
		Percentage Erring	88.7% Test Against 50% $p < .001$	59.3% Test Against 50% $p = .027$			$\chi^2(1) = 32.03$, $p < .001$
Smallest Positive Difference Probability	0.19	Average Estimate	0.06 (0.09)	0.20 (0.19)	$t(149) = -17.21$, $p < .001$	$t(149) = 0.55$, $p = .586$	$t(298) = -8.10$, $p < .001$
		Percentage Erring	90% Test Against 50% $p < .001$	51.3% Test Against 50% $p = .807$			$\chi^2(1) = 52.25$, $p < .001$
Maximum Difference Probability	0.03	Average Estimate	0.20 (0.19)	0.09 (0.14)	$t(149) = 11.05$, $p < .001$	$t(149) = 5.25$, $p < .001$	$t(298) = 6.00$, $p < .001$
		Percentage Erring	78% Test Against 50% $p < .001$	43.3% Test Against 50% $p = .121$			$\chi^2(1) = 36.33$, $p < .001$

4. General Discussion

In this paper, we examined how consensus information from polls affects peoples' inferences about utility differences between the poll-choice options. In our empirical investigation, we focused on understanding (1) how people draw inferences from consensus information, (2) why these inferences may deviate from reality, (3) how said inferences can impact people' decisions, and (4) how inferences can be debiased. Findings from 6 pre-registered experiments demonstrate that people overestimate utility differences inferred from consensus information (experiments 1a, 1b, 1c, 3, and 4). This occurs even when participants are sophisticated in the judged domain (experiment 1a), when they have their own subjective beliefs about the choice options (experiment 1b), when participants have substantial background knowledge on data analytics (experiment 1c), and when financial incentives are at stake (experiments 1a, 1c, 3, and 4). Our data demonstrate that consensus information causes people to overestimate how much better a majority preferred option is than the minority preferred option as they intuitively gravitate toward large differences and neglect small differences (experiments 3 and 4). These overestimations lead to shifts in choice shares depending on whether preference data are displayed in consensus format or as average preference ratings (experiment 2). And the biases are greatly reduced when people are given a visual reference, albeit an obvious one, which causes them to adjust their beliefs about large and small differences (experiment 4).

In addition to the experiments reported above, we also explored if participants overestimate utility differences from consensus information because they focus too much on the majority's preferences. To this end, we ran an experiment (see appendix experiment S1) where we randomly assigned participants to focus on the majority, minority, both, or neither (a control), by asking them to report ratings for the different group's liking of each option. We expected that when doing this only for the minority group, compared to the control, the degree of overestimation would be attenuated. However, we instead found that compared to the control condition, participants in all conditions, whether they focused on the majority, the minority, or both, reduced the degree of overestimation. The reduction in overestimation, however, was tiny compared to the debiasing effects observed in experiment 4.

4.1. Theoretical Contributions

The theoretical contribution of our work is threefold. First, while a vast amount of research has examined the psychological processes by which people draw inferences about a target (e.g., Broniarczyk and Alba 1994; Dick, Chakravarti, and Biehal 1990; Evangelidis and van Osselaer 2019; Ford and Smith 1987, Ganzach and Krantz 1990, 1991; Gunasti and Ross 2009; Jaccard and Wood 1988; Johnson and Levin 1985; Kardes et al. 2004a; Kardes, Posavac, and Cronley 2004b; Slovic and MacPhillamy 1974), it has not examined whether said inference processes lead to accurate inferences, estimates, or judgments. In many cases it is difficult—or even impossible—to examine the accuracy of people' inferences because there are no normative benchmarks to which said inferences can be compared. Our settings allow for such comparisons because it is possible to compare estimates to observed distributions of real-world ratings, or to simulate a normative benchmark—the likelihood distribution of utility differences—for a given consensus level. Our work thus extends prior research on decision-makers' inferences by demonstrating a robust bias that arises from inference-making processes.

Second, we contribute to the understanding of social influence, specifically of social proof (Cialdini 2007). Social proof denotes the phenomenon that people conform to the behavior of others. For instance, hotel guests are more likely to reuse their towels when learning that a majority of other hotel guests do so (Goldstein, Cialdini and Griskevicius 2008). As the name "social proof" indicates, such conforming behavior is explained by people' beliefs that others may have more accurate preferences or superior information (Burnkrant and Cousineau 1975; Kelley 1967), especially in ambiguous situations in which they are uncertain about their own preferences (Young, Vosgerau, and Morewedge 2014). Since people conform out of a belief that others' responses provide diagnostic information, social proof often leads to private acceptance as well as public compliance (Cialdini 2007). Consensus information in polls apprises people of the option being preferred by the majority, so it constitutes a form of social proof (Cialdini 2007). Our findings hence suggest an additional explanation why social proof is so powerful in making people conform to the preferences and choices of others. People not only believe others' preferences to be diagnostic, they are also likely to overestimate others' utility differences, thereby wrongly inferring that the majority-preferred option is better than it actually is, or inferring that the minority-preferred option is worse than it actually is. Thus, social proof consists not only of conforming to the preferences and choices of others, it also involves exaggeration of others' preference differences.

Our third theoretical contribution concerns the consequences of our findings for preference inferences over time. Conforming to the preferences of others over time can result in self-reinforcing social influence effects whereby the number of people whose preferences are mimicked grows with the number of people conforming to others' preferences, leading to so-called "preference cascades." Investors, for example, are more likely to follow other investors' coverage of a firm as the number of investors following that firm increases (Rao, Greve, and Davis 2001), and higher levels of scarcity of a good can increase further sales of the scarce good (Banerjee 1992; Van Herpen, Pieters, and Zeelenberg 2009). According to our findings, exaggeration of social influence effects not only happen over time as in the case of preference cascades, they can also occur instantaneously as people exaggerate the preferences to which they conform. Preference cascades, in turn, may be further accelerated by people’ exaggeration of the preferences that they conform to.

4.2. Implications for Polls with more than Two Options

In our experiments, we examine polls and consensus information in which two options face off against one another. Polls, however, may contain more than two options. In multiple option polls, people will still conflate consensus with utility differences and infer the latter from the former. Because there are more than two options, however, there are also more comparisons, which increases the number of erroneous inferences that people may make. Studying the inferential processes and resulting biases in multiple option polls may be an interesting avenue for future research.

4.3. Managerial and Public Policy Implications

As the results of experiment 2 show, managers, politicians, public policy makers—in short anyone interested in influencing people’ judgments and choices—can strategically choose how to display aggregate preferences. In many cases, displaying others’ preferences as consensus rather than average ratings that show a utility difference will make the majority-preferred option more attractive, and hence increase its choice share. A company engaging in time-consuming data collection by eliciting ratings of their own and competitor products may instead conduct simple polls. Poll results (consensus information) will in many cases be more persuasive than more elaborate average ratings. As mentioned in the introduction, Pepsi has been using this strategy successfully for the last 30 years to gain market share from its main rival.

People, on the other hand, should be wary of the inferences they draw from consensus information and seek out more continuous measures that can help to better inform their choices. For example, an individual wanting to bet on the spread (the amount of points the winning team is favored by) of the next Super Bowl might go to ESPN.com and see that one team is predicted by 72% of experts to win. Our research would predict that this person will likely infer that the winning team will easily cover the spread and thus place a sizeable wager. If the bettor also read the predicted scores that generated that consensus level, they might be surprised to learn that the experts think the winning team will only win by less than a touchdown (a close game). Knowing this, the bettor is likely to adjust their wager downward given the additional perceived risk.

Obviously, this is not restricted to sports betting but is applicable to any domain in which people infer the utility difference from consensus information, such as polls about public opinion concerning political candidates (e.g., “Which candidate do you trust more to revive the economy?”), sports contenders (“Who will win the Super Bowl?”), the next best product to create or keep (“Cast your DEWcision for the flavor to keep on shelves.”; prnewswire.com 2016), or product choices (“78% prefer our product over our competitor’s product”). As people tend to form unrealistic expectations about political candidates, athletes/teams, and products when learning consensus information, they are likely to be disappointed once their favorite candidate is voted into office, they have cast a bet on their preferred contender, or have bought their preferred product.

There are, however, situations where consensus is actually more informative than knowledge of the underlying preferences. For example, in the 2020 US democratic primary, democrats were trying to figure out which candidate, Bernie Sanders or Joe Biden, would have better chances of beating Donald Trump in the presidential election. Sanders had a minority of Democrats passionately supporting him and feeling very lukewarm about Biden. Biden had the majority of Democrats' support, but most were not very passionate about his candidacy. Chances of beating Trump in the presidential election are here indicated by consensus information, not by the underlying strength of preferences for each candidate. In general, which format for expressing aggregate consumer preferences is more advantageous depends on the goal at hand. Probability of superiority is indicated by consensus information, whereas satisfaction with and betting on actual outcomes is better predicted by continuous preference measures.

4.3. Implications for Preference Elicitation

Researchers use a variety of methods to elicit preferences, such as choice, preference ratings, or WTP. In doing so, they need to be aware of the kind of information that is acquired in the aggregate. When measuring consumer preferences for two options through choice, researchers learn about the ordinal preference ranking of individual people, and through aggregation about consensus, that is the ordinal preference ranking by the majority. When assessing preferences through ratings, researchers learn about utility differences. Note that the two convey different information. Consensus cannot be inferred from average ratings (even when the skew of both distributions is known). To infer consensus, a researcher needs to have access to the individual ratings by each consumer. Likewise, average ratings cannot be inferred from consensus information, as we have shown in this paper.

In choice experiments, it is quite common for researchers to test choice shares against 50%. In case of a null result, many researchers state that respondents are indifferent between the choice options. As we have demonstrated this is an unwarranted inference because indifference—a utility difference—cannot be inferred from consensus information. A 50% choice share may indicate indifference, but may also be indicative of uniform or bipolar preference distributions. In fact, any symmetrical preference distribution will lead to a 50%-50% choice share[4]. Consider the example that we discussed in the theory section about the US presidential election in 2016. It would have been foolish to infer from a poll showing 52% of voters preferring Trump over Hillary that the average voter likes Trump as much as Hillary.

[4] We thank Irene Scopelliti for pointing out this application of our findings.

References

Autoguide.com (2017). url: https://www.autoguide.com/auto-news/2017/09/poll-toyota-camry-or-honda-accord-.html. accessed on 6/18/2020.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, *107*, 797-817.

Breitbart.com (2021). url: https://www.breitbart.com/politics/2021/02/16/poll-donald-trump-favorability-75-among-republicans-mitch-mcconnell-underwater-at-15/. accessed on 10/20/2021.

Broniarczyk, S. M., & Alba, J. W. (1994). Theory versus data in prediction and correlation tasks. *Organizational Behavior and Human Decision Processes*, (1), 117-139.

Burnkrant, R. E., & Cousineau, A. (1975). Informational and normative social influence in buyer behavior. *Journal of Consumer Research*, *2*, 206-215.

Cialdini, R. B., & Cialdini, R. B. (2007). *Influence: The Psychology of Persuasion* (Vol. 55, p. 339). New York: Collins.

Dick, A., Chakravarti, D., & Biehal, G. (1990). Memory-based inferences during consumer choice. *Journal of Consumer Research*, *17*, 82-93.

ESPN (2020). url: https://www.espn.com/nfl/story/_/id/28576384/super-bowl-liv-score-predictions-espn-experts-pick-49ers-chiefs. accessed on 6/18/2020.

Evangelidis, I., & Van Osselaer, S. M. (2018). Points of (dis) parity: Expectation disconfirmation from common attributes in consumer choice. *Journal of Marketing Research*, 55, 1-13.

——— and ——— (2019). Interattribute Evaluation Theory. *Journal of Experimental Psychology: General*, 148, 1733-1746.

Fernberger, S. W. (1920). Interdependence of judgments within the series for the method of constant stimuli. *Journal of Experimental Psychology*, *3*, 126.

Fisher, M., & Keil, F. C. (2018). The binary bias: A systematic distortion in the integration of information. *Psychological Science*, *29*, 1846-1858.

Fisher, M., Newman, G. E., & Dhar, R. (2018). Seeing stars: How the binary bias distorts the interpretation of customer ratings. *Journal of Consumer Research*, *45*, 471-489.

Forbes (2021). url: https://www.forbes.com/sites/alisondurkee/2021/06/25/majority-of-americans-support-abortion-poll-finds---but-not-later-in-the-pregnancy/?sh=24f9d2fb5074 . accessed on 11/3/2021.

Ford, G. T., & Smith, R. A. (1987). Inferential beliefs in consumer evaluations: An assessment of alternative processing strategies. *Journal of ConsumerRresearch*, *14*, 363-371.

Ganzach, Y., & Krantz, D. H. (1990). The psychology of moderate prediction: I. Experience with multiple determination. *Organizational Behavior and Human Decision Processes*, *47*, 177-204.

——— and ——— (1991). The Psychology of Moderate Prediction: II. Leniency and Uncertainty. *Organizational Behavior and Human Decision Processes*, 48, 169-92.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, *4*, 133-151.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*, 472-482.

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, *9*, 1.

Gunasti, K., & Ross, W. T. (2009). How inferences about missing attributes decrease the tendency to defer choice and increase purchase probability. *Journal of Consumer Research*, *35*, 823-837.

Harper, Maxwell F. and Joseph A. Konstan (2015). The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5, 4: 19:1–19:19. <https://doi.org/10.1145/2827872>

Heintz, R. K. (1950). The effect of remote anchoring points upon the judgment of lifted weights. *Journal of Experimental Psychology*, *40*, 584.

History.com (2020). url: https://www.history.com/news/cola-wars-pepsi-new-coke-failure . accessed on 10/07/2020.

Huber, J., Ariely, D., & Fischer, G. (2002). Expressing preferences in a principal-agent task: A comparison of choice, rating, and matching. *Organizational Behavior and Human Decision Processes*, *87*, 66-90.

Huh, Y. E., Vosgerau, J., & Morewedge, C. K. (2014). Social defaults: Observed choices become choice defaults. *Journal of Consumer Research*, *41*, 746-760.

Jaccard, J., & Wood, G. (1988). The effects of incomplete information on the formation of attitudes toward behavioral alternatives. *Journal of Personality and Social Psychology*, *54*, 580.

Johnson, R. D., & Levin, I. P. (1985). More than meets the eye: The effect of missing information on purchase evaluations. *Journal of Consumer Research*, *12*, 169-177.

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

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Psychology for behavioral economics. *Heuristics and Biases: The Psychology of Intuitive Judgment*, 49-81.

Kardes, F. R., Cronley, M. L., Kellaris, J. J., & Posavac, S. S. (2004). The role of selective information processing in price-quality inference. *Journal of Consumer Research*, *31*, 368-374.

Kardes, F. R., Posavac, S. S., & Cronley, M. L. (2004). Consumer inference: A review of processes, bases, and judgment contexts. *Journal of Consumer Psychology*, *14*, 230-256.

Keeney, R. L., Raiffa, H., & Meyer, R. F. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge University Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In *Nebraska Symposium on Motivation*. University of Nebraska Press.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*, 215-233.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*S, 85.

Pepsimax.ie (2019). url: https://pepsimax.ie/. accessed on 10/07/2020.

Phillymag.com (2018). url: https://www.phillymag.com/foobooz/2018/01/23/vote-best-cheesesteak-2018/. accessed on 6/18/2020.

Prnewswire.com (2016). url: https://www.prnewswire.com/news-releases/dear-dew-nation-dewcision-2016-gives-you-the-power-to-choose-between-two-legendary-dew-flavors-300255397.html. accessed on 10/07/2020.

Quentinandre.net (2022). url: https://quentinandre.net. DOI: 10.5281/zenodo.166736, accessed on: 10/26/2022.

Rao, H., Greve, H. R., & Davis, G. F. (2001). Fool's gold: Social proof in the initiation and abandonment of coverage by Wall Street analysts. *Administrative Science Quarterly*, *46*, 502-526.

Sharpe, W. F., Goldstein, D. G., & Blythe, P. W. (2000). The distribution builder: A tool for inferring investor preferences. *preprint*.

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, *55*, 150.

Slovic, P., & MacPhillamy, D. (1974). Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance*, *11*, 172-194.

Van Herpen, E., Pieters, R., & Zeelenberg, M. (2009). When demand accelerates demand: Trailing the bandwagon. *Journal of Consumer Psychology*, *19*, 302-312.

Wever, E. G., & Zener, K. E. (1928). The method of absolute judgment in psychophysics. *Psychological Review*, *35*, 466.

Web Appendix for “Consumers Believe if 90% Prefer A Over B, A Must be Much Better than B. They Are Wrong.”

Appendix A: Simulations

In this section, we explain our simulations and display their results graphically. In our experiments, we preregistered our normative benchmark based on a simulation with uniform distributions. For robustness, we conducted additional simulations where we used truncated normal distributions and, additionally, we varied the degree of correlation between the utilities for each option within a group. Across specifications, the key takeaways remain unchanged. For our simulations, we directly generate the average utilities for each option by each group according to whichever distribution that simulation uses. From this we can then compute utility differences, plot them, and calculate summary statistics. Finally, we also examine how bimodal distributions affect our results.

Appendix A.1: Simulation Using Uniform Distributions

To simulate utility differences, we start with generating four variables to represent four mean utilities. We need four mean utilities because we have two choice options A and B, and two groups, those who prefer A over B—henceforth denoted by [A]—and those who prefer B over A—denoted by [B]. We now label the four variables that we will generate in the following way: Variable “[A]A” indicates the mean utility for option A by those who prefer A over B, variable “[A]B” indicates the mean utility for option B by those who prefer A over B, [B]A indicates the mean utility for option A by those who prefer B over A, and finally [B]B indicates the mean utility for option B by those who prefer B over A.

To generate all possible permutations of utility means for options A and B by those who prefer A and those who prefer B, we start by generating all permutations of mean utilities with replacement. The total number of permutations depends on how many values the means can take. For example, if we measured utilities on a scale from 1 to 10, then mean utilities could take on any value from 1 to 10 inclusive any number of decimal places. The more values we allow for by increasing the number of decimal places, the more fine grained will be the simulation of mean utilities. At the same time, the more values we generate the larger becomes the number of total permutations, quickly posing a computational burden. To balance the two opposing effects, we chose to generate values between 1 and 100 (so mean utilities can take on integers between 1 and 100), which resulted in 100^4 = 100,000,000 number of permutations. Values between 1 and 100 are equivalent to measuring utilities on a scale from 1 to 10 and allowing for one decimal place. Our dataset thus contains four variables with 100,000,000 rows.

In a second step, we now select those permutations that satisfy the preference-ordering implied by the labeling of the four variables. Data rows in which mean utilities satisfy the rank ordering [A]A > [A]B and [B]B > [B]A are kept, and if they do not, they are deleted.

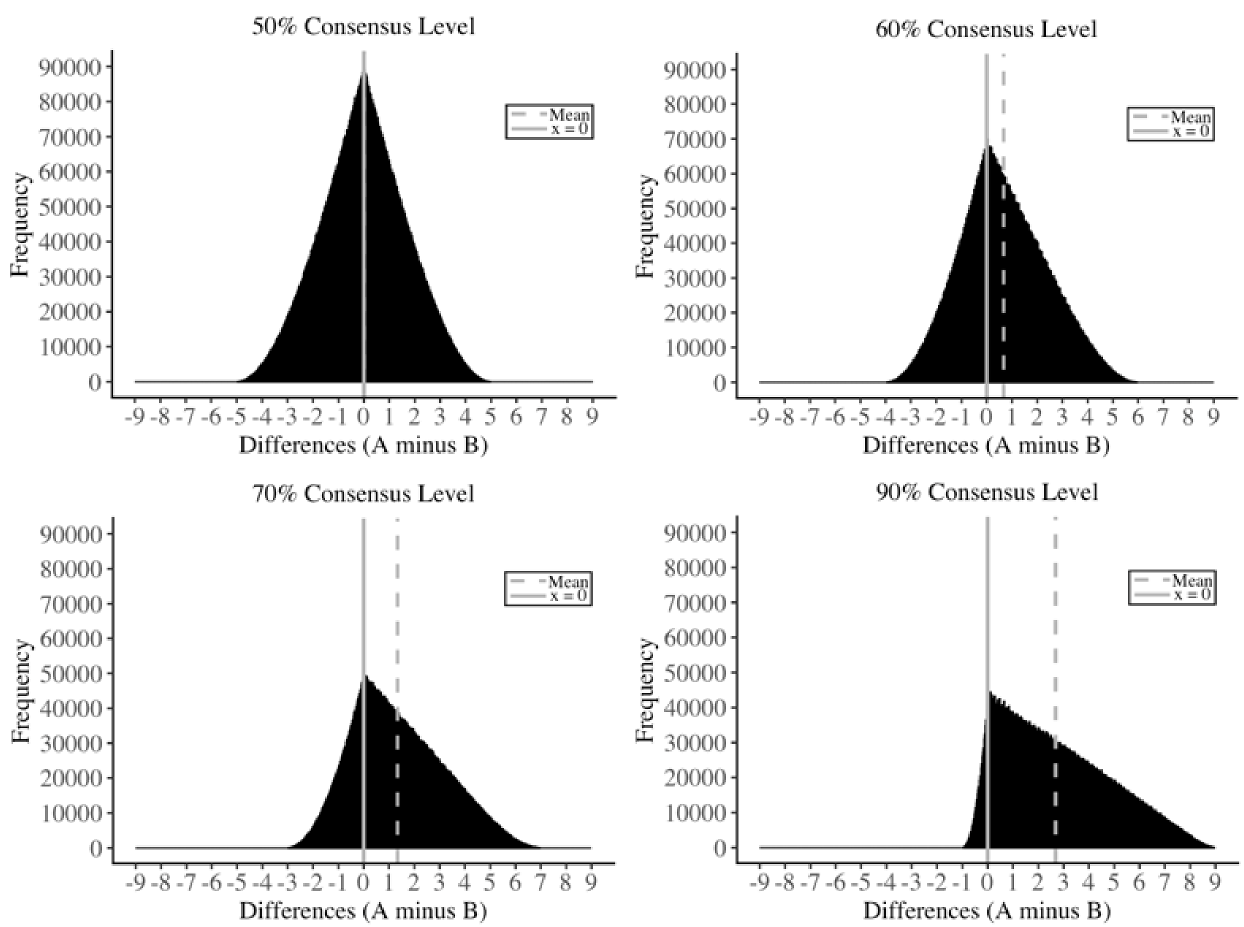
In a third step, we generate four consensus levels, 50%, 60%, 70%, and 90%. The consensus levels determine the weight with which mean utilities are weighted when calculating differences in mean utilities. Because consensus levels for B are one minus the consensus level for A, the weights of the utilities of [A] are given by $w_{[A]}$ and the weights of utilities of [B] are given by $1 - w_{[A]}$. With four weights $w_{[A]} = 0.5, 0.6, 0.7,$ and 0.9, we can now generate the mean utilities for option A and option B with the following formulas:

$$A = (w_{[A]}) * [A]A + (1 - w_{[A]}) * [B]A$$

$$B = (w_{[A]}) * [A]B + (1 - w_{[A]}) * [B]B$$

Finally, we calculate the difference in mean utilities of A and B. We divide utilities by 10 to make the resulting graphs analogous to the 10-point rating scale. Each row in our final dataset is a unique permutation of [A]A, [A]B, [B]A, and [B]B, so each row represents an answer to the question: “How much better is A than B?” These differences in mean utilities are depicted on the x-axis of Figure S2, the y-axis shows the frequency of each answer.

Figure S2. Simulation of the frequency of differences in average utilities (average rating of majority preferred option A minus average rating of minority preferred option B, ratings are on a 1-10 point scale) for four consensus levels. The vertical dashed line marks the mean of the distribution and the vertical solid line marks a difference of zero (i.e. the case where A and B are equivalent).



Appendix A.2: Simulation Using Truncated Normal Distributions with Specified Correlations

To simulate utility differences derived from truncated normal distributions requires specifying the variance and the mean of the distributions. In addition to this, we specify the degree and sign of the correlation between a group's average utility for their preferred option and their non-preferred option. We create two datasets, one for those who like A and one for those who like B, each containing that group's average utility for option A and option B drawn from a truncated normal multivariate distribution. The process for creation is identical for each dataset. Like in our uniform distribution, we create a fine-grained estimate by using 100 point scales. We set the means of the distributions at the midpoint (55), and the standard deviation at 10.3. Then we specify correlations of preferences at the two extremes, highly negative and highly positive. In one set of simulations, preferences of A and B within groups correlate with about −0.8, and in another, they correlate with about +0.5. We then randomly draw 100,000 pairs of observations from a two-variable, multivariate normal distribution. Each pair consists of the average utility for option A and the average utility for option B. We then implement the preference constraint for that group (i.e., A[A] > A[B] for those who like A and B[B] > B[A] for those who like B). Next, we truncate the distribution by dropping all observations where the rating for option A or B was outside our rating scale (i.e., less than 1 or greater than 10 on a 10-point scale). Doing that process twice generates the two datasets, one for each group. We then check to make sure both datasets are of the same size since they are independently and randomly created, the process of implementing the preference constraint and truncation can eliminate a different number of observations for each of them. After determining which dataset has more observations, we randomly drop observations from the larger dataset until it is the same size as the other. We then merge the two datasets together. The resulting dataset now contains four columns, A[A], A[B], B[B], and B[A], one for each average rating of an option by a group. Then we create and add three more columns, one for the overall rating of A, one for the overall rating of B, and one for their difference, according to the same formulas written above in the uniform simulation. Finally, we divide utilities by 10 to make the resulting graphs analogous to the 10-point rating scale. Mainly due to imposing preference constraints which cuts the size of the dataset nearly in half, the resulting datasets contain more than 49,000 rows. Like the uniform distribution, each row in our final dataset is a permutation of [A]A, [A]B, [B]A, and [B]B, so each row represents an answer to the question: "How much better is A than B?" Compared to our uniform distribution, extreme differences are even less probable due to the use of truncated normal distributions. These differences in mean utilities are depicted on the x-axis of Figure S3 (positive correlation) and S4 (negative correlation), the y-axis shows the frequency of each answer.

There are a few key takeaways. 1) The results of both distributions are very similar to the uniform simulation and the means are slightly lower. This occurs because the truncated normal distribution limits the frequency of extreme values, making them less probable compared to other values, whereas a uniform distribution does not. If we increase the variance and make the truncated normal distribution more and more like the uniform, the means will increase. Nevertheless, all these distributions, uniform, truncated normal with a positive correlation, and truncated normal with a negative correlation, are strikingly similar in their means. 2) Importantly, they all have means and modes that are very small with respect to their consensus level. 3) The more negative the correlation is between A and B for each group, the more frequently larger utility differences appear. However, even with a negative correlation of −0.81, the mean is still low compared to the maximum of the scale. For example, in Figure S4 the mean at a 90% consensus is 2.14 on a 10-point scale.

Figure S3. Simulation of the frequency of differences in average utilities (average rating of majority preferred option A minus average rating of minority preferred option B, ratings are on a 1-10-point scale) for four consensus levels using a within group correlation of .54. The vertical dashed line marks the mean of the distribution and the vertical solid line marks a difference of zero (i.e,. the case where A and B are equivalent).

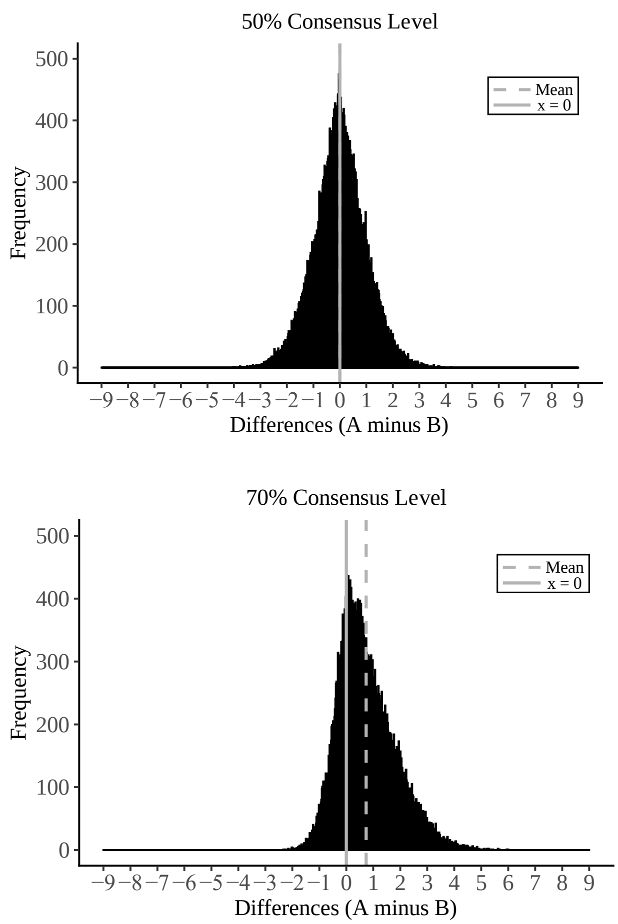
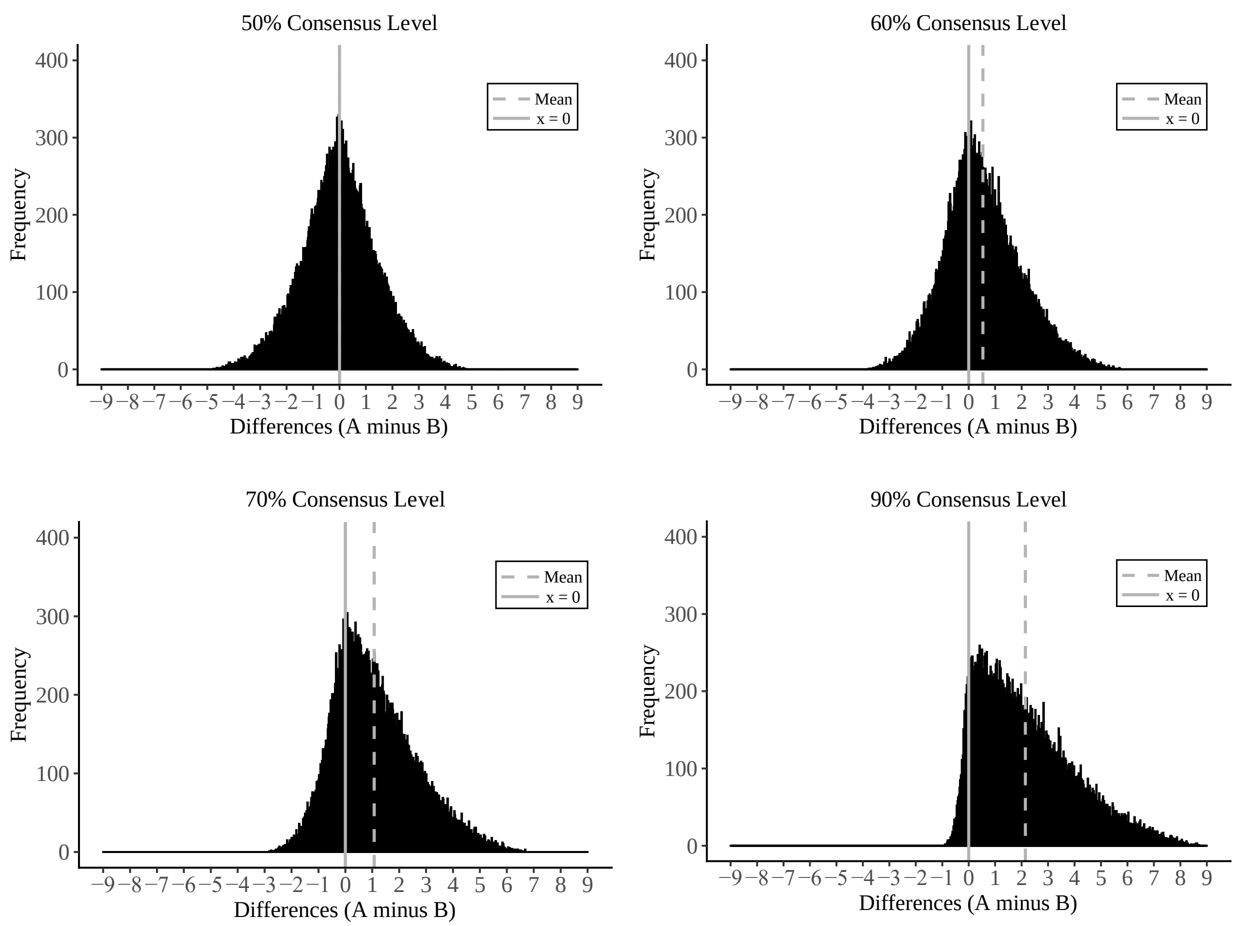


Figure S4. Simulation of the likelihood of differences in average utilities (average rating of majority preferred option A minus average rating of minority preferred option B, ratings are on a 1-10-point scale) for four consensus levels using a within group correlation of −0.81. The vertical dashed line marks the mean of the distribution and the vertical solid line marks a difference of zero (i.e., the case where A and B are equivalent).



Appendix A.3: Bimodal Simulation Using Truncated Normal Distributions with Specified Correlations

To simulate utility differences derived from bimodal distributions, we use the same method as our truncated normal distributions but further allow the means for A[A], A[B], B[B], and B[A] to vary from one another. The only requirements are the usual preference constraints. Since four means can vary, there are now many possibilities to use as inputs that will produce varying results. However, there are a few key takeaways. First, there is a general principle: as the difference in utilities between the preferred option and non-preferred option grows, so too will the mode, median, and mean of the resulting distribution. Second, unsurprisingly consensus levels determine how much this matters. As the level of consensus increases, the more the average utility difference reflects the majority group's preferences. In the extreme of 100% consensus only the majority's preferences and how far apart they are matter. For demonstration purposes, let's consider an extreme case. Those who like A and those who like B each have a mean utility for their preferred option of 90/100 and 10/100 for their non-preferred option, and an extreme, negative correlation of -.86 (SD around 11.2). Furthermore, 90% of people like option A. In this extreme case, the reader might be surprised to learn that the median of the distribution of utility differences is 52.47 and the mean is 49.69, both near the midpoint of the 100-point scale despite there being 90% consensus and a love, hate relationship. If we further shrink the frequency of possible outcomes by reducing the standard deviation to about 3.1, the median is 63.98 and the mean is 63.97. Remember, this is only one possibility in the inputs for our means and there are more possibilities compared to this one where the utilities are closer, and the average and modal utility difference would be smaller, than there are ones where the utilities are far apart, and the average and modal utility difference would be greater. Thus, although larger differences become more likely when distributions are bimodal and consensus levels are very high, the expected difference in utilities is only slightly higher compared to our previous specifications.

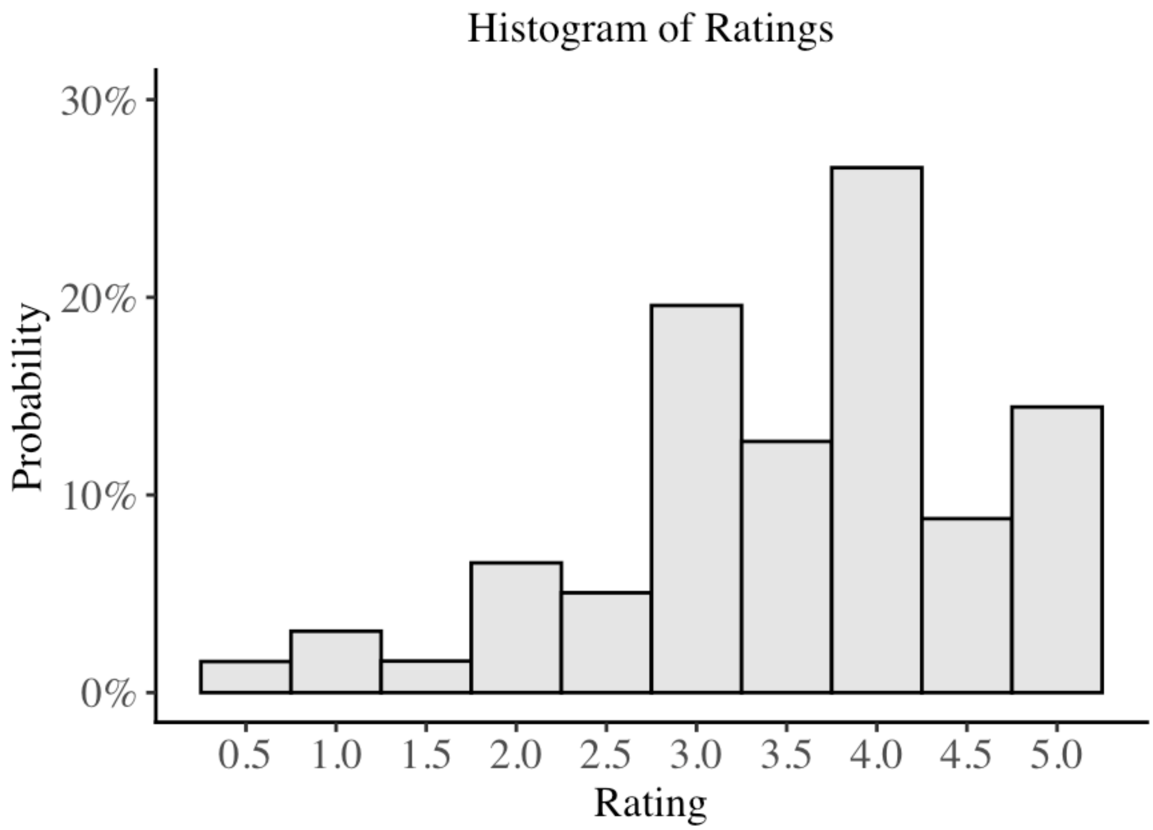
Appendix B: Jokes, Beers, and Movie Rating Dataset Descriptions

The jokes dataset comes from a project to build a joke recommendation system and is populated by anyone who visits the website (Goldberg, Roeder, Gupta, and Perkins, 2001; https://goldberg.berkeley.edu/jester-data/). The website has evolved over time but individuals who visit the website (http://eigentaste.berkeley.edu/) view one joke at a time and can rate the joke using a continuous sliding scale (at the time of our data) from less funny (encoded as −10) to more funny (encoded as 10). The website's joke recommendation system learns the prior jokes' ratings from each user and presents each successive joke to a user based of their ratings. Users can stop rating jokes and leave the website whenever they like. The data set we use comprises 100 jokes. The data include ratings from 24,938 users, who, on average, rated 24 jokes each.

The beer dataset comes from user ratings of beers on beeradvocate.com and can be found here: https://www.kaggle.com/rdoume/beerreviews. Reviewers can rate many aspects of the beer (e.g., taste, aroma, appearance) but also give an overall review, which we use as our rating. The overall review is rated from 1 to 5 points in 0.5 increments with 5 being the best. The scale and labels for each value seem to have changed over time but a post from 2014 reveals that when the scale was done in 0.25 increments the labels were "awful" to "world-class" (https://www.beeradvocate.com/community/threads/how-to-review-a-beer.241156/). The dataset contains 66,051 beers rated by 33,387 reviewers with more than 1.5 million reviews.

The movie dataset consists of ratings from movielens.org which is a movie recommendation site (https://grouplens.org/datasets/movielens/25m/). The dataset contains about 25 million reviews from 162,541 users of 59,047 movies that were rated from 1995 to 2019. A detailed summary can be found in the readme file available at the above link. Ratings are given from 0.5 to 5.0 with 0.5 increments. In this dataset, integer ratings are more common than near decimal ratings (e.g., ratings of 3, 4, or 5 are much more common than 3.5 or 4.5; see Figure S1 below).

Figure S1. Histogram of ratings for movies (N= 25000095).



Appendix C: Modal Analysis

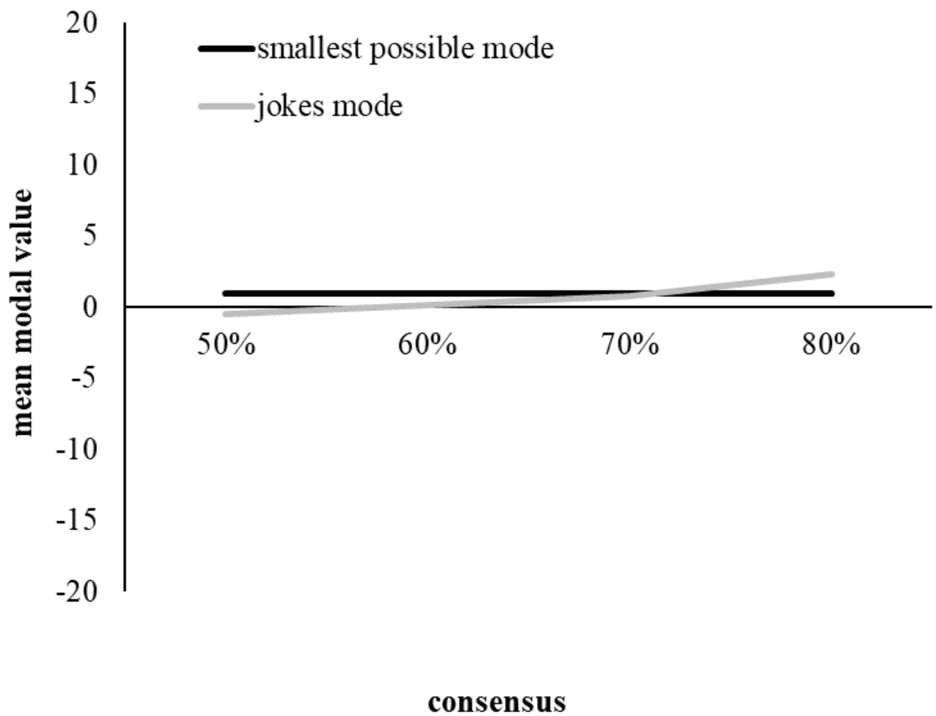
In this section, we examine the observed modes of rating differences in our real-world datasets (jokes, beers, and movies) to see how they compare to the predicted modes from our simulations. The beers and movies datasets contain many options (66,051 and 59,047 respectively) which means that there is an extremely large number of possible combinations. So, for both datasets, we reduce the number of options by keeping only beers with more than 500 reviews (65th percentile) and movies with more than 10,000 reviews (53rd percentile). We also eliminate reviewers who rated only one option. The resulting beers dataset has 589 beers with 173,166 unique combinations and the resulting movies dataset has 588 movies with 172,578 unique combinations. For the jokes data there are 100 jokes resulting in 4,950 unique combinations. Out of these combinations for each dataset, we only examine those with 50 or more observations to get stable estimates of the properties of the utility difference distribution for each combination.

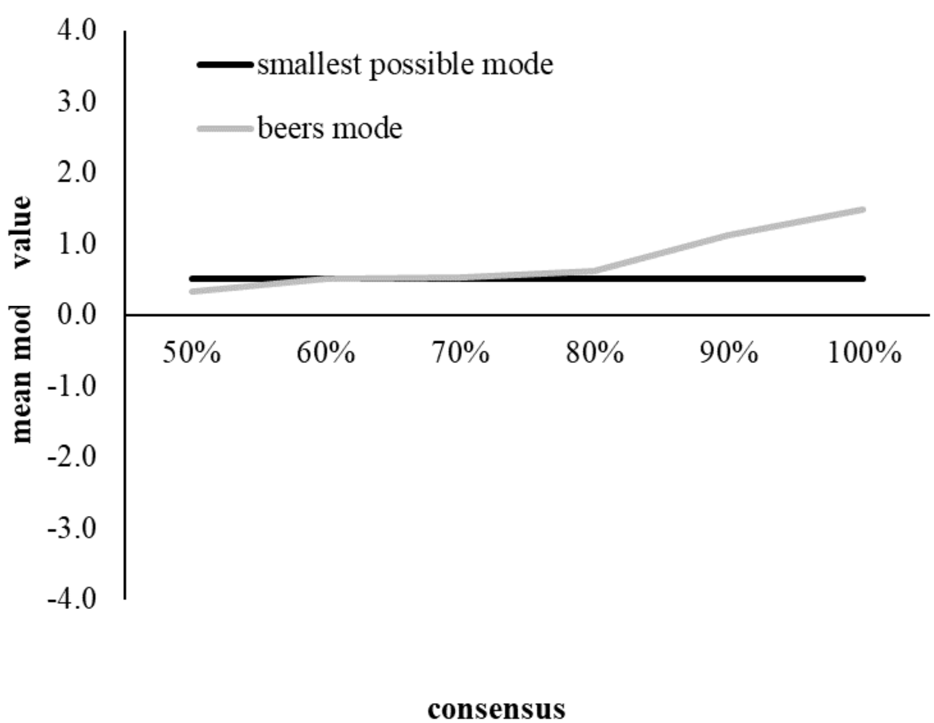
For each combination, we calculate the consensus level, average utility difference (excluding zeros as in our simulations), and number of observations. We determine the mode(s) for each combination and record the minimum modal value and the maximum modal value, which is the same in unimodal distributions but different in multimodal distributions. For multimodal distributions, we also calculate and record how many modes there are. We then create six bins for our consensus levels. The 50% bin contains consensus levels from including 50% up to and including 59%, the 60% bin contains consensus levels from including 60% up to and including 69% and so on up until the final 100% bin which contains only 100% consensus levels. Not all datasets contain all the highest bins because either no combinations are produced in that bin or because the combinations that are produced contain less than 50 observations.

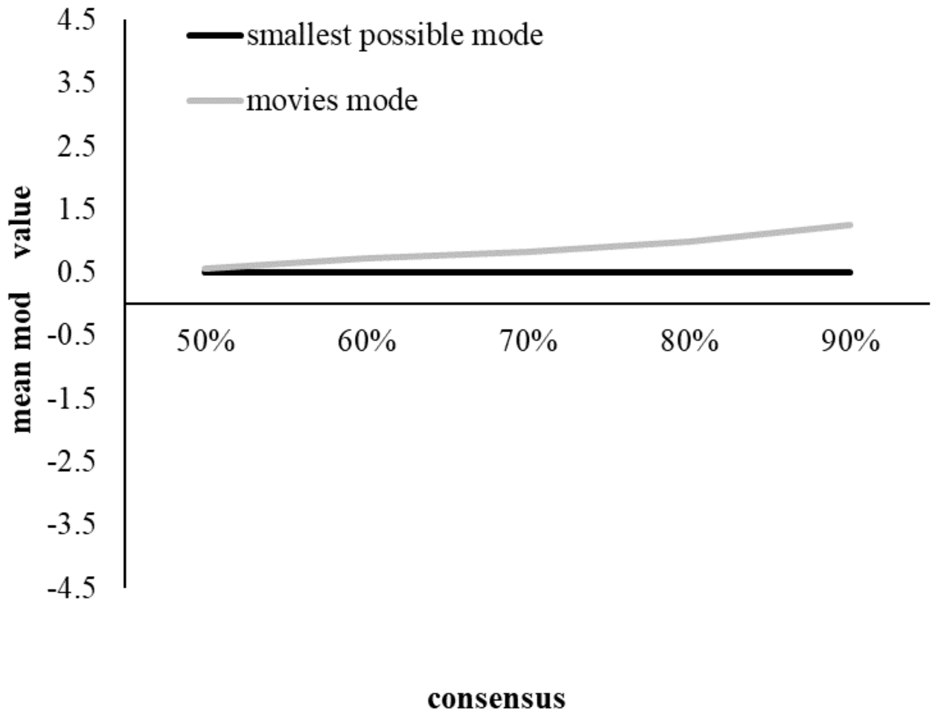
Because the overwhelming majority of distributions of rating differences is unimodal (over 90%), in the following we concentrate on these unimodal distributions. For each dataset, we plot the average mode by consensus level and compare them to the smallest possible positive mode (the grey line in Figure S5). The smallest mode is the lowest, positive, theoretically possible utility difference that could occur on the respective scale (i.e. 0.5 for beers and movies, and 1 for jokes). The closer the observed average modes are to this smallest possible mode, the closer our simulations mirror reality.

Our results reveal a few general key facts across datasets. First, the overwhelming majority of combinations (i.e., more than 90%) are unimodal. Second, the average maximum modal value (excluding unimodal distributions) with respect to each consensus level and dataset is small. For example, for beers at 60% it is 0.8 and at 90% it is 1.64, for movies at 60% it is 0.99 and at 90% it is 1.81, and for jokes at 60% it is 1.87 and at 80% it is 6. This is the most conservative approach we can take to examining the modal value. Third, as consensus increases so do the average modal values. Finally, and most importantly, within a consensus level, the average minimum and maximum modal values (without exclusions or examining only unimodal distributions) are quite small compared to what our experiments suggest that people intuit.

Figure S5. Average modes for unimodal distributions of rating differences for jokes, beers, and movies. The black line represents the smallest possible positive mode and the grey line is the actual minimum modes.







Appendix D: Experiment S1

In experiment S1, participants were asked to take the perspective of a poll's majority, minority, or of both groups. Consumers may overestimate utility differences because they may focus on a poll's majority and fail to consider how the minority may have rated both options, which would result in inflated estimates of utility differences. Accordingly, we predicted that participants instructed to take the perspective of the poll's minority would provide lower and more accurate estimates of utility differences than participants in a control condition and participants instructed to take the perspective of the majority.

Appendix D.S1.1. Method

We recruited 800 participants (US, UK, Canada, and Ireland) via Prolific to complete our online study in exchange for £0.38. Following our preregistration, we dropped anyone who took our survey more than once or failed to complete it, leaving us with 798 participants (59.5% female, M_{age} = 36.5, *SD* = 14.01).

Participants read that “500 people rated both Chardonnays below (Wine A and Wine B) on a scale from 1 to 9 (1 = poor quality and 9 = very high quality).” They saw images of the two bottles and learned that “80% rated Wine A higher than Wine B” and “20% rated Wine B higher than Wine A.” On the same screen, participants had to correctly select which scale the wines were rated on before being able to continue. Participants were then randomly assigned to one of four conditions.

In the ‘majority focus’ condition, participants were asked to think about the 80% of people who had rated Wine A as better, and to estimate how much they thought that majority on average may have liked both Wine A and Wine B. Participants were asked to enter values between 1 and 9 with up to one decimal, and to enter a higher value for Wine A than Wine B. The order of the questions was randomized. Analogously in the ‘minority focus’ condition, participants estimated how the 20% minority may have rated Wine A and Wine B. In the ‘both’ condition, participants were asked to provide estimates for both the majority and the minority (counterbalanced). In the ‘control’ condition, participants proceeded directly to the dependent variable without doing any of the rating tasks.

Participants in all conditions then saw the stimuli and consensus information again and answered, “What do you think is the most likely difference between the ratings of Wine A and

Wine B?" They were given two examples and then reported their answers on a scale from 0 to 8 with at most one decimal.

In addition to our predictions about how perspective taking will influence estimates of the most likely rating difference, we predicted that the majority of participants in the 'control' condition will provide estimates greater than the median of 1.7 from our simulations.

Appendix D.S1.2. Results and Discussion

As predicted, the majority of participants in the 'control' condition overestimated the most likely rating difference and provided a value higher than the median (98%, Pearson $\chi^2(1) = 189.31$, $p < .001$). Compared to the 'control', the percentage of participants overestimating was reduced in the 'majority' condition (98% vs. 91.9%, Pearson $\chi^2(1) = 8.09$, $p = .004$), the 'minority' condition (98% vs. 85.1%, Pearson $\chi^2(1) = 22.04$, $p < .001$), and the 'both' condition (98%% vs. 90%, Pearson $\chi^2(1) = 11.68$, $p = .001$). Likewise, the percentage of participants overestimating was lower in the 'minority' condition compared to the 'majority' condition (91.9% vs. 85.1%, Pearson $\chi^2(1) = 4.39$, $p = .036$). The 'minority' and 'both' conditions did not differ (85.1% vs. 90%, Pearson $\chi^2(1) = 2.21$, $p = .137$).

A slightly different pattern was observed for the degree of overestimation. Compared to the 'control' condition ($M_{\text{control}} = 4.62$, $SD = 1.92$), the degree of overestimation was reduced in all experimental conditions (all $p < .001$). However, none of the treatments differed significantly from one another ($M_{\text{majority}} = 3.83$, $SD = 1.91$ vs. $M_{\text{minority}} = 3.54$, $SD = 1.88$, $t(390) = 1.49$, $p = .138$; $M_{\text{minority}} = 3.54$, $SD = 1.88$ vs. $M_{\text{both}} = 3.70$, $SD = 1.82$, $t(394) = 0.85$, $p = .393$; $M_{\text{majority}} = 3.83$, $SD = 1.91$ vs. $M_{\text{both}} = 3.70$, $SD = 1.82$, $t(396) = 0.67$, $p = .500$).

In sum, we find that both the percentage of people overestimating, and their degree of overestimation are reduced when they consider either group's or both groups' utility for each option. The findings suggest that taking the perspective of a poll's minority, but also the perspective of a poll's majority can reduce the extent to which consumers overestimate average utility differences inferred from consensus information.

Chapter 2

Consumers (Wrongly) Believe that Others Like an Option Less When Those Others Express Indifference

Abstract

Consumers often make decisions not solely based on their own desires and opinions, but also those of others. Yet, relatively little is known about how consumers interpret and understand other people's preferences as a function of how those others express their preferences. We investigate how solicitors of others' preferences ("requestors") make inferences about solicitees' ("responders") liking of the options depending on responders' preference expressions. We compare two common expressions of preference: a preference, where responders pick one option out of the consideration set, with indifference, where responders equate all (or some of) the options under consideration by not preferring a single option. We theorize and demonstrate that requestors learning indifference (wrongly) believe that responders like a given (target) option less compared to when the target is the only option preferred. This effect persists even when responders explicitly state that they like the alternatives when expressing indifference. We demonstrate that perceptions of uncertainty underlie our results. When responders express indifference, requestors infer that responders are uncertain about how much they like the options, leading to devaluations of the target option. Our findings hold important implications for theories of preference formation, inference-making, and dyadic decision-making.

Consumers frequently solicit and use others' opinions as inputs for their own, individual, decisions as well as joint decisions. When shopping one might ask a friend to help them decide between jackets. When travelling one might ask an acquaintance from the area what local spots they prefer. When picking an activity for a date or day out with friends, one might desire to know their preference between the options. Ultimately one wishes to know how happy their date will be with the chosen option or which option their friend really likes and thinks is best.

Recent research in marketing has found that, whereas solicitors of others' preferences (hereafter referred to as "requestors") typically covet preference expressions, solicitees (hereafter referred to as "responders") often express indifference instead of a preference (Liu and Min 2020). For example, consider the case of a consumer deciding between two restaurants for a dinner with a friend. The consumer ("requestor") asks their friend ("responder") about their preferences. Whereas the requestor likely wishes that the responder conveys a preference for one restaurant over the other, the responder answers that they like both restaurants instead. Interestingly, while prior research has examined requestors' and responders' proclivity to use—and their attitude towards—preference expressions, relatively little is known about the impact of said expressions on requestors' inferences about the responders' liking for the options.

In this article, we examine requestors' inferences about how much responders like a given option as a function of different preference expressions. We compare two common expressions of preference. Specifically, we compare when a responder expresses a (singular) preference, picking one option out of the consideration set, with when a responder expresses indifference, equating all (or some of) the options under consideration by not expressing a preference for a single option. We examine how the use of indifference as opposed to (singular) preference influences requestors' beliefs about how much the responder likes the options under consideration. Our key hypothesis is that requestors infer responders like a given (target) option less when those responders express indifference (e.g., "I like both A and B") versus preference (e.g., "I prefer A"). Importantly, we argue that these inferences are misguided. Responders may in fact like a given option to the same extent when they express indifference as opposed to preference for that option. Thus, our research uncovers a novel source of misperceptions that may arise when consumers make inferences about others (e.g., Barasz and Kim 2022; Barasz, Kim, and Evangelidis 2019; Reit and Critcher 2020).

THEORETICAL FRAMEWORK

Misperceptions About Others' Preferences

Recent literature in marketing and psychology has revealed that consumers systematically err when making inferences about other people's preferences (for an overview see Barasz and Kim 2022). For example, Jung, Moon, and Nelson (2020) reported a series of experiments showing that consumers believe that others value and enjoy the same experiences more than they do (see also Frederick 2012). In their first experiment, Jung et al. observed that participants expected that others would both enjoy and be willing to pay significantly more to watch the movie Dog Days than they themselves would. Relatedly, Reit and Critcher (2020) demonstrated that consumers overestimate how often others will choose common products over rarer, but appealing alternatives. For example, participants in the first study of Reit and Critcher overpredicted how many of their counterparts would choose an orange juice over a passion fruit juice. Further, Barasz, Kim, and Evangelidis (2019) reported that consumers may overinfer the importance of an attribute in others' decisions when chosen alternatives have extreme values on that attribute. For example, in the context of lightbulb purchases, participants in Study 6 of Barasz et al. (2019) overinferred how much other consumers actually cared about eco-friendliness relative to other attributes, such as price, wattage, and lifespan when they were told that those others had chosen an option with a relatively higher eco-friendliness score.

Barasz and Kim (2022) trace many of these mistakes to errors in how consumers perceive others (referred to as "interpersonal errors"). They contend that misperceptions seem to arise because it is inherently difficult for consumers to take the perspective of others (Eyal, Steffel, and Epley 2018; Samuel, Cole, and Eacott 2020). According to Barasz and Kim (2020), difficulties in perspective-taking are particularly pronounced when consumers lack in-depth information about others' preferences. In those situations, consumers often impute missing information by anchoring on their own preferences or those of others who they know.

Importantly, this view of interpersonal errors suggests that misperceptions should be less likely to arise in situations where consumers can solicit information directly from others about their preferences. That is, consumers are expected to make more accurate inferences about others when they are provided with relevant information, such as in the case of dyadic interactions between requestors and responders. However, we hypothesize that similar misperceptions may arise even when consumers (i.e., requestors) solicit information from others (i.e., responders) about their preferences. We argue that these misconceptions can arise because requestors may misinterpret the information communicated by the responders.

Misperceptions About Others' Preferences In Dyadic Interactions

While research on how consumers interpret others' preference expressions in dyadic interactions is scarce, some literature on joint consumption has examined asymmetries in preference expression desires (Kim et al. 2020; Liu and Min 2020). The work by Liu and Min (2020) reveals that requestors desire expressions of (singular) preference from responders (e.g., "let's do X") because said expressions can diminish decision difficulty and assist requestors in reaching a decision. However, responders typically refrain from expressing a preference, and opt to express indifference instead (e.g., "I like both X and Y"). This happens because responders prioritize coming across as easygoing to increase their likeability. Thus, responders expressing indifference believe this will make the decision easier, but requestors learning indifference believe it will make the decision harder, resulting in an asymmetry in perceived decision difficulty between the two parties. This can also lead to requestors believing that responders have a hidden preference (Kim et al. 2020). The aforementioned asymmetry between the two parties can lead to requestors liking responders less and being less inclined to initiate future joint consumption when indifference (vs. preference) is expressed (Kim et al. 2020; Liu and Min 2020).

Building on this research, we theorize that, beyond their negative impact on the relationship between requestors and responders, expressions of indifference can also sway requestors' inferences about the responders' true liking of the options. We conjecture that responders may like a given option to the same extent when they express indifference (e.g., "I like both A and B") as opposed to preference for that option (e.g., "I prefer A"). This is particularly likely to be true when responders invoke indifference to express the fact that they really like more than one option (and that they have no preference between these options). However, we propose that requestors will systematically misinterpret responders' expressions of indifference. Specifically, we hypothesize that expressions of indifference can lead requestors to (mistakenly) infer responders like a given (target) option less compared with when responders express a preference for the same option. We believe that this will be true even when responders clearly signal that they like the alternatives when expressing indifference (e.g., "I like both A and B").

We hypothesize that expressions of indifference (vs. preference) will sway requestors' perceptions about the responders' liking of a given option because indifference can signal that responders are uncertain about their own preferences. Indeed, prior literature shows that consumers may defer choice when they experience subjective feelings of difficulty (Novemsky et al. 2007) or uncertainty about the value of the options (Greenleaf and Lehmann 1995). We hypothesize that, akin to deferral, expressions of indifference may be seen by requestors as a coping mechanism that responders employ to cope with uncertainty about the options' value. Consequently, we propose that, when responders communicate indifference (vs. preference), requestors will infer responders are uncertain about the extent to which they like the options under consideration. In turn, this uncertainty will have a deleterious effect on requestors' inferences about responders' liking of the target, an effect that may persist even when positive cues, such as knowing that the options are generally liked, are present (Fischer, Luce, and Jia, 2000; Luce, Jia, Fischer, 2003). Therefore, we hypothesize that expressions of indifference will lead requestors to infer responders like the target option less compared with expressions of preference for the target.

EMPIRICAL OVERVIEW

Across seven, preregistered studies we provide evidence that expressions of indifference lead requestors to infer responders like the target less compared with expressions of preference for the same option. For the sake of generalizability, across our studies we use a range of scenarios that involve different types of interactions between requestors and responders, and wording of responders' preference expressions. Study 1 provides a conservative test of our main hypothesis. It shows that even a positive expression of indifference (i.e., "I like both A and B") can lead to inferences that the responder likes the target less compared with an expression of preference. Study 2 attempts to replicate these results and test a downstream consequence of preference expression on choice. We replicate our prior results in a new scenario and, interestingly, in a follow-up task, we further find that indifference (vs. preference) increases requestors' proclivity to choose outside alternatives (i.e., options that are not included in the original consideration set). Thus, not only do requestors infer responders like the target less when the latter express indifference as opposed to preference, but they are also willing to forgo the target—and switch to an outside alternative—in an attempt to appease responders.

Study 3 extends the findings of Studies 1 and 2 in two ways. First, it provides evidence for our key result in an individual decision-making task as opposed to joint consumption decisions. Second, it demonstrates that expressions of indifference (vs. preference) can lead to higher incidence of choice deferral when requestors make decisions for themselves.

Study 4 introduces another manipulation of indifference whereby requestors learn that responders are indifferent between two alternatives but also prefer both over a third option (e.g., A = B > C). Even in this case, we replicate our basic result whereby requestors infer responders like the target less compared with when they express preference for a single option (A > B = C).

Studies 5 and 6 provide evidence for our theoretical account. Study 5 demonstrates that uncertainty mediates the effect of preference expression on inferences about responders' liking of the options. Consistent with our framework, we find that, when responders communicate indifference (vs. preference), requestors infer responders are more uncertain about the extent to which they like the options under consideration. In turn, uncertainty dilutes requestors' inferences about the extent to which responders like the target option. Study 6 provides further evidence for our account by experimentally manipulating perceptions of uncertainty.

Finally, Study 7 is a two-stage experiment which provides evidence that requestors' inferences about responders' liking are erroneous. At the first stage of Study 7, we collect preference expression and liking data from participants in the role of responders. During the second stage of Study 7, we ask requestors to infer responders' liking based on those responders' preference expressions. We provide evidence that requestors' inferences are misaligned with responders' actual liking. All code, data, preregistrations, and survey materials are available on ResearchBox (https://researchbox.org/660&PEER_REVIEW_passcode=NQHPCL).

STUDY 1: BASIC EFFECT

Study 1 tests our basic hypothesis that expressing indifference between multiple options, compared to a preference for a single option, leads requestors to infer responders like the target option less.

Method

We recruited 452 participants (47.1% female, Mage = 38.4) via Amazon Mechanical Turk. According to our preregistration, we dropped anyone who took our survey more than once leaving us with 446 participants.

All participants first read a scenario stating that they recently met a new person whom they got along with while hanging out with a group of friends and decided to meet up again. They then read, “You decide to organize the meet up and are debating some different activities you two could do together. Since you don't know what this person likes or dislikes, you suggest two different activities around the same price, going to a pool hall to play billiards or going to a museum.” Following this, participants were then randomly assigned to one of three conditions—preference, indifference, or positive indifference—in which they learned the new person’s preference over the two options. In the preference condition participants read, “After reviewing the activities, they respond that they prefer playing billiards.” In the indifference condition participants read, “After reviewing the activities, they respond that playing billiards and going to museum are the same to them.” In the positive indifference condition participants read, “After reviewing the activities, they respond that they like both playing billiards and going to a museum.”

Participants were then asked to “Please rate how much you think they will like each activity.” They did so for both billiards and a museum on sliding scales from 0, “extremely dislikes,” to 100, “extremely likes,” with 50 labeled as “neutral.” The slider’s starting position was set to 50.

Results and Discussion

As predicted, we found that requestors (i.e. participants) believed that responders liked the target option (“billiards”) significantly less when responders expressed indifference compared with preference (Mpreference = 78.96, SD = 14.90 vs. Mindifference = 64.90, SD = 17.00; $t(291) = 7.53$, $p < .001$). This result persisted even when responders explicitly stated that they liked both options in the positive indifference condition ($M_{preference}$ = 78.96, SD = 14.90 vs. $M_{positive}$ = 71.44, SD = 19.19; t(300) = 3.79, p = .001).

FIGURE 1

RESULTS OF STUDY 1

Note. Study 1 effect of preference expression on inferences of responders' liking for the target option. Bars represent standard errors.

The results from Study 1 reveal that requestors believe that responders like the target option less when responders express indifference compared to when they express preference. This effect persists even when responders clearly signal that they actually like both options when they express indifference. In Study 2 we seek to replicate these findings and explore downstream consequences.

STUDY 2: DOWNSTREAM CONSEQUENCES ON CHOICE

Study 1 provided evidence for our main hypothesis that requestors assume responders like the target option less when the latter express indifference as opposed to preference. In Study 2, we additionally test the consequences of expressing indifference on subsequent choices. Moreover, we use a control, no information, condition to better understand how indifference is perceived.

Method

We recruited 601 participants (39.3% female, Mage = 37.3) via Amazon Mechanical Turk. All participants first read a scenario where they invited one of their friends to go out for dinner and suggested two restaurant options, A and B, to their friend. Following this, participants were then randomly assigned to one of four conditions—baseline (no information), preference, indifference, or positive indifference—in which they learned their friend's expressed preference over the two options. In the no information, control, condition participants read, "After reviewing the restaurants, your friend's phone broke and they had no way to reach you." In the preference condition participants read, "After reviewing the restaurants, your friend responds that they would choose A." Accordingly, the target option is "A." In the indifference condition participants read, "After reviewing the restaurants, your friend responds that A and B are the same to them." In the positive indifference condition participants read, "After reviewing the restaurants, your friend responds that they like both A and B."

Participants were then asked to "Please rate how much your friend would like each restaurant." They did so for both A and B on sliding scales from 0, "extremely dislikes," to 100, "extremely likes," with 50 labeled as "neutral." The slider's starting position was set to 50.

Finally, participants were told to "Imagine there is another restaurant, option C, that you could choose as well but you forgot to tell your friend about it. Now you need to decide at which restaurant to eat." Participants then answered, "To which restaurant would you choose to take your friend to dinner?" The answer choices were A, B, or C. Lastly, participants filled out demographic information.

Results and Discussion

Replicating the findings of Study 1, we found that requestors (i.e. participants) inferred the responder liked the target option A significantly less when the responder expressed indifference as opposed to preference (Mpreference = 83.91, SD = 12.81 vs. Mindifference = 66.20, SD = 16.75; t(300) = 10.32, p < .001). Again, this result persisted even when requestors were told that the responder liked both options in the positive indifference condition (Mpreference = 83.91, SD = 12.81 vs. Mpositive = 72.62, SD = 17.88; t(301) = 6.32, p < .001).We also find that requestors inferred the responder liked the target option A significantly less when the responder could not reply in the no information condition as opposed to the preference condition (Mpreference = 83.91, SD = 12.81 vs. Mno_information = 67.26, SD = 17.26; t(296) = 9.48, p < .001).

FIGURE 2

RESULTS OF STUDY 2 (LIKING)

Note. Study 2 effect of preference expression on inferences of responders' liking for the target option. Bars represent standard errors.

Subsequently, requestors were more likely to choose the outside option, C, when learning indifference compared to preference (40.4% vs. 4%, □2(1) = 58.02, p < .001; see Figure 2). Strikingly, this result persisted even when requestors learned both options were liked in the positive indifference condition (22.4% vs. 4%, □2(1) = 22.37, p < .001). The latter result is striking since a substantial proportion of participants elected to forgo an alternative that they knew responders liked in favor of an outside option for which they lacked information about responders' preferences. Requestors were also more likely to choose the outside option in the no information condition than in the preference condition (32% vs. 4%, $\chi^2(1) = 11.44$, $p = .001$).

FIGURE 3

RESULTS OF STUDY 2 (CHOICE)

Note. Study 2 downstream consequence of inferences from expressed preference on choice of an outside option.

Studies 1 and 2 demonstrate that requestors believe responders like a target option less when responders express indifference as opposed to preference. This effect persists even when the options are known to be liked (positive indifference). Ultimately, more requestors in the indifference conditions (vs. preference) chose an outside option in a subsequent choice. Thus, not only do requestors infer responders like the target less when responders express indifference as opposed to preference, but they are also willing to forgo the target—and switch to an outside alternative—in an attempt to appease responders. Interestingly, requestors elect to forgo the target option even when responders explicitly state that they like all options.

STUDY 3: INDIVIDUAL CONSUMPTION

Study 3 extends previous findings in three ways. First, while Studies 1 and 2 tested our hypothesis in joint consumption scenarios, Study 3 tests our hypothesis in an individual consumption domain wherein only the requestor will be consuming the options under consideration. Second, we test whether preference expressions bear on another consequence: choice deferral. Third, Study 3 seeks to further expand the generalizability of previous results by using a new scenario, new operationalizations of preference expression, as well as a different sample.

Method

We recruited 300 participants (48.7% female, Mage = 18.8) via a European university's lab pool. All participants first read a scenario asking them to imagine they are planning a trip to Greece and have time in their schedule to visit one of the Greek islands. Luckily, one of their friends is Greek, so they ask them which of two islands, A or B, they should visit. Following this, participants were then randomly assigned to one of two conditions, preference or indifference. In the preference condition participants read, "Your friend replies that they would go to island A." In the indifference condition participants read, "Your friend replies that they would go to either island, A or B."

Participants then decided if they would choose to go to island A, island B, or "keep reading about other islands." Next, they were asked to "Please rate how much you think your friend likes each island." They did so for both islands on sliding scales from 0, "extremely dislikes," to 100, "extremely likes," with 50 labeled as "neutral." The slider's starting position was set to 50.

Results and Discussion

Replicating our previous findings, we found that requestors (i.e. participants) inferred that responders liked the target option, island A, significantly less when responders expressed indifference as opposed to preference ($M_{preference} = 82.19$, $SD = 11.04$ vs. $M_{indifference} = 65.35$, $SD = 17.46$; $t(298) = 9.99$, $p < .001$).

Further, requestors were more likely to defer their choice, instead opting to review other potential islands, when others expressed indifference as opposed to preference (64% vs. 45%, $\chi^2(1) = 11.38$, $p = .001$).

FIGURE 4

RESULTS OF STUDY 3 (CHOICE)

Note. Study 3 choice shares of each island and deferral.

Study 3 generalized our previous result to an individual consumption setting besides joint consumption settings. Importantly, it showed that preference expressions have important downstream consequences for one's own decision-making. Specifically, data of Study 3 show that indifference (vs. preference) expressions boost requestors' proclivity to defer choice. Taken together, the results of Studies 1, 2, and 3 show that requestors infer responders like a given option less when the latter express indifference between multiple options as opposed to preference for a single option. In turn, upon receiving information that responders are indifferent between multiple options, requestors are more likely to search for (Study 3)—and select (Study 2)—outside options. This result persists even when responders explicitly state that they actually like the alternatives under consideration, such that choice of existing options may be satisfactory.

STUDY 4: INDIFFERENCE AS A DUAL PREFERENCE

All the studies thus far have tested our basic hypotheses using an operationalization of indifference which communicated responders were indifferent between all options under consideration. The purpose of this study is to test another, common expression of indifference, whereby a responder states that they are indifferent between some of the options, while they prefer those options over alternative(s). That is, responders may express preference for a subset of the options under consideration. For instance, a responder may state that, while they prefer both A and B over C, they do not have a preference between two options, A and B. We will refer to this type of indifference as dual preference. Importantly, to create a dual preference, we move from consideration sets containing two options, A and B, to sets containing three options, A, B, and C. The introduction of a third option now allows us to compare three expressions: $A > B = C$ (preference), $A = B > C$ (dual preference), and $A = B = C$ (indifference). Our prediction here is that requestors will infer responders like the target option (A) less when those responders express either a dual preference or indifference compared with when they express preference only for option A.

Method

We recruited 451 participants (55.9% female, M_{age} = 38.2) via Amazon Mechanical Turk. According to our preregistration, we dropped anyone who took our survey more than once leaving us with 445 participants. All participants first read the same scenario as in Study 2. Following this, participants were then randomly assigned to one of three conditions—preference, indifference, or dual preference—in which they learned the new person's preference over the three options. In the preference condition participants read, "After reviewing the restaurants, your friend responds that they prefer A." In the dual preference condition participants read, "After reviewing the restaurants, your friend responds that they prefer either A or B." In the indifference condition participants read, "After reviewing the restaurants, your friend responds that they prefer either A, B, or C." Unlike previous studies, in this study every condition contains the word, "prefer".

Participants were then asked to "Please rate how much you think your friend will like each restaurant." They did so for restaurant options A, B, and C on sliding scales from 0, "extremely dislikes," to 100, "extremely likes," with 50 labeled as "neutral." The slider's starting position was set to 50. Lastly, participants filled out demographic information.

Results and Discussion

Replicating our basic finding, requestors (i.e. participants) inferred that responders liked the target option A less when responders expressed indifference as opposed to preference ($M_{preference} = 85.63$, $SD = 13.96$ vs. $M_{indifference} = 70.21$, $SD = 15.69$; $t(295) = 8.94$, $p < .001$). Importantly, this result persisted, albeit slightly weakened, when responders expressed indifference between only two of the options, as in the dual preference condition ($M_{preference} = 85.63$, $SD = 13.96$ vs. $M_{dual\_preference} = 78.20$, $SD = 13.96$; $t(293) = 4.57$, $p < .001$). Finally, we also found that inferences about A's liking differed between the dual preference condition and the indifference condition ($M_{dual\_preference} = 78.20$, $SD = 13.96$ vs. $M_{indifference} = 70.21$, $SD = 15.69$; $t(296) = 4.65$, $p < .001$).

FIGURE 5

RESULTS OF STUDY 4

Note. Study 4 effect of preference expression on inferences of responders' liking for the three options. Bars represent standard errors.

The results of Study 4 extend our findings to a larger consideration set and importantly show that requestors learning indifference, compared to preference, infer responders like the target less even when indifference involves preferring a subset of the options. One potential explanation for our results may involve the number of forgone options. In this, and previous studies, the number of forgone options increases when moving from indifference to preference. For instance, in Study 4, there are zero forgone options in the indifference condition, one in the dual preference condition, and two in the preference condition. We wondered whether the number of forgone options acts as a costly signal causing requestors to infer responders like the target more the more options are forgone. To test this account, we ran another study with the same design as Study 4, but additionally manipulated whether the consideration set was three or five options. When the size of the consideration set increases, so do the number of forgone options, and according to this account, so should the inferred liking of the target option. We found no consistent evidence for this account (see Web Appendix B).

STUDY 5: THE MEDIATING ROLE OF UNCERTAINTY

Across a variety of contexts and consumption modes, we have shown that requestors infer responders like the target option less when the latter express indifference compared with preference. In our theory section we proposed that this would occur because of requestors' beliefs about the extent to which responders experience preference uncertainty. Specifically, we conjectured that requestors would believe that responders are more uncertain about the value of the options when they express indifference compared with when they express preference. In this study, we test our theoretical account directly by measuring whether responders expressing indifference are judged to experience more uncertainty than those expressing preference. In turn, feelings of uncertainty may mediate the impact of preference expression on inferences about the extent to which responders like the target option. Further, for generalizability, Study 5 features a novel consumption scenario, shopping with friends.

Method

We recruited 452 participants (71.2% female, Mage = 35.7) via Prolific. All participants first read that they are out shopping with a friend and found three jackets they are interested in and are debating if they should buy one. So, they ask for their friend's opinion. Following this, participants were then randomly assigned to one of three conditions—preference, indifference, or dual preference—in which they learned the new person's preference over the three options.

In all conditions they read, "Your friend asks you to try on each jacket." In the indifference condition they then learn "Your friend then says they like the first jacket, the second jacket, and the third jacket." In the dual preference condition participants read, "Your friend then says they like both the first jacket and the second jacket." In the preference condition participants read, "Your friend says they like the first jacket."

Afterwards, participants answered our key dependent measures. As in previous studies, participants were asked to "please rate how much you think your friend likes each jacket." They did so for the first, second and third jacket on 101-point sliding scales (0 = extremely dislikes, 50 = neutral, 100 = extremely likes; the slider's starting position was set to 50). We measured uncertainty by asking participants, "How uncertain do you think your friend is about their preference?" (1 = very uncertain, 7 = very certain). We counterbalanced the order of our dependent measures.

Results and Discussion

Replicating previous findings, we observed that requestors (i.e. participants) inferred that responders liked the target option A less when responders expressed indifference than when they expressed preference ($M_{preference} = 83.39$, $SD = 14.34$ vs. $M_{indifference} = 64.91$, $SD = 15.26$; $t(297) = 10.79$, $p < .001$). Further, replicating Study 4, we again found that this result persisted when responders expressed indifference between two preferred options ($M_{preference} = 83.39$, $SD = 14.34$ vs. $M_{dual\_preference} = 75.73$, $SD = 12.92$; $t(299) = 4.87$, $p < .001$). Finally, we also found that inferences about A's liking also differed between the dual preference condition and the indifference condition ($M_{dual\_preference} = 75.73$, $SD = 12.92$ vs. $M_{indifference} = 64.91$, $SD = 15.26$; $t(302) = 6.68$, $p < .001$).

Importantly, turning to the measure of uncertainty, we found requestors believed responders experienced more preference uncertainty when responders expressed indifference as opposed to preference ($M_{preference} = 5.32$, $SD = 1.17$ vs. $M_{indifference} = 2.98$, $SD = 1.59$; $t(297) = 14.51$, $p < .001$). This result persisted, yet, similar to the effect on inferences about liking above, was relatively smaller when participants expressed indifference between a subset of the options ($M_{preference}$ = 5.32, SD = 1.17 vs. $M_{dual\_preference}$ = 4.73, SD = 1.25; t(299) = 4.24, p < .001). Finally, inferences about preference uncertainty also differed significantly between the dual preference and indifference conditions ($M_{dual\_preference}$ = 4.73, SD = 1.25 vs. $M_{indifference}$ = 2.98, SD = 1.59; t(302) = 10.69, p < .001).

We ran a simple mediation model with uncertainty mediating the effect of preference expression on inferences about responders' liking of the target using 10,000 bootstrap samples. We report bias corrected 95% confidence intervals for the indirect effects. Compared to preference, the indirect effect of expressing indifference through uncertainty on the inferred liking of the target was significantly different from zero ($ß_{indifference}$ = -9.74; SE = 1.46, 95% LLCI = -12.79, ULCI = -7.02) and so was the indirect effect of expressing a dual preference ($ß_{dual\_preference}$ = -2.46; SE = 0.66, 95% LLCI = -3.90, ULCI = -1.28). Furthermore, compared to expressing indifference, the indirect effect of expressing dual preference through uncertainty on inferred liking of the target was significant ($ß_{dual\_preference}$ = 7.28; SE = 1.20, 95% LLCI = 5.13, ULCI = 9.87).

In conclusion, the results of Study 5 corroborate our theorical account. Consistent with our claims, requestors believe that responders are more uncertain about their preferences when they express indifference compared with when they express preference. In turn, perceived uncertainty mediates the impact of preference expression on responders' beliefs about the extent to which responders like the target option. In our next study, we further test our theoretical account by experimentally manipulating uncertainty.

STUDY 6: MANIPULATING UNCERTAINTY

We previously found that requestors infer responders like a target option less when they express indifference versus preference because they associate indifference with preference uncertainty. Building on this finding, we examined whether experimentally manipulating uncertainty would moderate our key result. Specifically, we examined whether our results would be attenuated when we told participants that responders were uncertain about the extent to which they liked the option(s). We surmised that our key result would be weakened under the additional uncertainty manipulation because the latter would induce beliefs that responders experience uncertainty in both preference and indifference conditions. That is, introducing uncertainty should lower perceived liking of the target more so for the preference condition than for the indifference condition, leading to an attenuation of our result.

Method

We recruited 803 participants (39.9% female, Mage = 26.5) via Prolific. All participants first read that they are out shopping with a friend and found three jackets they are interested in and are debating if they should buy one. So, they ask for their friend's opinion. Following this, participants were then randomly assigned to one of four conditions in a 2 (Expression: Dual preference vs. Preference) x 2 (Uncertainty: Control vs. Uncertainty) between-subjects design.

All participants first read that "Your friend asks you to try on each jacket." The preference condition read, "Your friend says they like the first jacket." The dual preference condition read, "Your friend says they like both the first and second jacket." Participants in the control conditions received no additional information. In contrast, those in the uncertainty preference [dual preference] condition additionally read: "However your friend is unsure if they like [both] the first jacket [and the second jacket] just a little or if they really love it [them]."

Participants were then asked to “Please rate how much you think your friend likes each jacket.” They did so for the first, second, and third jacket on sliding scales from 0, “extremely dislikes,” to 100, “extremely likes,” with 50 labeled as “neutral.” The slider’s starting position was set to 50.

Results and Discussion

As predicted, we found a significant interaction between our two experimental factors on the inferred liking of the target option (i.e., the first jacket; ßinteraction = 7.67; t(799) = 3.75, p < .001). Replicating our previous finding, in the control conditions, requestors (i.e., participants) inferred responders liked the target less when responders expressed a dual preference compared with when they expressed preference for a single option (Mpreference = 84.9, SD = 14.57 vs. Mdual_preference = 75.72, SD = 14.46; t(400) = 6.34, p < .001). However, this result was eliminated in the uncertainty conditions (Mpreference = 72.34, SD = 12.78 vs. Mdual_preference = 70.83, SD = 14.46; t(399) = 1.04, p = .298).

FIGURE 6

RESULTS OF STUDY 6

Note. Study 5 effect of uncertainty on inferences of responders’ liking for the target option (first jacket). Bars represent standard errors.

Taken together, data of Studies 5 and 6 provide robust support for our theoretical account. Study 5 showed that requestors infer responders like a target option less when they express indifference versus preference because they associate indifference with preference uncertainty. Building on this insight, Study 6 showed that our basic finding is eliminated when we experimentally manipulate preference uncertainty, such that responders in both preference and indifference conditions convey that they are uncertain about how much they like the target. When both preference and indifference are coupled with preference uncertainty, we no longer observe substantial differences in requestors' inferences about the extent to which responders like the target. Consequently, beliefs about preference uncertainty play a key role in the association between preference expression and perceptions about how much others like a given (target) option.

STUDY 7: IS OUR BASIC RESULT A FORM OF MISPERCEPTION?

So far, we have consistently shown that requestors infer responders like a target option less when those responders' express indifference as opposed to preference. In our Introduction and Theoretical Framework sections, we argued that these inferences are misguided because responders may in fact like a given option to the same extent when they express indifference as opposed to preference for that option. Thus, we hypothesized that requestors' inferences do not accurately reflect responders' true preferences. Study 7 attempts to empirically demonstrate our propositions using actual data on responders' liking that we collected in a two-stage design. In the first stage, participants take on the role of the responder and visit the Yelp pages of two restaurants their friend has proposed to them. After reviewing the two restaurants, responders select the expression that best matches their preferences and then rate their liking of each restaurant. In the second stage, different participants take on the role of the requestor and are asked to infer responders' liking of the options for a given preference expression. Further, to make responses consequential, participants in stage 2 were incentivized to predict correctly. Importantly, we should note that, in this study, requestors were provided with concrete information about the choice options, such that they could see that alternatives' average rating (and additional information) on Yelp. Seeing the average Yelp ratings of the options should serve as a powerful piece of information that can inform requestors about how much responders would like each option, narrowing down the set of plausible inferences. Thus, Study 7 provides a rather conservative test of our hypotheses.

Method

All participants in stage 1 first read a scenario asking them to imagine they are on a post-COVID trip with a friend in Memphis, Tennessee which is known for BBQ and Blues music. They decide that they should try some BBQ and their friend goes on Yelp and proposes two BBQ restaurants to them. On the next screen, they were provided with the links to two BBQ restaurants, Corky's BBQ and Central BBQ, and asked to spend a couple of minutes checking out each link. They were also told, "after this there will only be a few questions to answer." This was done in the hope that participants would spend more time reviewing the restaurants. Furthermore, on the same screen we asked them what the banner image was on each restaurant's Yelp page. Again, this was done to further nudge participants to visit the links and meaningfully review them. We did not use this as a means of exclusion. Participants were then asked to "Please choose the option that best states what you would tell your friend." The options (randomly ordered) were:

1. I would prefer Central BBQ

2. I would prefer Corky's BBQ

3. I would like both Central and Corky's BBQ

4. Central and Corky's BBQ are the same to me

5. I don't like either option

Next, they were asked to "Please rate how much you think you would like each BBQ restaurant." They did so for both Corky's BBQ and Central BBQ on 101-point sliding scales on 101-point sliding scales (0 = extremely dislikes, 50 = neutral, 100 = extremely likes; the slider's starting position was set to 50).

In stage 2, participants assumed the role of requestors tasked with predicting the liking of responders from stage 1. All participants saw the exact task and number of participants who completed stage 1 and were told that they will see the choices of some of these participants and predict how much those participants like the options. They also learned that "the 5 closest predictions to the average of prior participants' ratings would receive a 3£ bonus" (base payment was 0.75£). After reading the same scenario and visiting the same Yelp pages as participants from stage 1, they saw the question and choices that participants from stage 1 could choose. Participants were then randomly assigned to focus on only one group of stage 1 participants. The focused on either those who preferred Corky's BBQ (preference), preferred Central BBQ (preference), liked both of them (positive indifference), or viewed them as the same (indifference). They were then asked to rate how those in their assigned condition, on average, rated each restaurant using the same scales as the responders (stage 1 participants). They were informed that the scales were the same as the ones used by prior participants and again reminded of the potential bonus.

Our key prediction was that the interaction between participants' role and preference expression would be significant such that requestors would overestimate responders' liking for the target option when responders expressed a preference for as opposed to positive indifference. That is, requestors would overestimate responders' true difference in liking for the target option when those responders' expressed preference versus positive indifference.

A couple of notes are worth discussing. Due to the nature of the first stage where participants selected their preference expression, we could not control how many participants would select into each condition (choice of preference expression) that is used in the second stage. This led us to be woefully underpowered to detect an interaction the first time we ran this study if we were to remain faithful to our preregistrations. Realizing this only afterward, we decided to run a second wave of both stage 1 and stage 2 and combine the two waves. In our analysis, we deal with this choice to run a second wave in two ways, which we preregistered before running the second wave. First, we include wave fixed effects in our regression to control for any differences in the two waves. We also verified that the Yelp pages of each restaurant during each wave of stage 1 and stage 2 were the same by visiting the pages and reviewing the information displayed. Second, we lowered our alpha level to account for the choice to collect additional data. A formal correction of the alpha level would lead to a lowering from .05 to .03 (Pocock 1977). Preregistrations were done for both waves and are available in our Researchbox. After (pre-registered) exclusions, combining waves yielded 2,411 participants (71.9% female, Mage = 32.4) recruited via Prolific.

Results and Discussion

In both stages of our study, we did not force participants to provide ratings that were logically consistent with the preference expression category that they identified with (stage 1) or that they were assigned to (stage 2). Due to this, we observed several inconsistencies in our data. Specifically, 201 participants out of 1,174 in the preference condition either reported that the target option was liked less or the same as the alternative. We deal with these violations in two ways. First, in our analyses below we drop these participants. Second, instead of dropping these participants data, we run a thought experiment asking: what would have happened if these participants were in the conditions that are consistent with their reported liking of the options? To answer this, we reassigned these participants to the condition that is consistent with their reported liking of the options and ran the analyses (see appendix for details). The results are similar across these two methods. Neither of these methods were preregistered as we simply did not think about the occurrence of such violations prior to data collection.

As preregistered, we collapsed the two preference conditions, prefer Corky's and prefer Central, into one preference condition to increase power and because we had no a priori reason to expect differences. We create the liking of the target option by taking the maximum liking rating between Corky and Central for each participant.

Using OLS with robust standard errors, we regressed the liking of the target option on role (0 for responder, 1 for requestor), expression (0 for preference, 1 for positive indifference), their interaction, and a dummy for wave ($N = 1,623$). We found a significant interaction between role and expression ($ß_{interaction} = -2.48$; $t(1,618) = -1.86$, $p = .026$; see Figure 6). Responders (i.e., stage 1 participants) expressing positive indifference compared to those expressing a preference liked the target to a similar extent ($M_{positive\_indifference} = 80.57$, $SD = 10.32$ vs. $M_{preference} = 81.92$, $SD = 11.51$; $t(731) = 1.63$, $p = .104$). However, replicating previous findings, requestors (i.e. stage 2 participants) wrongly inferred that responders liked the target less when responders expressed positive indifference as opposed to preference ($M_{positive\_indifference}$ = 78.33, SD = 11.62 vs. $M_{preference}$ = 82.17, SD = 10.09; $t(888) = 5.21$, $p < .001$). Interestingly, this misprediction comes solely from requestors underestimating responders' liking of the target in the positive indifference conditions ($M_{responders}$ = 80.57, SD = 10.32 vs. $M_{requestors}$ = 78.33, SD = 11.62; $t(648) = 2.57$, $p = .01$) and not from requestors overestimating responders' liking of the target in the preference conditions ($M_{responders}$ = 81.92, SD = 11.51 vs. $M_{requestors}$ = 82.17, SD = 11.22; $t(971) = .35$, $p = .724$).

FIGURE 7

RESULTS OF STUDY 7

Note. Study 7 effect of preference expression on inferences of responders' liking for the target option. Bars represent standard errors.

Though not preregistered, we also ran the same regression but this time comparing preference to indifference. Interestingly, the interaction between role and expression was not significant ($ß_{interaction}$ = 2.49; $t(1,555) = 1.61$, $p = .108$). Taken together, our results show that requestors accurately assess liking of the target option when indifference is expressed. However, when an additional positive cue is present, such as when responders explicitly state that they like both options (i.e., positive indifference condition), requestors end up underestimating how much responders like the target.

In conclusion, using an incentivized task that resembles what a dyad would face in the real-world, Study 7 provides further evidence that requestors infer responders like a target option less when the latter express indifference as opposed to preference. Importantly, our data further show that said inferences can be misguided when responders explicitly state that they like the alternatives when expressing indifference. Our data suggest that our effects occur because requestors underestimate responders' liking of the target when responders' express indifference as opposed to preference (and not because requestors overestimate responders' liking of the target when responders express preference). Thus, data of Study 7 suggest that expressions of indifference can lead to misperceptions about the extent to which others like a given (target) option.

GENERAL DISCUSSION

Consumers are constantly requesting the opinions of others before reaching a decision. In this paper, we examined how the inferences requestors draw about responders' liking of a target option vary as a function of preference expression. Data from seven preregistered experiments corroborate that learning expressions of indifference (vs. preference) lead requestors to infer responders like the target option less. This occurs in both joint (Studies 1, 2, and 4) and individual consumption (Studies 3, 5, 6, and 7) settings. Importantly, this empirical regularity is predicated on requestors' beliefs about responders' feelings of preference uncertainty. Our data show that, when responders communicate indifference (vs. preference), requestors infer responders are more uncertain about the extent to which they like the options under consideration. Consequently, uncertainty reduces requestors' inferences about the extent to which responders like the target option (Study 5). Study 6 provides further evidence for our framework by experimentally manipulating uncertainty. Importantly, Study 7 demonstrates that requestors' inferences about responders' liking of the target option can be inaccurate, a finding which suggests that preference expressions may produce noise in interpersonal communications between consumers.

Theoretical Contributions

Our research contributes to three distinct research streams within social sciences. First, our research contributes to prior literature in marketing and psychology which reports that consumers often err when making inferences about others' preferences (e.g., Barasz et al. 2019; Jung et al. 2020; Reit and Critcher 2020). As we previously argued, Barasz and Kim (2022) trace many of these empirical findings to errors in how consumers perceive others, presumably because it is particularly hard for consumers to take other consumers' perspective (Eyal et al. 2018; Samuel et al. 2020). Importantly, prior research postulates that such misperceptions should be less common when consumers can solicit information directly from others about their preferences. In other words, consumers should be able to make accurate inferences about others' preferences when they possess relevant information about said preferences. Our contribution to this research is twofold. First, we demonstrated a novel type of error that arises when responders express indifference as opposed to preference in response to requestors. Second, our research suggests that interpersonal misperceptions may occur even when consumers have information about others' preferences. Thus, our data suggest that interpersonal misperceptions may be more common than previously thought, potentially introducing noise to a wide range of communications.

Second, our research contributes to prior consumer literature on dyadic interactions on joint consumption decisions. For instance, Liu and Min (2020) showed that requestors typically want responders to express preference instead of indifference because preference expressions can facilitate decision-making. However, contrary to requestors' desires, responders typically express indifference instead of preference because they prioritize coming across as easygoing and likeable. This tendency may further cause requestors to believe that responders have a hidden preference, which, in turn, can lead to reduced desire to engage in joint consumption in the future (Kim et al. 2020; Liu and Min 2020). Our work contributes to that literature in three main ways. First, our data show that, beyond their negative impact on the relationship between requestors and responders, expressions of indifference can influence requestors' inferences about responders' preferences. Second, to our knowledge, our study is the first to empirically demonstrate that requestors' inferences about responders' preferences may, in fact, be inaccurate (Study 7). Third, our research question allows us to study preference expressions not only in the context of joint consumption but also in the case of individual consumption (cf. Kim et al. 2020; Liu and Min 2020). By doing so, our research suggests that preference expressions can exert a powerful impact across a wider range of settings and decision-making domains than previously thought.

Third, our work contributes to prior decision-making literature on preference uncertainty. According to this literature, consumers tend to experience uncertainty about their preferences when choosing between options that present trade-offs, such as when an option scores higher on one attribute but lower on another compared to the alternative(s) (Dhar 1997; Novemsky et al. 2007). Similarly, consumers may experience preference uncertainty when faced with options that have both pros and cons (Fischer, Luce, Jia 2000; Luce, Jia, Fischer 2003). Our research extends this literature by examining how consumers make inferences about the extent to which others experience uncertainty based on their stated preferences. Our data suggest that consumers are more likely to infer that others experience uncertainty when the latter express indifference (vs. preference), even when those others explicitly state that they actually like the options under consideration. Thus, our research shows that preference uncertainty is not restricted to those that experience it. Rather, preference uncertainty can be the object of inference-making processes when speculating about the preferences of others.

Directions For Future Research

Our data can inspire several directions for future research. First, while our work varied the relationship of the responder with the requestor across scenarios, we did not formally study how this relationship might impact our key results. Prior research has shown that relationship types affect various consumer decisions (Gorlin and Dhar, 2012; Hamilton et al., 2020; Wight et al., 2022). For example, while we reasonably assume friends and partners have our best interests at heart, it might be more difficult to trust marketers who we think are just trying to sell us something. In this case, a consumer may believe that a marketer stating a preference (vs. indifference) might have ulterior motives. In turn, the consumer may be less likely to follow on the marketer’s recommendation because they may feel that a given (target) option is being pushed on them. Thus, the type of the relationship between the requestor and the responder might moderate the effects presented in our studies.

Second, in our work, we only examined scenarios where one responder states their preference expression. However, many decisions are made with the input of many responders. For example, when shopping for a wedding dress, the bride to be often takes an entourage of family and friends with her. Our results show that stating a preference is associated with more certainty and this certainty, being valued by decision makers, would likely cause expressions of preference to be weighted more heavily than expressions of indifference. If preference is weighted more heavily when aggregating responders' opinions, then it might also cause more difficulty reaching a conclusion when responders' preferences oppose. This makes for an interesting prediction: a group of responders where some responders are divided on their preferred option may produce more decision uncertainty than a group of responders who are all indifferent among the options. If this is true, it suggests that the individual level effects presented in our paper may be eliminated—or reversed—when requestors need to aggregate multiple responders' preference expressions.

Third, even though our studies contained participants from different cultures across numerous countries, we should note that most of our participants were from individualistic rather than collectivistic cultures. This point is important to consider because expressions of indifference may be more common in collectivistic cultures where people focus more on the desires of others and may have a stronger desire to come across as easygoing. As such, in collectivistic cultures, requestors may be more prone to (correctly) infer that responders do not like an option less when the latter express indifference as opposed to preference. Future research could explore how collectivistic versus individualistic cultures potentially differ in their usage and views of preference expressions.

Practical Implications

Our research demonstrates an important tradeoff that responders may face when they like two (or more) options equally, such that they consider expressing indifference. On the one hand, expressing indifference may seem like a good idea because it allows responders to accurately convey their true preferences (i.e., the fact that they like two options equally). On the other hand, expressing indifference can backfire because indifference may prompt the requestors to forgo extant alternatives and consider other, outside options that responders might enjoy less. According to our data, this may occur even when expressions of indifference clearly convey that the requestor likes the focal alternatives (e.g., "I like both A and B"). Thus, expressing indifference is a strategic decision not only for appearing amenable but also for its potential impact on the option that is eventually chosen. For requestors, our results suggest that when they learn expressions of indifference, they should better incorporate additional cues, such as the presence of positive indicators (e.g., "like"), which signal that indifference need not imply that their counterparts like the focal alternative(s) less.

For marketers our findings suggest that salespersons, whose opinions the consumer trusts, need to be aware that they too are responders and expressing a preference versus indifference to a customer can have varying effects on them. For example, a salesperson advising a consumer to buy a given product may be more likely to produce a sale compared to a salesperson that advises a consumer that multiple products are equally good. This is because salespersons communicating a preference for a single product may appear to be more certain about their liking of the products, which, in turn, may prompt consumers to make a purchase. As mentioned above, these effects might diminish for salespersons whose opinions are not trusted by the consumer.

Conclusion

Requesting the opinions of others is an important and often overlooked part of consumer decision-making. Friends, dates, partners, salespersons, and so on, shape our beliefs about the options that we are considering based on how they express their preferences. To this end, in this paper we investigated how different forms of preference expression can drive requestors' inferences about other people's preferences. By doing so, we hope that our investigation will inspire future research with the aim of uncovering how consumers can better communicate their own—but also interpret other consumers'—preferences.

References

Barasz, Kate, Tami Kim, and Ioannis Evangelidis (2019), "I Know Why You Voted For Trump: (Over) Inferring Motives Based on Choice," Cognition, 188, 85-97.

Barasz, Kate, and Tami Kim (2022), "Choice Perception: Making Sense (and Nonsense) of Others' Decisions," Current Opinion in Psychology, 43, 176-81.

Dhar, Ravi (1997), "Consumer Preference for a No-Choice Option," Journal of Consumer Research, 24 (2), 215-31.

Eyal, Tal, Mary Steffel, and Nicholas Epley (2018), "Perspective Mistaking: Accurately Understanding the Mind of Another Requires Getting Perspective, Not Taking Perspective." Journal of Personality and Social Psychology, 114 (4), 547-71.

Fischer, Gregory W., Mary Frances Luce, and Jianmin Jia (2000), "Attribute Conflict and Preference Uncertainty: Effects on Judgment Time and Error," Management Science, 46 (1), 88-103.

Frederick, Shane (2012), "Overestimating Others' Willingness to Pay," Journal of Consumer Research, 39 (1), 1-21.

Gorlin, Margarita and Ravi Dhar (2012), "Bridging the Gap between Joint and Individual Decisions: Deconstructing Preferences in Relationships," Journal of Consumer Psychology, 22 (3), 320-23.

Greenleaf, Eric A., and Donald R. Lehmann (1995), “Reasons for Substantial Delay in Consumer Decision Making,” Journal of Consumer Research, 22 (2), 186-99.

Hamilton, Ryan, Rosellina Ferraro, Kelly L. Haws, and Anirban Mukhopadhyay (2021), “Traveling with Companions: The Social Customer Journey,” Journal of Marketing, 85 (1), 68-92.

Jung, Minah H., Alice Moon, and Leif D. Nelson (2020), “Overestimating the Valuations and Preferences of Others,” Journal of Experimental Psychology: General, 149 (6), 1214.

Kim, Nicole Y., Alixandra Barasch, Yonat Zwebner, and Rom Y. Schrift (2020), “You Must Have a Preference: The Impact of No Preference Communication on Joint Decision Making,” Available at SSRN 3650352, accessed on 09/18/2020.

Liu, Peggy J., and Kate E. Min (2020), “Where Do You Want To Go For Dinner? A Preference Expression Asymmetry in Joint Consumption,” Journal of Marketing Research, 57 (6), 1037-54.

Luce, Mary Frances (1998), “Choosing to Avoid: Coping with Negatively Emotion-Laden Consumer Decisions,” Journal of Consumer Research, 24 (4), 409-33.

Luce, Mary Frances, Jianmin Jia, and Gregory W. Fischer (2003), “How Much Do You Like It? Within-Alternative Conflict and Subjective Confidence in Consumer Judgments,” Journal of Consumer Research, 30 (3), 464-72.

Novemsky, Nathan, Ravi Dhar, Norbert Schwarz, and Itamar Simonson (2007), “Preference Fluency in Choice,” Journal of Marketing Research, 44 (3), 347-56.

Pocock, Stuart J. (1977), "Group Sequential Methods in the Design and Analysis of Clinical Trials," Biometrika, 64 (2), 191-9.

Reit, Emily S., and Clayton R. Critcher (2020), "The Commonness Fallacy: Commonly Chosen Options Have Less Choice Appeal Than People Think," Journal of Personality and Social Psychology, 118 (1), 1-21.

Samuel, Steven, Geoff G. Cole, and Madeline J. Eacott (2020), "Two Independent Sources of Difficulty in Perspective-Taking/Theory of Mind Task," Psychonomic Bulletin & Review, 27 (6), 1341-47.

Wight, Kelley Gullo, Peggy J. Liu, James R. Bettman, and Gavan J. Fitzsimons (2022), Social Relationships and Consumer Behavior.

WEB APPENDIX FOR

"Consumers (Wrongly) Believe that Others Like an Option Less When Those Others Express Indifference"

WEB APPENDIX A: ADDITIONAL RESULTS IN STUDIES 1-7

STUDY 1: BASIC EFFECT

We also found that requestors (i.e., participants) inferred the responder liked the non-preferred option, the museum, significantly more when the responder expressed indifference as opposed to preference ($M_{preference}$ = 47.13, SD = 23.02 vs. $M_{indifference}$ = 62.48, SD = 19.98;

t(291) = 6.09, p < .001). This result persisted when comparing the preference condition to the positive indifference condition (Mpreference = 47.13, SD = 23.02 vs. Mpositive = 71.11, SD = 18.26; t(300) = 10.04, p < .001).

After answering rating each option, participants were told, “After hearing back, you learn that the only billiards place in town is closed for renovations. You remember there is also a comedy club in town. You need to buy tickets now for either the museum or the comedy club. There are only 5 tickets left for the comedy club and for the museum. Both activities cost the same amount. You don't have time to ask about their preferences as you need to buy the tickets now before they sell out.” Participants then answered, “Which activity would you buy tickets for?” The answer choices were comedy club or museum. This measure provides an additional test of our theoretical account, whereby we could examine whether the new option (i.e., comedy club) is chosen more frequently when responders expressed preference (vs. indifference).

We found that requestors were less likely to choose the new, outside option, the comedy club, when learning indifference compared to preference (34% vs. 77%, □2(1) = 55.34, p < .001; see Figure 4). This result was similar when requestors learned that responders liked both options as stated in the positive indifference condition (29% vs. 77%, □2(1) = 69.15, p < .001). We also observed that the two indifference conditions did not differ in the percent of requestors choosing the outside option (34% vs. 29%, □2(1) = 0.73, p = .393).

STUDY 2: DOWNSTREAM CONSEQUENCES ON CHOICE

We also found that requestors (i.e. participants) inferred the responder liked the non-preferred option B significantly more when the responder expressed indifference as opposed to preference (Mpreference = 59.15, SD = 20.21 vs. Mindifference = 66.20, SD = 17.25; t(300) =

3.26, $p = .001$). This result persisted when comparing the preference to both the positive indifference ($M_{preference} = 59.15$, $SD = 20.21$ vs. $M_{positive} = 71.84$, $SD = 17.78$; $t(301) = 5.81$, $p < .001$) and the baseline (no information) condition ($M_{preference} = 59.15$, $SD = 20.21$ vs. $M_{no\_information} = 65.82$, $SD = 17.88$; $t(296) = 3.16$, $p = .002$).

STUDY 3: INDIVIDUAL CONSUMPTION

Like studies 1 and 2, requestors inferred their friend likes the non-preferred option significantly more when the responder expressed indifference as opposed to preference ($M_{preference} = 52.09$, $SD = 14.85$ vs. $M_{indifference} = 65.78$, $SD = 17.20$; $t(298) = 6.29$, $p < .001$).

STUDY 4: INDIFFERENCE AS A DUAL PREFERENCE

We found that responders inferred their friend likes the non-preferred option C significantly more when the responder expressed indifference as opposed to preference ($M_{preference} = 58.37$, $SD = 18.83$ vs. $M_{indifference} = 64.35$, $SD = 17.63$; $t(295) = -5.47$, $p = .005$). This result persisted, and was larger in magnitude, when responders expressed indifference between only two of the options, as in the dual preference condition ($M_{preference} = 58.37$, $SD = 18.83$ vs. $M_{dual\_preference} = 36.89$, $SD = 24.18$; $t(293) = 4.57$, $p < .001$). Note that when the target is the non-preferred option C, then option C is the only foregone option in the dual preference condition, and is not the only foregone option in the preference condition. Thus, there is indifference with the non-preferred option C in the preference condition and not in the dual preference condition. From this perspective, these results mirror the results for Option A, and again point to a general finding: as more indifference is introduced, the less extreme are inferences of liking of the respective option.

STUDY 7: IS OUR BASIC RESULT A FORM OF MISPERCEPTION?

We collapsed the two preference conditions, prefer Corky's and prefer Central, into one preference condition to increase power and because we had no a priori reason to expect differences. We create the liking of the non-preferred option by taking the minimum liking rating between Corky and Central for each participant.

Using OLS, we regressed the liking of the non-preferred option on role (0 for responder, 1 for requestor), expression (0 for preference, 1 for indifference), their interaction, and a dummy for wave (N = 1,560). The interaction between role and expression was not significant (ßinteraction = .70; t(1,557) = 0.39, p = .696).

WEB APPENDIX B: S1 MANIPULATING THE NUMBER OF FOREGONE OPTIONS

Supplementary Study 1 uses the same paradigm as Study 4 but additionally tests whether increasing the number of foregone options will increase inferred liking. In a consideration set of three options, two options are forgone when responders express a (singular) preference, one option is forgone when responders express dual preference, and zero options are forgone when responders express indifference. If we increase the set size to five options, four options are forgone when responders express a (singular) preference, three options are forgone when responders express dual preference, and zero options are forgone when responders express indifference. If the number of foregone options increases the liking of the target option A, then for both preference and dual preference, requestors should infer that responders like the target more when preferences are expressed in the context of a five-option set than a three-option set.

Method

We recruited 1,020 participants (53.2% female, Mage = 39.2) via Amazon Mechanical Turk.

All participants first read the same scenario as in Studies 2 and 4. Following this, participants were then randomly assigned to one of six conditions in a 2(Set Size: Three, Five) x2(Expression: Preference, Dual Preference, Indifference) between-subjects design. In the three-option set, participants read that they presented their friend with three restaurant options: A, B, and C. In the five-option set, participants read that they presented their friend with three restaurant options: A, B, C, D, and E. In both set size conditions, the preference condition participants read, "After reviewing the restaurants, your friend responds that they prefer A." In both set size conditions, the dual preference condition participants read, "After reviewing the restaurants, your friend responds that prefer either A or B." In the three-option set, indifference condition participants read, "After reviewing the restaurants, your friend responds that prefer either A, B, or C." In the five-option set, indifference condition participants read, "After reviewing the restaurants, your friend responds that prefer either A, B, C, D or E."

Participants were then asked to "Please rate how much you think your friend will like each restaurant." They did so only for restaurant options A, B, and C on sliding scales from 0, “extremely dislikes,” to 100, “extremely likes,” with 50 labeled as “neutral.” The slider’s starting position was set to 50. Lastly, participants filled out demographic information.

Results and Discussion

We ran an OLS regression with the inferred liking of option A regressed on a dummy for expression (indifference as the reference group), a dummy for set size (three-option set as the reference group), and their interaction. Neither the interaction with dual preference ($ß_{interaction\_dual\_preference} = 2.98$; $t(1,116) = 1.32$, $p = .187$) nor with preference were significant ($ß_{interaction\_ preference} = -1.62$; $t(1,116) = -0.74$, $p = .458$). To compare preference to dual preference, we rerun the regression with dual preference as the reference group and did find a significant interaction ($ß_{interaction\_ preference} = -4.60$; $t(1,116) = -2.24$, $p = .025$). Furthermore, when running the regression without the interaction term, there was no main effect of set size ($ß_{set\_size} = -1.09$; $t(1,117) = -1.\ 23$, $p = .219$). Thus, there is no consistent evidence that the number of foregone options influences our results (see Figure S1 below). Instead, we replicate the main findings of Study 4.

FIGURE S1

RESULTS OF STUDY S1

Note. Study S1 effect of preference expression and set size on inferences of responders’ liking for the target option. Bars represent standard errors.

WEB APPENDIX C: STUDY 7 REASSIGNMENT ANALYSIS

Study 7 in the main text had 201 participants across the preference conditions (i.e., prefer Corky's BBQ or prefer Central BBQ) who reported liking ratings that violated their stated preference if they were the responder or the information they learned if they were the requestor. Of the 201 participants, 51 participants reported equally liking both options and 150 reported a liking for the non-preferred option that was higher than the preferred option. In the main text, we dealt with this by dropping these participants who committed violations. In this section, we perform a thought experiment where instead of dropping these participants we reassign them to the appropriate condition consistent with their reported liking ratings. Those whose like Central BBQ more than Corky's BBQ were reassigned to the Central BBQ preference condition, and vice versa for those who like Corky's BBQ more. For the 51 participants who reported liking each equally but were assigned to a preference condition or selected a preference, we still drop these participants like in the main text since it is unclear whether to assign them to indifference or positive indifference which requires setting an arbitrary threshold above the scale midpoint to qualify as positive indifference. We then rerun the same analyses reported in the main text.

Results and Discussion

Using OLS, we regressed the responder's inferred liking of the preferred option on role (0 for responder, 1 for requestor), expression (0 for preference, 1 for positive indifference), their interaction, and a dummy for wave (N = 1,773). We found a marginally significant interaction between role and expression because we set our alpha level to .03 ($ß_{interaction}$ = -2.17; t(1,768)

= -1.99, p = .046). Responders (i.e., stage 1 participants) expressing positive indifference compared to those expressing a preference liked the target to a similar extent ($M_{positive\_indifference}$ = 80.57, SD = 10.32 vs. $M_{preference}$ = 81.75, SD = 11.42; t(772) = 1.45, p = .149). However, replicating previous findings, requestors (i.e. stage 2 participants) wrongly inferred that responders liked the target less when responders expressed positive indifference as opposed to preference ($M_{positive\_indifference}$ = 78.33, SD = 11.62 vs. $M_{preference}$ = 81.68, SD = 10.27; t(997) = 4.68, p < .001). Interestingly, this misprediction comes solely from requestors underestimating responders' liking of the target in the positive indifference conditions ($M_{responders}$ = 80.57, SD = 10.32 vs. $M_{requestors}$ = 78.33, SD = 11.62; t(648) = 2.57, p = .01) and not from requestors overestimating responders' liking of the target in the preference conditions ($M_{responders}$ = 81.75, SD = 11.42 vs. $M_{requestors}$ = 81.67, SD = 10.27; t(1121) = .11, p = .913).

Similarly, to the above analysis, we determine the liking of the non-preferred option by taking the minimum liking rating reported between Corky and Central for each participant. Using OLS, we regressed the liking of the non-preferred option on role (0 for responder, 1 for requestor), expression (0 for preference, 1 for indifference), their interaction, and a dummy for wave (N = 1,710). The interaction between role and expression was not significant ($ß_{interaction}$ = 1.02; t(1,705) = 0.58, p = .561).

Chapter 3

SCARCITY IS ALL RELATIVE: HOW THE VALUATION OF SCARCE GOODS DEPENDS ON THE PRESENCE OF ALTERNATIVES

ABSTRACT

Commodity theory predicts that consumers value goods to the extent to which they are unavailable. The current research is the first to show how the relative scarcity between goods influences valuations, even when objective scarcity (i.e., the number of items left) does not change. Specifically, this research posits that consumers attend to the comparative availability between goods—not only a good's own availability—when making scarcity judgments and purchasing decisions. Although consumers typically consider multiple goods at once, prior work on the evaluation of scarce goods has focused on contexts in which goods are evaluated in isolation. The current research asks how the presence of other goods affects how scarce and abundant goods are evaluated. Six pre-registered experiments demonstrate that valuations of a scarce good depend on the scarcity of alternative goods independent of its own level of unavailability. In particular, people evaluate a scarce good more favorably in the presence of a less scarce alternative than in isolation. The authors discuss the process underlying this effect, as well as the moderators of the results.

Key words: scarcity, abundance, reference points, contrast effects, comparative evaluation

Imagine that someone is headed to a dinner party and would like to bring a bottle of wine that the host and other guests will enjoy. On the way to the party, she visits a small wine shop and asks the owner to help her choose a nice bottle of white wine in the $20 range. The owner gladly selects a couple of different bottles and brings them to the front. The first bottle the owner shows is a bottle of Chardonnay from Chablis. When the owner tells the guest-to-be that the shop has only 11 bottles of this Chablis left, she thinks she has found a fairly strong option worthy of consideration. The owner then picks up the second bottle—a Chardonnay from Bourgogne—and says that only 36 bottles of this varietal remain. This bottle from Bourgogne also seems quite good, but she now views the scarcer Chablis in an even more favorable light. Even though the Chablis's level of scarcity (i.e., availability) has not changed, seeing the relatively more abundant bottle of Bourgogne makes this guest sure that the Chablis is the right choice. In the current research, we are the first to explore this paradox and, in doing so, shed light on the role of different reference points of availability on perceptions of scarcity and subsequent choice.

As in our scenario above, consumers often encounter cues of scarcity or abundance when choosing whether to make a purchase. Traditional, online, and app-based retailers can present goods and services being sold as either abundant or scarce, and several explicit or implicit cues can be employed to alter consumers' perceptions of scarcity. In some cases, like our scenario in which the dinner guest views the Chablis for the first time, consumers may infer that a product is scarce or abundant based on the product's *own* level of availability—for example, by observing empty shelf space in a standalone display, receiving communication from a retailer that a wish list item's supply is limited, or noting explicit cues given by a retailer (e.g., number of rooms available in a hotel, number of wine bottles remaining in the warehouse, number of airline tickets remaining "at this price"). In other cases, consumers—like the person in our scenario evaluating the Chablis after the Bourgogne—may infer that a product is scarce by noting that *other* products are abundant. In the current research, we distinguish between these sources of information to unpack how consumers' inferences of scarcity influence their choices.

There is a vast literature on the effect of scarcity on consumer valuation. Most of this literature studies how products tend to be valued to the extent that they are unavailable (Gierl and Huettl 2010; Hamilton et al. 2019; Lynn 1991, 1992; Sevilla and Redden 2014; Verhallen and Robben 1994; Worchel, Lee, and Adewole 1975; Zhu and Ratner 2015). In particular, these studies have focused on how a good in isolation (i.e., in "separate evaluation mode") is valued more or less as a function of its own level of availability. From a theoretical perspective, we extend this literature by asking how the scarcity of an alternative option affects the preference for a scarce or abundant good (henceforth "joint evaluation mode"), even when objective scarcity levels remain unchanged. Indeed, people take several implicit or explicit reference points into account when making evaluations, and past research has been mute about the potential moderating role of the scarcity of alternatives. The current research asks whether joint versus separate evaluation modes influence preferences for scarce and abundant options. In addition, from a practical point of view, given the ubiquity of multiple-option consideration in consumer choice, we extend this literature by asking how the presence of other products influences the effect of scarcity on evaluation.

It is worth noting that some prior work has suggested the possibility that people differentiate between scarce and abundant goods when they are posed as alternatives. For example, when comparing a relatively scarce and relatively abundant product on adjacent shelves in a retail setting, consumers tend to favor the relatively scarce product, choosing the item that has fewer units on the shelf (Parker and Lehmann 2011; van Herpen, Pieters, and

Zeelenberg 2009). Because participants in these studies evaluate scarce or abundant goods in joint evaluation mode only, it is not possible to distinguish whether another good's scarcity level affects demand for any given product, nor whether consumers are merely sensitive to a given good's current level of availability relative to its initial level (i.e., how much shelf space a given good occupies). The studies in the current research aim to disentangle the effects of these two different reference points, specifically testing how joint evaluation mode affects choice when current availability is held constant.

A novel contribution of this work arises from disentangling these two reference points; specifically, we test the extent to which people use between-good in addition to within-good comparisons. While consumers may contrast a good's current availability with its initial availability, we show that beyond this, they rely on another good's availability in their evaluations of a given item. One might expect that compared to being viewed in isolation, a scarce good may be evaluated more favorably in the presence of an alternative abundant good simply because the addition of an abundant good makes the scarcity of the focal good easier to evaluate (Evangelidis and van Osselaer 2019; Hsee 1996; Hsee et al. 1999; Hsee and Zhang 2010). Yet our studies show that such differences in evaluability do not account for our results. Moreover, it is also plausible that the presence of *any* additional reference point could give rise to a contrast effect (Benary 1924; Campbell, Lewis, and Hunt 1958; Dehaene 1998; Dhar, Nowlis, and Sherman 1999; Fernberger 1920; Gilchrist et al. 1999; Heintz 1950; Mussweiler 2003; Schwarz and Bless 1991; Sherif, Taub, and Hovland 1958; Wever and Zener 1928). However, we find that it is not the presence of merely *any* given reference point, but specifically a contrast effect in which goods are compared against one another that affects evaluations of scarce goods.

In sum, past research has not investigated whether viewing a scarce good alongside another relatively scarce or abundant good affects valuations of the good. We address this gap and, in doing so, disentangle the influence of two scarcity-relevant reference points on consumers' judgments and decisions: the good's initial availability and the current availability of an alternative good. Our work sheds light on the key comparisons consumers make in many real-world settings and contributes to the literature on scarcity by clarifying when and why scarcity effects are likely to arise.

In this work, we label a good as *scarce* if its actual availability is lower than that of a referent good, and we label a good as *abundant* when its actual availability is higher than that of a referent good. We hypothesize that as in our scenario, evaluating a scarce good alongside an abundant one (i.e., joint evaluation mode) will cause consumers like the party-goer to value the scarce good more than if they had evaluated the scarce good by itself (i.e., separate evaluation mode). Analogously, we also predict that valuation of an abundant good in joint evaluation mode will be lower than that in separate evaluation mode. In addition to these two key predictions, we hypothesize that these effects arise due to a comparison between the current availabilities of the focal good and an abundant alternative good. This comparison, in turn, results in an asymmetric contrast effect whereby scarcer goods are more attractive when juxtaposed with relatively abundant ones.

The remainder of the paper is organized as follows. We first review prior work on how scarcity affects valuation. Then we discuss a previously unexplored potential mechanism: contrast effects. In outlining this mechanism, we generate our key hypotheses that a scarce (abundant) good will be valued more alongside an abundant (scarce) good than in isolation.

Finally, we leverage comparison and contrast effects to explain why the effect of scarcity on valuation is driven, in large part, by a contrast of current availabilities.

SCARCITY EFFECTS ON VALUATION

Scarcity has been shown to influence consumers' valuations and behaviors in a myriad of ways. On the one hand, scarcity has been operationalized as a mindset characterized by feeling that one has too few resources (e.g., Roux, Goldsmith, and Bonezzoni 2015; Shah, Mullainathan, and Shafir 2012; Shah, Shafir, and Mullainathan 2015) or as a state in which one has limited time to make a purchase (e.g., Aggarwal, Jun, and Huh 2011). On the other hand, we home in on a more basic type of scarcity that is specific to the product: the mere number of available units (i.e., limited availability; Zhu and Ratner 2015).

The effects of limited availability on valuation have been demonstrated through both numerical and qualitative formulations of scarcity. In the research stream that has focused on numerical formulations of scarcity, people have typically been prompted to evaluate a good that is either scarce (e.g., has 10 units available) or abundant (e.g., has 50 units available), and the findings have shown that people tend to value scarce goods more. Although limited availability can arise due to supplier choices as well as to demand, people value scarce goods to a greater extent when limited availability is attributable to consumer demand (Worchel, Lee, and Adewole 1975). Similarly, a scarce good on a shelf is chosen more often than an abundant one because of popularity (Parker and Lehmann 2011), which can lead to demand cascades (van Herpen, Pieters, and Zeelenberg 2009). Furthermore, people are more likely to select their favorite options among alternatives when their favorite is more scarce than abundant (Zhu and Ratner 2015).

Research that uses more qualitative manipulations of scarcity finds that limited availability as represented by non-numerical, abstract statements (e.g., adequate supply vs. few available) also affects consumers' valuations. Specifically, the choice of the scarce product depends on the extent to which scarcity arises from market conditions (Verhallen and Robben 1994) and whether the good is suitable for conspicuous consumption (Gierl and Huettl 2010). In addition, when a product is seasonal or otherwise naturally scarce, consumers value it more and satiate more slowly (Sevilla and Redden 2014).

The foundation for the effects of scarcity on valuation—whether based on numerical cues or qualitative statements—is commodity theory, which states that "any commodity will be valued to the extent that it is unavailable" (Brock 1968, p. 246). From this perspective, perceptions of scarcity and the resulting effects of scarcity on valuation should depend on the availability of a given good. Because environmental cues influence judgments and preferences, which are often constructed on the spot (Bettman, Luce, and Payne 1998; Simonson 2008), scarcity judgments related to any good likely depend on considerations of other goods' scarcity levels. That is, when we hold constant the availability of a given good and thus rule out the account of commodity theory, evaluating the good in the presence of an alternative good may still influence perceptions of scarcity. In particular, when evaluating a scarce or abundant good, people may recognize the relative scarcity or abundance of other available goods, which in turn results in natural comparisons between the relative availabilities of the focal good and an available alternative. In the next section, we review the basic tenets of comparison effects—highlighting contrast effects in particular—to explain why consideration of the scarcity of an alternative good might influence valuations of a scarce or abundant good.

CONTRAST EFFECTS IN JUDGMENTS

When consumers make comparisons between options, contrast effects or assimilation effects can emerge and often do so automatically and unconsciously (Dehaene 1998; Dhar, Nowlis, and Sherman 1999; Mussweiler 2003; Schwarz and Bless 1991). A contrast effect occurs when differences between objects are emphasized, or more generally, when the judgment of a target shifts away from that of a reference point after its introduction. Early work on contrast effects, for example, examined how the perceived weight of an object (Fernberger 1920; Heintz 1950; Sherif, Taub, and Hovland 1958; Wever and Zener 1928), the sound of a tone (Campbell, Lewis, and Hunt 1958), or the lightness of a color (Benary 1924; Gilchrist et al. 1999) varied as a function of the presence of comparative stimuli. Specifically, an object's weight was judged as heavier (lighter) when the comparator was lighter (heavier). More recently, consumer research has demonstrated contrast effects in a myriad of ways. For example, when discriminating among portable media players, people perceive a target portable media player to be less expensive in the presence of a more expensive alternative (Cunha and Shulman 2011). However, an assimilation effect occurs when similarity is emphasized, or, more generally, when the judgment of a target shifts towards that of a reference point after its introduction. For example, when people are motivated to focus on common features of portable media players, then learning that the product category is priced above average made the target player seem more expensive (Cunha and Shulman 2011). Thus, whether a contrast or an assimilation effect emerges in a particular context is a function of the motivation or goal of the consumer. When the evaluator seeks to differentiate between options, as is the case in many purchasing contexts in which consumers want to obtain the value-maximizing options, a contrast effect is likely to emerge.

While separate evaluation mode induces consumers to evaluate an option with respect to its own attributes, joint evaluation mode prompts comparisons between goods on attributes that differ. Therefore, when evaluating a scarce or abundant good in joint evaluation mode, consumers are more likely to consider the relative availability of the alternative good than they would be in separate evaluation mode. Because contrast effects emerge from a motivation to differentiate, and joint evaluation mode—unlike separate evaluation mode—necessitates differentiation between goods, then it follows that a contrast effect will emerge more strongly in joint evaluation mode than in separate evaluation mode. Thus, in contexts in which a focal good and an alternative are substitutes, we predict that differing levels of current availability will produce a contrast effect whereby the scarce good appears scarcer in the presence of an abundant alternative, and vice versa. To the extent that scarcity judgments influence valuation, this contrast effect, in turn, should lead to a divergence in valuations such that the scarce good is valued more and the abundant good is valued less in joint evaluation mode than in separate evaluation mode.

H_{1a}: A scarce good will be valued more in joint evaluation mode than in separate evaluation mode.

H_{1b}: An abundant good will be valued less in joint evaluation mode than in separate evaluation mode.

Normatively, evaluation mode should not affect perceptions of scarcity or subsequent evaluations of a given good. If a given good's level of availability—which remains constant across separate and joint evaluation modes—determines the effect of scarcity on its valuation, then the mere presence of an alternative good that is more scarce or abundant should not affect the evaluation of a focal good. However, if a contrast effect comparing the current availabilities of the focal and alternative goods arises, then the focal good may be valued more or less in the presence of an alternative, even if its own level of availability remains unchanged. As consumers incorporate available cues from their environment into their choices (Bettman, Luce, and Payne 1998; Simonson 2008), the salient reference point (i.e., the current availability of the alternative good) introduced in joint evaluation mode provides a point of comparison and differentiation that gives rise to a contrast effect.

H_2: Divergence in valuations of the scarce and abundant goods in joint evaluation mode versus separate evaluation mode occurs because consumers focus on the scarcity *differences* between the two goods, rather than the actual scarcity levels of each good.

Importantly, the introduction of an alternative good that is more scarce or abundant could lead to divergence in valuations not only due to a contrast effect, but also by helping consumers understand the meaningfulness of the scarcity level (i.e., level of availability) they encounter. Indeed, compared to viewing an option in isolation (i.e., separate evaluation mode), evaluating the option in the presence of an alternative (i.e., joint evaluation mode) can make attributes more evaluable, which can lead to a divergence in preferences (Evangelidis and van Osselaer 2019; Hsee 1996; Hsee et al. 1999; Hsee and Zhang 2010). In particular, evaluability is “the extent to which a person has relevant reference information to gauge the desirability of target values and map them onto evaluation” (Hsee and Zhang 2010, p. 344-345). While evaluability naturally plays a role in any setting in which joint and separate evaluation modes are varied, a contrast effect can emerge above and beyond that of evaluability. We provide indirect and direct evidence that the pattern of data observed across studies is due to a contrast effect as opposed to evaluability by holding the evaluability of attributes constant across separate and joint evaluation modes, particularly in our latter studies.

OVERVIEW OF STUDIES

To unpack which reference points consumers attend to when evaluating scarce goods, six pre-registered studies examined how the presence of other scarce or abundant goods affects the valuation of a focal good. Studies 1-3 examined whether and to what extent the presence of another good (joint evaluation mode) increases the valuation of a scarce good and decreases the valuation of an abundant good relative to that of the focal good evaluated in isolation (separate evaluation mode). Specifically, Study 1 borrowed from prior literature by manipulating scarcity through varying whether the current availability of a good is high or low. Here, we found evidence that compared to separate evaluation mode, joint evaluation mode increases valuations of a scarce good and decreases valuations of an abundant good. Studies 2a and 2b built on the naturalistic design of Study 1 by introducing a consequential choice and giving participants initial availabilities as a reference point, thereby demonstrating robustness across both ratings-based and choice-based measures. Study 3 bolstered the findings of the first three studies by examining the effect of joint versus separate evaluation modes on scarce goods using different stimuli (service-based rather than goods-based stimuli) containing choice-relevant cues that helped the participants evaluate the options. Notably, in the three studies that provided initial availabilities as a reference point, we found robust evidence that, relative to separate evaluation mode, joint evaluation mode increases valuations of the scarce good (H1a), but we did not find that it decreases valuations of the abundant good (H1b). In Study 4, we provided a conceptual test of our mechanism–namely, that the comparison of the current availabilities (as opposed to the initial availabilities) of each good results in a contrast effect, causing scarce goods to be evaluated more favorably in joint evaluation mode than in separate evaluation mode. Furthermore, Study 4 demonstrated that this effect persists even when the good is extremely scarce and in high demand, a somewhat intuitive boundary condition. Finally, in Study 5 we directly tested our mechanism and found support for the hypothesis that consumers are heavily weighting between-good scarcity information, which moderates the effect of evaluation mode. Studies 3-5 together effectively controlled for evaluability as an alternative explanation for the results.

In all of our studies, we report all measures, manipulations, and exclusions. All studies were pre-registered on AsPredicted.org (see Web Appendix for links), and we perfectly followed all of our pre-registered analysis plans and exclusion rules. The Web Appendix also reports the results of analyses without attention check exclusions (these analyses were not pre-registered). The study materials and data are available at this website: https://osf.io/e4hxz/?view_only=6e51de93af774136b5284ef07f332935

STUDY 1: THE PRESENCE OF A SCARCE OR ABUNDANT ALTERNATIVE

Study 1 aimed to test whether the mere presence of information about another good's scarcity can influence the valuation of a good. Thus, for the first time ever, we test whether scarcity information can impact valuations even when objective scarcity does not change by isolating the effect of one good's scarcity on another. Specifically, we ask how the valuation of a scarce (abundant) good changes in the presence of an alternative abundant (scarce) good. In line with the types of decisions consumers typically make in several online purchasing platforms, we created a naturalistic online shopping scenario in which consumers must decide on a bottle of wine. Participants were presented with numerical scarcity information about a focal bottle of wine (i.e., the remaining number of bottles available) before evaluating it. While some participants evaluated this bottle in isolation, others evaluated it in the presence of another bottle, which was either relatively scarcer or relatively more abundant. This study allowed us to test whether another good's scarcity cues affect valuations of the focal good asymmetrically, increasing the value of the scarce good and decreasing the value of the abundant good. Furthermore, the paradigm is similar to that used in a plethora of studies in which no explicit reference point is given, such as an initial level of availability against which to compare the current level.

Method

We recruited 402 adults (40.8% female, M_{age} = 34, SD = 10.2) via Amazon Mechanical Turk to complete the survey in exchange for $.20. As pre-registered, we dropped 73 participants who failed the attention check and 5 participants who took the study more than once, leaving 324 participants for subsequent analyses.

This experiment followed a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) between-subjects design. All participants read a scenario in which they were asked to imagine they were looking to purchase a bottle of Chardonnay as a friend's birthday gift, and that they visited an online wine shop that offered thousands of wine varieties. Across all conditions, participants were asked to evaluate a focal bottle of wine that was either scarce (i.e., "2 bottles available") or abundant (i.e., "60 bottles available"), which appeared with its name and vintage. Stimuli consisted of an image of a real bottle of wine with its actual name and vintage. A pretest (N = 50) in which several bottles were presented simultaneously showed that judgments of quality ("How would you rate the quality of this product?"; 1 = "very low quality," 4 = "average quality," 7 = "very high quality") did not differ between the two bottles (M_1 = 4.78, SD = 1.13; M_2 = 4.86, SD = 1.07; t(49) = 0.41, *p* = .687), and that both were perceived to be above average quality (comparisons to midpoint of 4; bottle 1: t(49) = 4.88, *p* < .001; bottle 2: t(49) = 5.69, *p* < .001).

In the separate evaluation mode condition, participants saw a single bottle of wine that was either scarce or abundant. To measure participants' valuation of the bottle, we asked them to indicate how likely they would be to add the bottle to their online shopping cart (1 = "very unlikely," 9 = "very likely"). In the joint evaluation mode condition, participants also saw the focal (scarce or abundant) bottle of wine, but with one key difference: an additional bottle was present. Specifically, those evaluating the scarce bottle in the joint evaluation mode condition were also shown the image, name, vintage, and current availability of the abundant bottle, and those evaluating the abundant bottle were given the analogous information for the scarce bottle. Therefore, participants evaluated a single bottle—of which there were either 2 or 60 bottles remaining—in either the presence or absence of the other bottle. Figure 1 depicts example stimuli for the abundant good presented in the separate evaluation mode and joint evaluation mode conditions.

Results and Discussion

We conducted a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) ANOVA on the likelihood of adding the focal wine to the shopping cart. The results revealed the predicted significant interaction of scarcity level and evaluation mode (F(1, 320) = 24.06, $p < .001$; see Figure 2) as well as a simple effect of evaluation mode (F(1, 322) = 13.74, $p < .001$). Consistent with our predictions, the scarce good was valued more in the presence of the abundant good than by itself (M_{joint} = 6.80, SD = 1.91 vs. $M_{separate}$ = 6.08, SD = 1.92; t(164) = 2.38, $p = .018$). Conversely, the valuation of the abundant good was lower in the presence of the scarcer good than by itself (M_{joint} = 5.01, SD = 2.09 vs. $M_{separate}$ = 6.33, SD = 1.46; t(156) = 4.62, $p < .001$). In contrast to prior work documenting the effect of availability on evaluation (Gierl and Huettl 2010; Hamilton et al. 2019; Lynn 1991, 1992; Sevilla and Redden 2014; Verhallen and Robben 1994; Worchel, Lee, and Adewole 1975; Zhu and Ratner 2015), we did not find an effect of scarcity level within the separate evaluation mode condition (M_{scarce} = 6.08 vs. $M_{abundant}$ = 6.33, t(150) = .91, $p = .366$). The fact that we do not observe an effect of availability on evaluation here could be due to the specific scarcity parameters chosen, the scenario or paradigm, or an absence of other moderators tested in other research settings.

The results of Study 1 yielded support for our main predictions. Namely, we found that people are more likely to purchase a scarce good when an abundant good is present and are less likely to purchase an abundant good when a scarce good is present. It is worth noting that we did not find evidence for the basic prediction of commodity theory—that the mere availability level of a single good drives valuation of the good. Importantly, we demonstrate that although the relative scarcity between goods affects valuations, changes in a good’s current availability do not.

STUDIES 2A AND 2B: THE PRESENCE OF AN EVALUABLE ALTERNATIVE

Study 1 provided evidence for our first two hypotheses in a naturalistic, externally valid setting. In particular, a scarce (abundant) good was valued more (less) in joint evaluation mode than in separate evaluation mode. However, it is possible that joint evaluation mode led people to find the scarce good more attractive because the presence of the abundant good gave them a benchmark against which to evaluate the scarcity of the focal good. That is, it is possible that the asymmetric effect of joint evaluation mode on valuations of scarce and abundant goods arises from the fact that joint evaluation mode makes any scarcity level more evaluable (Hsee 1996; Hsee et al. 1999; Hsee and Zhang 2010). Study 1 participants who evaluated a wine bottle of which only two remained had more information in joint evaluation mode than they did in separate evaluation mode, since knowledge of a more abundant bottle provides an additional reference point to which the scarcity of the focal bottle can be compared. However, if the initial availability of the bottles had been provided, then the presence of another scarce or abundant bottle—no matter how objectively scarce—does not provide additional information about the objective availability of any focal bottle. Thus, in order to make the choice context more evaluable, we provided participants with a reference point regarding the initial availability of units available (i.e., 100). Doing so makes all given scarcity levels equally evaluable along an interval scale of scarcity (i.e., 0-100 available).

Several design features of Studies 2a and 2b are worth noting. First, to better assess real behavior and achieve greater external validity, we employ incentive-compatible choice designs. Thus, in contrast to Study 1, these studies ask participants to choose for themselves rather than for a friend. Second, to rule out the possibility that participants might infer that scarcity is a function of retailers' strategic stocking decisions, participants were informed that the online wine shop restocks its wines regularly and had recently restocked at the beginning of the month, ensuring that both goods were replenished at the same rate and time. Finally, these studies used different scarcity or abundance parameters. In Study 2a, scarcity levels were 5 and 85 (both out of 100). In Study 2b, scarcity levels were 51 and 91 (both out of 100), ensuring that both scarcity levels had the same number of digits (i.e., two) and that both goods were fairly abundant.

Method

Both studies recruited participants via Amazon Mechanical Turk to complete the survey for $.30. For Study 2a, we recruited 805 adults (40.9% female, M_{age} = 37.8, SD = 12.8). As pre-registered, we dropped 133 participants who failed the attention check and 4 participants who took the study more than once, leaving 668 participants for subsequent analyses. For Study 2b, we recruited 811 adults (40.1% female, M_{age} = 37.4, SD = 12.4), and dropped 115 participants who failed the attention check and 4 participants who took the study more than once, leaving 692 participants for subsequent analyses.

This experiment followed a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) between-subjects design.

In contrast to Study 1, in Studies 2a and 2b participants were informed that some of them would be randomly chosen to receive their product choice in addition to their regular compensation. (In actuality, to ensure participants' privacy, some received a $25 gift card for an online wine store, Wine.com). To control for inferences participants might make about retailers' strategic stocking decisions, they were also informed that the wine website stocks 100 bottles of each wine at the beginning of each month. This also served as the initial reference point and created the implicit, evaluable scale of 0 to 100 bottles available.

To measure choice, all participants were asked, "Please select whether you would choose the [focal wine] or to look for a different bottle." Across both conditions, the dependent variable is the choice between the focal wine or continued search. Note that in the joint evaluation mode condition, continued search could include the non-focal wine presented. Because of the overwhelming number of possibilities embedded in an outside option, this serves as a conservative test for the effect of joint evaluation mode or of scarcity as consequential cues in choice.

Results and Discussion

For both studies, we ran a logistic regression of choosing the focal wine on evaluation mode (1 = joint, 0 = separate), scarcity level (1 = scarce, 0 = abundant), and their interaction, using robust standard errors. In Study 2a, the coefficient on the interaction was significant ($\beta_{interaction}$ = .78, SE = .33, Z = 2.36, p = .018; see Figure 3). Indeed, consistent with Study 1, the scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (68% vs. 31%, $\chi^2(1) = 47.01$, $p < .001$). However, in contrast to Study 1, the abundant good was also incidentally chosen more frequently in joint evaluation mode than in separate evaluation mode (47% vs. 28%, $\chi^2(1) = 12$, $p = .001$), though to a lesser extent than the scarce good. For completeness, we report all regression coefficients, including simple effects (i.e., the effect of the focal independent variable on the dependent variable conditional on the other, non-focal, independent variable in the interaction taking a value of 0). We also found a simple effect of evaluation mode (β_{joint} = .80, SE = .23, Z = 3.43, p = .001), and no simple effect of scarcity level (β_{scarce} = .12, SE = .23, Z = 0.49, p = .622; within separate evaluation mode: 31% vs. 28%, $\chi^2(1) = .24$, p = .622) on choosing the focal wine.

In Study 2b, the results mirrored those of Study 2a. The coefficient on the interaction was significant ($\beta_{interaction}$ = .77, SE = .32, Z = 2.40, p = .016). Indeed, consistent with Study 1, the scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (56% vs. 28%, $\chi^2(1) = 31.31$, $p < .001$). As in Study 2a, the abundant good was also incidentally chosen more frequently in joint evaluation mode than in separate evaluation mode (42% vs. 31%, $\chi^2(1) = 4.69$, $p = .030$), but to a lesser extent than the scarce good. We also found a simple effect of evaluation mode ($\beta_{joint} = .49$, SE = .23, Z = 2.16, $p = .031$), and no simple effect of scarcity level ($\beta_{scarce} = -.14$, SE = .23, Z = -.62, $p = .532$; within separate evaluation mode: 31% vs. 28%, $\chi^2(1) = .39$, $p = .532$) on choosing the focal wine. In contrast to the Study 1 results, the abundant good was also incidentally chosen more frequently in joint evaluation mode than in separate evaluation mode (42% vs. 31%, $\chi^2(1) = 4.69$, $p = .030$).

Studies 2a and 2b extended the results of Study 1 by demonstrating that in a real-choice task, the choice of a scarce good is greater in joint evaluation mode than in separate evaluation mode. Notably, this effect persists when people are given an initial availability—and thus are able to evaluate scarcity at any given level. Moreover, Study 2b showed that this effect holds even when the scarcer good is somewhat abundant. In these studies, we observed an asymmetric contrast effect in the effect of evaluation mode on choice—particularly with respect to scarce rather than abundant goods. That we find consistent evidence for the effect of evaluation mode on the scarce good and not the abundant good is consistent with the notion that consumers may be interested in seeking the best option rather than devaluing options perceived to be inferior.

STUDY 3: CHOOSING WITH INFORMATIVE VISUAL CUES

Studies 1-2b found consistent support that the scarcer good is valued to a greater extent in joint evaluation mode than in separate evaluation mode. This effect holds in both hypothetical and real-choice contexts, and when initial availabilities of goods are present or absent. Study 3 extends the findings of these initial studies to a context in which visual cues provide valuable information with which to evaluate options. In particular, this study asks participants to evaluate a service that is commonly booked on internet-based platforms: hotel rooms. In contrast to pictures of wine bottles, pictures of hotel rooms contain more information about the quality of the actual product or service, which makes information about scarcity potentially less valuable than it was in these initial studies. Two additional pieces of information were included to rule out potential confounds in the previous studies. First, because participants in the joint evaluation mode condition also have more information about the possible distribution of scarcity levels, we provided relevant distributional information that rules out the possibility that joint evaluation mode allows for more learning about different potential scarcity levels. In addition to the fact that availability is naturally bounded within an interval scale (e.g., 0-50 in this study), this distributional information effectively controls for an alternative account that scarcity is merely more evaluable in joint evaluation mode. Second, participants were told that all options were the same price to rule out an alternative account that participants choose scarce options due to affordability concerns (e.g., the option is scarce because it is on sale and thus cheaper).

Method

We recruited 803 adults (49.4% female, M_{age} = 36.6, SD = 11.4) via Amazon Mechanical Turk to complete the survey in exchange for $.35. As pre-registered, we dropped 130 participants who failed the attention check and 2 participants who took the study more than once, leaving 671 participants for subsequent analyses.

This experiment followed a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) between-subjects design. The design was substantively similar to that of Study 2b, with a few exceptions. First, we used new stimuli, hotels (see Figure 4). A pretest (N = 49) showed that judgments of quality ("How would you rate the quality of this hotel?"; 1 = "very low quality," 4 = "average quality," 7 = "very high quality") did not differ between the two hotels (M_1 = 5.50, SD = 1.08; M_2 = 5.51, SD = 0.82; t(48) = 0.13, p = .898) and that both were perceived to be above average quality (comparisons to midpoint of 4; hotel 1: t(48) = 9.63, $p < .001$; hotel 2: t(48) = 12.90, $p < .001$). Participants were given the following prompt: "Imagine that you are going on a trip and you need to book a hotel for this trip in the next couple of days. You decide to look on Booking.com for places to stay." Second, to show that the effect arises above and beyond that of evaluability, participants were given the following information regarding the distribution of scarcity levels: "As of right now, 20% of the hotels have 1 to 10 rooms left, 25% have 11 to 20 rooms left, and 55% have 20 to 50 rooms left." Third, to ensure that perceptions of price were held constant, participants learned that "all of these hotels are within your budget" and were again informed that the hotel(s) has (have) "the same price per night as other hotels within your budget." Finally, in this study, the scarce hotel had 11 (out of 50) available rooms, while the abundant one had 49 (out of 50) available rooms.

Results and Discussion

We ran a logistic regression of choosing the focal hotel on evaluation mode (1 = joint, 0 = separate), scarcity level (1 = scarce, 0 = abundant), and their interaction, using robust standard errors. Importantly, the coefficient on the interaction was (marginally) significant ($\beta_{interaction}$ = .66, SE = .35, Z = 1.89, p = .059). Consistent with Studies 1-2b, an asymmetric effect of evaluation mode on choice arose such that the scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (84% vs. 62%, $\chi^2(1) = 21.45$, $p < .001$). Incidentally, as in Studies 2a and 2b, the abundant good was also chosen more frequently in joint evaluation mode than in separate evaluation mode (64% vs. 52%, $\chi^2(1) = 5.15$, $p = .023$), albeit to a lesser extent than the scarce good. Additionally, the analysis revealed a simple effect of evaluation mode (β_{joint} = .52, SE = .23, Z = 2.26, *p* = .024) and a simple effect of scarcity level (β_{scarce} = .43, SE = .21, Z = 2.02, *p* = .044). Here, as in Study 1 (but in contrast to Studies 2a and 2b), we found an effect of scarcity level on choice within the separate evaluation mode condition (62% vs. 52%, $\chi^2(1) = 4.09$, *p* = .043).

Taken together, the results of Study 3 replicate the findings from the studies thus far using a different set of stimuli in which visual cues were informative. Moreover, the results held even when participants were given explicit information about the distribution of scarcity levels, which rules out the possibility that those in joint evaluation mode learn about the existence of goods with other scarcity levels.

Notably, over our first few studies, the scarce good has availabilities indicating extreme scarcity (i.e., 2 and 5). A priori, extreme scarcity may be considered a natural boundary of the effect that a scarce good is valued to a greater extent in joint evaluation mode than in separate evaluation mode: if a good is already *very* scarce, as indicated by a near-zero availability level (and thus perceived to be popular), then the effect of a relatively abundant alternative good on valuations of that scarce good should be small or near zero. Thus, this setting provides a basis for a conservative test of our predictions.

Finally, while the first four studies provided converging evidence that the scarcer good is valued to a greater extent in joint evaluation than in separate evaluation mode, the evidence for evaluations of the abundant good is mixed. As a result, in the subsequent two studies, we shift our focus to the scarce good to unpack why it is evaluated differently depending on evaluation mode.

STUDY 4: PREFERENCE FOR A SCARCE GOOD UNDER HIGH DEMAND

Thus far, we have demonstrated that a scarce good is valued more in joint evaluation than in separate evaluation mode, even when the objective scarcity level (i.e., its own level of availability) remains unchanged. We propose that this effect of evaluation mode on valuation arises from a contrast effect whereby another good's relative abundance makes the scarce good appear more attractive. Therefore, even under conditions of extreme demand (i.e., low availability relative to initial quantities) in which each good is quite scarce, we expect that the effect of joint versus separate evaluation mode would persist. If participants compare the current availabilities of goods and largely neglect to compare the current availability of a good to its initial quantity, then then a contrast effect between goods' current availabilities ought to persist even when initial availability is multiplied tenfold. This finding would imply that unlike traditional conceptualizations of the effect of scarcity on value, comparisons *between* goods are strong drivers of scarcity judgments and related valuations. This study examines whether the effect of evaluation mode on evaluations of a *scarce* (i.e., not abundant) good persists in the face of high or extremely high demand.

Method

We recruited 1,001 adults (47.3% female, M_{age} = 36.2, SD = 11.3) via Amazon Mechanical Turk to complete the survey in exchange for $.30. As pre-registered, we dropped 145 participants who failed the attention check (there were no duplicates), leaving 856 participants for subsequent analyses.

This experiment followed a 2 (evaluation mode: separate, joint) x 2 (initial availability: low, high) between-subjects design. In contrast to the previous studies, in Study 4 participants evaluated the scarce good only; that is, the focal good was always the scarce good, and the referent good was always abundant. Thus, across all conditions, participants evaluated a focal bottle of wine that had 5 remaining bottles. Specifically, they responded to a question about choosing wine to bring to a party: “Please select whether you would choose to purchase the Cristo wine or to continue searching for a different wine to purchase on the site.” In the joint evaluation condition, participants evaluated the scarce bottle alongside a referent bottle of wine of which 45 bottles remained. In the low initial availability condition, the initial availability given for the wines was 100, identical to that given in Studies 2a, 2b, and 3. In the high initial availability condition, we increased initial availability tenfold (to 1,000), thereby making both the focal and referent wine extremely scarce. Participants were also given distributional information similar to that given in Study 3. In particular, they were told: “As of today, 20% of the wines have 1 to 20 bottles available, 25% have 21 to 45 bottles available, and 55% have greater than 45 bottles available. All of these wines are within your budget.”

Results and Discussion

We ran a logistic regression of choosing the scarce wine on evaluation mode (1 = joint, 0 = separate), initial availability level (1 = high, 0 = low), and their interaction, using robust standard errors. As predicted, the interaction was not significant ($\beta_{interaction}$ = .18, SE = .31, Z = .58, *p* = .565). Despite the tenfold increase in initial availability—and thus a tenfold increase in the objective scarcity level of the scarce and abundant goods—the effect of joint evaluation mode on choosing the scarce good did not differ across low and high initial availability levels (see Figure 5). Indeed, we found an effect of a joint evaluation mode (vs. separate evaluation mode) on choosing the scarce good both when initial quantities were low and when they were high, and these effects were remarkably similar (low initial availability: 77% vs. 53%, $\chi^2(1)$ =

27.68, $p < .001$; high initial availability: 77% vs. 48%, $\chi^2(1) = 33.74$, $p < .001$). We thus found a simple effect of evaluation mode ($\beta_{joint} = 1.08$, SE = .21, Z = 5.18, $p < .001$), though no simple effect of initial availability ($\beta_{high} = -.18$, SE = .12, Z = -.91, $p = .360$).

These findings show that even under conditions of extreme demand (i.e., low current availability relative to initial quantities), the effect of joint versus separate evaluation mode on choosing the scarce good persists. The results thus far suggest that when evaluating scarce options alongside abundant ones, people rely on comparisons between the current availabilities of goods regardless of whether the differences between initial and current quantities are extremely large. In the next study, we provide direct evidence that the effect of evaluation mode on valuation is driven by a contrast between the scarcity of the focal good and relative abundance of another good.

STUDY 5: FOCUSING ON THE RELATIVE SCARCITY BETWEEN GOODS

Studies 1-3 provided consistent evidence that even when objective scarcity remains unchanged, a scarce good is valued to a greater extent in joint evaluation mode than in separate evaluation mode. In Study 4, this effect persisted even under conditions of extreme demand (i.e., high initial availability), providing suggestive evidence that comparisons *between* goods' current availabilities are the strongest drivers of scarcity judgments and related valuations. Study 5 aimed to provide direct evidence of this process by manipulating people's attention to the differences in current availability. We propose that joint evaluation mode causes people to draw comparisons between the scarcity levels (i.e., current availabilities) of a focal good and another good, which results in a contrast effect above and beyond that of evaluability. As a corollary to this proposition, considering the scarcity level of any alternative good with greater availability ought to make the scarce good more attractive in separate evaluation mode. Thus, in Study 5, we manipulated whether participants focused on an alternative, more abundant good and attended to the differences in current availability between the focal and alternative good. We predicted that the effect of evaluation mode on choice would be attenuated when participants were asked to focus on the relative scarcity differences between the scarcer focal good and an abundant alternative.

Method

We recruited 703 adults (42% female, M_{age} = 38.1, SD = 12.4) via Amazon Mechanical Turk to complete the survey in exchange for $.30. As pre-registered, we dropped 92 participants who failed the attention check and 2 participants who took the study more than once, leaving 609 participants for subsequent analyses.

This experiment followed a 2 (evaluation mode: separate, joint) x 2 (reference: focus on differences, control) between-subjects design. Participants were asked to imagine that they were visiting an online wine shop to prepare for an upcoming dinner party they planned to attend. They were then assigned to one of the four conditions. The control conditions of Study 5 were the same as the low initial availability condition in Study 4 in which participants evaluated a scarce, focal wine (with 5 out of 100 bottles available) either by itself or alongside an abundant alternative wine (with 45 out of 100 bottles available). In the focus-on-differences condition, participants were additionally asked to explicitly consider an abundant alternative wine (with 45 out of 100 bottles available). Specifically, those in the separate evaluation mode condition read, “While viewing the wine below, please take a moment to think about another wine for sale which has 45 remaining bottles;” those in the joint evaluation mode condition read, “While viewing the wine below, please take a moment to think about each wine for sale.” Then, participants were asked to “consider the difference between the availabilities, 45 and 5, of these two wines. Mentally take the difference between these two availabilities.” Note that the focus-on-differences condition in separate evaluation mode became more similar, though not identical, to joint evaluation mode: participants were asked to think about an alternative but were not presented with one as an actual alternative option. Finally, all participants chose between the focal wine (which was scarce) or continuing to search for a different wine, as in Study 4.

Moreover, as in Studies 3 and 4, the inclusion of an initial availability level in conjunction with information about the distribution of scarcity levels ensures that the joint evaluation mode and separate evaluation mode conditions are largely informationally equivalent and thus equally evaluable.

Results and Discussion

We ran a logistic regression of choosing the focal wine on evaluation mode (1 = joint, 0 = separate), reference (1 = difference, 0 = control), and their interaction, using robust standard errors. We found a significant interaction ($\beta_{interaction} = -.70$, SE = .36, Z = 1.97, $p = .049$; see Figure 6). As predicted, in the control, we replicated our previous findings that the scarce good is chosen more frequently in joint evaluation mode than in separate evaluation mode (75% vs. 50%, $\chi^2(1) = 19.73$, $p < .001$). However, when participants focused on the differences in scarcity levels between the two goods, the effect was attenuated (75% vs. 67%, $\chi^2(1) = 2.25$, $p = .133$). Additionally, the analysis revealed simple effects of evaluation mode ($\beta_{joint} = 1.09$, SE = .25, Z = 4.37, $p < .001$) and of reference ($\beta_{difference} = .70$, SE = .24, Z = 2.97, $p = .003$).

Thus, the effect of joint evaluation mode on preferences for the scarce good was attenuated when participants were prompted to think about differences between current availabilities of the goods. This process, in turn, partially explains why joint evaluation mode causes people to draw comparisons between the scarcity levels (i.e., current availabilities) of a focal good and another good. We therefore find direct evidence that comparisons *between* goods' current availabilities drive valuations. Importantly, this evidence emerges even though the scarcer good is already quite scarce and the initial availability is held constant—and thus scarcity levels are equally evaluable across conditions.

GENERAL DISCUSSION

Consumers tend to value products more highly when they are scarce than when they are abundant. While the impact of scarcity on consumers' valuations has been widely studied (Gierl and Huettl 2010; Hamilton et al. 2019; Lynn 1991, 1992; Sevilla and Redden 2014; Verhallen and Robben 1994; Worchel, Lee, and Adewole 1975; Zhu and Ratner 2015), past work has failed to emphasize the fact that consumers consider alternative goods when evaluating scarce and abundant goods. Given that the presence of an alternative abundant or scarce good can create a contrast effect (Dehaene 1998; Dhar, Nowlis, and Sherman 1999; Mussweiler 2003; Schwarz and Bless 1991), we examined whether valuations of a scarce or abundant good depend on the scarcity of alternative goods, even when the focal good's level of unavailability is held constant.

Findings from six pre-registered studies demonstrated that people value a scarce good more when they evaluate it alongside an alternative, abundant good than when they view it by itself (Studies 1-5). That is, a scarce good is valued more highly in joint evaluation mode than in separate evaluation mode. This is because comparing the current availabilities (as opposed to the initial availabilities) of each good results in a contrast effect, causing scarce goods to be valued more highly in joint evaluation mode than in separate evaluation mode (Studies 4 and 5). This asymmetric effect of joint evaluation mode on valuations of scarce and abundant goods cannot be explained by the fact that joint evaluation mode makes scarcity levels more evaluable (Studies 3-5). Furthermore, the increased valuation of scarce goods that results from joint evaluation mode persists even in the face of extremely high demand (i.e., low initial availability), suggesting that people attend to contrasts between current availability levels (Study 4). Furthermore, our findings are robust across different measures of evaluation—namely, across ratings and choice. Thus, we provide evidence for an asymmetric contrast effect of scarcity on valuation that is not accounted for by differences in evaluability. These findings build on the prior literature by highlighting that goods are valued not only to the extent to which they are scarce, but also to the extent to which alternative goods are abundant.

In sum, we demonstrate that a scarce good is valued more in joint evaluation mode than in separate evaluation mode due to the contrast between a relatively abundant good's current availability and its own. We show that this effect occurs even when the objective availabilities are held constant, and even though the scarce good is fully evaluable.

Theoretical Contributions

This research makes important contributions to the literature on scarcity and reference points. Prior work has largely examined how consumers evaluate scarce or abundant products when they are viewed in isolation (Gierl and Huettl 2010; Hamilton et al. 2019; Lynn 1991, 1992; Sevilla and Redden 2014; Verhallen and Robben 1994; Worchel, Lee, and Adewole 1975; Zhu and Ratner 2015), and the scant research regarding how consumers view scarce and abundant products together has been mute about which scarcity-related cues drive the choice for a scarce product (Parker and Lehmann 2011; van Herpen, Pieters, and Zeelenberg 2009). First, we document a novel effect, demonstrating that an alternative good's scarcity plays a critical role in consumers' scarcity judgments and related valuations of goods. Whereas past literature treats scarcity as a mere function of a good's own level of unavailability, we show that the relative scarcity of alternative goods has a notable impact on scarcity judgments.

Second, we unpack the mechanism underlying past work on shelf-based scarcity. In particular, consumers' preference for products that appear on relatively empty shelves rather than relatively full ones could have been driven by the extent to which each individual good's shelf was empty or full rather than the comparison between shelves (Parker and Lehmann 2011; van Herpen, Pieters, and Zeelenberg 2009). Our findings shed light on this puzzle by showing that while a good's own empty shelf space may influence consumer choices, it is the relative unavailability of an alternative good that consistently and robustly influences scarcity-based choices.

Finally, we contribute to the literature on the consideration of multiple reference points in consumer decision making. In particular, the current research disentangles the effects of two important scarcity-related reference points: a good's initial quantity available and an alternative good's current quantity available. While prior research suggests that a good's own level of unavailability (e.g., 5 remaining vs. 100 initially) drives scarcity judgments, we provide evidence that the latter reference point (e.g., 5 remaining of good A vs. 45 remaining of good B) has a disproportionate effect on evaluations. Although prior work has examined the differential effects of absolute versus relative reference points (Lockhead 2004; Olson, Goffin, and Haynes 2007; Simonson 2008; Stewart, Brown, and Charter 2005) as well as the general consideration of multiple reference points (Boles and Messick 1995; Kahneman 1992; Ordóñez, Connolly, and

Coughlan 2000; Schweitzer 1995; Sullivan and Kida 1995), to our knowledge, we are the first to unpack the differences between within-good and between-good comparisons that affect scarcity judgments and subsequent evaluations.

Directions for Future Research

While the effect of evaluation mode on valuations of scarce goods was robust with respect to a variety of scarcity levels, whether choosing a good or a service, and whether choosing for oneself and/or for others, it is worth noting that these studies examined settings in which consumers had some ex ante uncertainty about the options under consideration. One source of uncertainty is the extent to which consumers have prior knowledge about which good represents the best choice. Scarcity cues can resolve this uncertainty either as a form of social proof (Banerjee 1992; Cialdini 2007; Goldstein, Cialdini, and Griskevicius 2008) whereby one can infer that the scarce choice is in greater demand, or as a signal of limited supply that signals high quality. In the case of social proof, the party-goer mentioned earlier might rely on others' choices to determine which wine to select. But in the case of limited supply, if the party-goer were to learn that the scarce wine was produced in a small batch of 100 bottles, she may be more inclined to select it. In both cases, the scarcity cue provides meaningful information about the good, which is likely an important moderator of the effect of scarcity on evaluation as well as the effect of evaluation mode on judgments of a scarce good. One might expect that if the referent good's scarcity cue were not meaningful (e.g., half of the case was broken in transit), then the presence of an abundant alternative may not affect preferences for the scarce good. However, research on anchoring effects has demonstrated that even irrelevant anchors influence judgments and choices (Chapman and Johnson 2002; Simonin and Ruth 1998). Thus, while the studies in this paper abstract away from the source of scarcity, future research should examine how the source of an alternative good's scarcity impacts preferences for scarce and abundant goods.

Another source of uncertainty arises from the extent to which consumers perceive the options to be differentiable. That is, for the presence of an alternative good to affect demand for a scarce product, consumers ought to perceive some qualitative differences between alternatives in the product category. In our studies, we expected consumers to believe that there are subtle or possibly drastic differences in how wines taste or how comfortable hotel rooms are. Yet if a consumer were considering a bus in which 5 seats remained, it is unlikely that learning that a bus leaving one hour later had 15 open seats would influence his choice. One might predict that as the goods become increasingly similar, the scarcity level of an alternative good is less meaningful. In line with the literature, we examined settings in which consumers perceive differentiation between products. Future research could explore the extent to which homogeneity within a product category might influence the effect of evaluation mode on valuations of scarce goods.

While we limited our research to cases in which consumers choose among similar products, the effect of evaluation mode on valuations of scarce goods may vary as a function of the similarity between goods. Specifically, we examined cases in which two options were directly comparable (i.e., substitutable). Further research should examine cases in which consumers encounter scarcity between goods that are complementary. There may be settings in which another product's scarcity could increase demand for a given good. For instance, when shoppers are planning a barbecue, a shelf of hot dog buns that seems relatively empty may lead to a hurried scramble toward the aisle containing hot dogs. In line with spillover effects (Oppenheimer, LeBouef, and Brewer 2008), one might expect that the relative scarcity or abundance of a complementary good could enhance or depress demand for a given focal option, respectively. Moreover, it is possible that the effect of an alternative on preferences for a scarce good could vary depending on whether the alternative is a premium or generic option. Suppose someone is considering a generic pair of Oakley sunglasses (of which a few pairs remain) and then notices on the adjacent shelf only one remaining pair of Oakley's high-end sunglasses. Spillover effects might occur such that an assimilation effect (rather than a contrast effect) emerges; the relative scarcity of the high-end product could signal that the Oakley brand is highly valued, thereby causing the generic pair of the same brand to be valued more highly in joint evaluation mode than it would be in separate evaluation mode.

Importantly, to allow for clean comparisons in objective scarcity, we focused chiefly on numerical rather than qualitative scarcity cues. Yet consumers regularly encounter qualitative scarcity cues that could bear on the findings reported. Product listings and marketing communications regularly employ certain words or phrases that connote ideas of scarcity (e.g., "only," "limited"). For example, Etsy.com listings often display an image of an hourglass accompanied by the text "Almost gone. There's only 2 left"; Zalando, a major online apparel and accessory retailer, informs consumers about the scarcity of product sizes with phrases such as "Only 1 left." Although it is difficult to evaluate such phrases, we might expect that our findings would replicate if the scarce good were presented as qualitatively scarce while the abundant good were presented as qualitatively abundant. However, it is possible that these qualitative cues—when accompanied by numerical ones—could dampen the impact of the numerical cues such that the effect of an alternative good on evaluations of a focal option is attenuated. It would be interesting for future research to examine how quantitative and qualitative manipulations of scarcity might interact with one another to differentially influence how scarcity reference points shape valuation.

Finally, one reason scarcity cues affect valuation is that they provide an implicit recommendation by way of popularity (Brock 1968; Parker and Lehmann 2011). However, both online and traditional retailers frequently use explicit recommendations to move consumers toward a particular choice (Eliashberg and Shugan 1997; Gershoff, Mukherjee, and Mukhopadhyay 2003). Although explicit recommendations are often effective, they also have the propensity to backfire—especially when they contradict consumers' initial impressions of the product (Fitzsimons and Lehmann 2004). Thus, there seem to be two opposing predictions. On the one hand, if the abundant reference good is accompanied by a recommendation, then the effect of evaluation mode on evaluations of the scarce good may be attenuated as the recommendation makes the abundant good more attractive. On the other hand, if recommendations backfire, we might expect that the recommended abundant good could make the scarce product more attractive, especially if the abundant good were judged to be less popular. Future research could examine the effect of a recommendation and its interaction with a referent good's scarcity to see which of these two predictions prevails.

Practical Implications

The current research has direct implications for marketing practice, specifically for marketers who must make strategic decisions about how to display goods in retail settings, how to construct relevant online search results, and how to communicate product availability directly to the consumer. For traditional brick-and-mortar retailers, our findings suggest that standalone displays of scarce goods might not always convey the relative scarcity of the given product.

Instead of placing a single scarce product display at the end of an aisle, stores should consider always placing them side-by-side with a more abundant product, particularly when the goal is to move the scarce product. In e-retail contexts, marketers can structure search results by making use of comparative recommendations. When consumers are viewing a product description, the retailer can suggest less scarce alternatives to bolster their attraction to the focal product. Such a feature could be integrated simply via existing recommendations such as displays of what other customers also bought or searched for recently. Whether in online or in physical retail settings, scarce products may be sold more readily when they are juxtaposed with relatively abundant alternatives.

All retailers—whether online or traditional—can leverage scarcity cues in their written and electronic communications with customers. Online retailers, in particular, often send reminders to customers about their most recent shopping experiences or make suggestions about which products may interest them given past search and purchase decisions. Rather than merely reminding customers that a particular item remains in their shopping cart, they can bolster this information by highlighting an abundant alternative. In a similar vein, retailers could benefit from conveying not only the limited availability of items recently viewed, but also the relative abundance of similar items. Such contrasts could be leveraged to increase customer follow-through or increase engagement with the retailer.

Conclusion

This research asked how consumers evaluate a scarce good in the presence or absence of an abundant alternative. Across six pre-registered studies, we found that a scarce good is valued more when presented alongside an abundant good than when viewed in isolation. This effect arises from a contrast between the goods' current levels of availability. Thus, perceived scarcity is a function not merely of a good's own objective availability, but also of the relative scarcity or abundance of the good's alternatives.

REFERENCES

Aggarwal, Praveen, Sung Youl Jun, and Jong Ho Huh (2011), "Scarcity Messages," *Journal of Advertising*, 40 (3), 19-30.

Banerjee, Abhijit V. (1992), "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107 (3), 797-817.

Benary, Wilhelm (1924), "Beobachtungen zu einem Experiment über Helligkeitskontrast," *Psychologische Forschung*, 5 (1), 131-42.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research,* 25 (3), 187-217.

Boles, Terry L. and David M. Messick (1995), "A Reverse Outcome Bias: The Influence of Multiple Reference Points on the Evaluation of Outcomes and Decisions," *Organizational Behavior and Human Decision Processes*, 61 (3), 262-75.

Brock, Timothy C. (1968), "Implications of Commodity Theory for Value Change," *Psychological Foundations of Attitudes,* Academic Press, 243-75.

Campbell, Donald T., Nan A. Lewis, and William A. Hunt (1958), "Context Effects with Judgmental Language that is Absolute, Extensive, and Extra-Experimentally Anchored," *Journal of Experimental Psychology*, 55 (3), 220-28.

Chapman, Gretchen B. and Eric J. Johnson (2002), "Incorporating the Irrelevant: Anchors in Judgments of Belief and Value," *Heuristics and Biases: The Psychology of Intuitive Judgment*, 120-38.

Cialdini, Robert B. (2007), *Influence: The Psychology of Persuasion*, New York: HarperCollins.

Cunha Jr., Marcus and Jeffery D. Shulman (2011), "Assimilation and Contrast in Price Evaluations," *Journal of Consumer Research*, 37 (5), 822-35.

Dehaene, Stanislas, Lionel Naccache, Gurvan Le Clec'H, Etienne Koechlin, Michael Mueller, Ghislaine Dehaene-Lambertz, Pierre-Francois van de Moortele, and Denis Le Bihan (1998), "Imaging Unconscious Semantic Priming," *Nature*, 395 (6702), 597-600.

Dhar, Ravi, Stephen M. Nowlis, and Steven J. Sherman (1999), "Comparison Effects on Preference Construction," *Journal of Consumer Research*, 26 (3), 293-306.

Eliashberg, Jehoshua and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?" *Journal of Marketing*, 61 (2), 68-78.

Evangelidis, Ioannis and Stijn M. J. van Osselaer (2019), "Interattribute Evaluation Theory," *Journal of Experimental Psychology: General*, 148 (10), 1733–46.

Fernberger, Samuel W. (1920), "Interdependence of Judgments Within the Series for the Method of Constant Stimuli," *Journal of Experimental Psychology*, 3 (2), 126-50.

Fitzsimons, Gavan J. and Donald R. Lehmann (2004), "Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses," *Marketing Science*, 23 (1), 82–94.

Gershoff, Andrew, Ashesh Mukherjee, and Anirban Mukhopadhyay (2003), "Consumer Acceptance of Online Agent Advice: Extremity and Positivity Effects," *Journal of Consumer Psychology*, 13 (1&2), 161-170.

Gierl, Heribert and Verena Huettl (2010), "Are Scarce Products Always More Attractive? The Interaction of Different Types of Scarcity Signals with Products' Suitability for Conspicuous Consumption," *International Journal of Research in Marketing*, 27 (3), 225-35.

Gilchrist, Alan, Christos Kossyfidis, Frederick Bonato, Tiziano Agostini, Joseph Cataliotti, Xiaojun Li, Branka Spehar, Vidal Annan, and Elias Economou (1999), "An Anchoring Theory of Lightness Perception," *Psychological Review*, 106 (4), 795-834.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008), "A Room With a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research*, 35 (3), 472-82.

Hamilton, Rebecca, Debora Thompson, Sterling Bone, Lan Nguyen Chaplin, Vladas Griskevicius, Kelly Goldsmith, Ronald Hill, Deborah R. John, Chiraag Mittal, Thomas O'Guinn, Paul Piff, Anuj Shah, and Meng Zhu (2019), "The Effects of Scarcity on Consumer Decision Journeys," *Journal of the Academy of Marketing Science*, 47 (3), 532-50.

Heintz, Roy K. (1950), "The Effect of Remote Anchoring Points Upon the Judgment of Lifted Weights," *Journal of Experimental Psychology*, 40 (5), 584-91.

Hsee, Christopher K. (1996), "The Evaluability Hypothesis: An Explanation for Preference Reversals Between Joint and Separate Evaluations of Alternatives," *Organizational Behavior and Human Decision Processes*, 67 (3), 247-57.

Hsee, Chrisopher K., George F. Loewenstein, Sally Blount, and Max H. Bazerman (1999), "Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis," *Psychological Bulletin*, 125 (5), 576-90.

Hsee, Christopher K. and Jiao Zhang (2010), "General Evaluability Theory," *Perspectives on Psychological Science*, 5 (4), 343-55.

Kahneman, Daniel (1992), "Reference Points, Anchors, Norms, and Mixed Feelings," *Organizational Behavior and Human Decision Processes*, 51 (2), 296-312.

Lockhead, Gregory R. (2004), "Absolute Judgments Are Relative: A Reinterpretation of Some Psychophysical Ideas," *Review of General Psychology*, 8 (4), 265-72.

Lynn, Michael (1991), "Scarcity Effects on Value: A Quantitative Review of the Commodity Theory Literature," *Psychology & Marketing*, 8 (1), 43-57.

Lynn, Michael (1992), "The Psychology of Unavailability: Explaining Scarcity and Cost Effects on Value," *Basic and Applied Social Psychology*, 13 (1), 3-7.

Mussweiler, Thomas (2003), "Comparison Processes in Social Judgment: Mechanisms and Consequences," *Psychological Review*, 110 (3), 472-89.

Olson, James M., Richard D. Goffin, and Graeme A. Haynes (2007), "Relative Versus Absolute Measures of Explicit Attitudes: Implications for Predicting Diverse Attitude-Relevant Criteria," *Journal of Personality and Social Psychology*, 93 (6), 907-26.

Oppenheimer, Daniel. M., Robyn A. LeBoeuf, and Noel T. Brewer (2008), "Anchors Aweigh: A Demonstration of Cross-Modality Anchoring and Magnitude Priming," *Cognition*, 106 (1), 13-26.

Ordóñez, Lisa D., Terry Connolly, and Richard Coughlan (2000), "Multiple Reference Points in Satisfaction and Fairness Assessment," *Journal of Behavioral Decision Making*, 13 (3), 329-44.

Parker, Jeffrey R. and Donald R. Lehmann (2011), "When Shelf-Based Scarcity Impacts Consumer Preferences," *Journal of Retailing*, 87 (2), 142-55.

Roux, Caroline, Kelly Goldsmith, and Andrea Bonezzi (2015), "On the Psychology of Scarcity: When Reminders of Resource Scarcity Promote Selfish (and Generous) Behavior," *Journal of Consumer Research*, 42 (4), 615-31.

Schwarz, Norbert and Herbert Bless (1992), “Constructing Reality and Its Alternatives: An Inclusion/Exclusion Model of Assimilation and Contrast Effects in Social Judgment,” In L. L. Martin & A. Tesser (Eds.), *The Construction of Social Judgment,* 217-245, Hillsdale, NJ: Erlbaum.

Schweitzer, Maurice (1995), “Multiple Reference Points, Framing, and the Status Quo Bias in Health Care Financing Decisions,” *Organizational Behavior and Human Decision Processes*, 63 (1), 69-72.

Sevilla, Julio and Joseph P. Redden (2014), “Limited Availability Reduces the Rate of Satiation,” *Journal of Marketing Research*, 51 (2), 205-17.

Shah, Anuj K., Sendhil Mullainathan, and Eldar Shafir (2012), “Some Consequences of Having Too Little,” *Science,* 338 (6107), 682-85.

Shah, Anuj K., Eldar Shafir, and Sendhil Mullainathan (2015), “Scarcity Frames Value,” *Psychological Science,* 26 (4), 402-12.

Sherif, Mufazer, Daniel Taub, and Carl I. Hovland (1958), “Assimilation and Contrast Effects of Anchoring Stimuli on Judgments,” *Journal of Experimental Psychology*, 55 (2), 150-5.

Simonin, Bernard L. and Julie A. Ruth (1998), “Is a Company Known by the Company It Keeps? Assessing the Spillover Effects of Brand Alliances on Consumer Brand Attitudes,” *Journal of Marketing Research*, 35 (1), 30-42.

Simonson, Itamar (2008), “Will I Like a “Medium” Pillow? Another Look at Constructed and Inherent Preferences,” *Journal of Consumer Psychology*, 18 (3), 155-69.

Stewart, Neil, Gordon D. A. Brown, and Nick Chater (2005), “Absolute Identification by Relative Judgment,” *Psychological Review*, 112 (4), 881-911.

Sullivan, Kathryn, and Thomas Kida (1995), “The Effect of Multiple Reference Points and Prior Gains and Losses on Managers′ Risky Decision Making,” *Organizational Behavior and Human Decision Processes*, 64 (1), 76-83.

Van Herpen, Erica, Rik Pieters, and Marcel Zeelenberg (2009), “When Demand Accelerates Demand: Trailing the Bandwagon,” *Journal of Consumer Psychology*, 19 (3), 302-12.

Verhallen, Theo M. M. and Henry S. J. Robben (1994), “Scarcity and Preference: An Experiment on Unavailability and Product Evaluation,” *Journal of Economic Psychology*, 15 (2), 315-31.

Wever, Ernest Glen and Karl Edward Zener (1928), “The Method of Absolute Judgment in Psychophysics,” *Psychological Review*, 35 (6), 466-93.

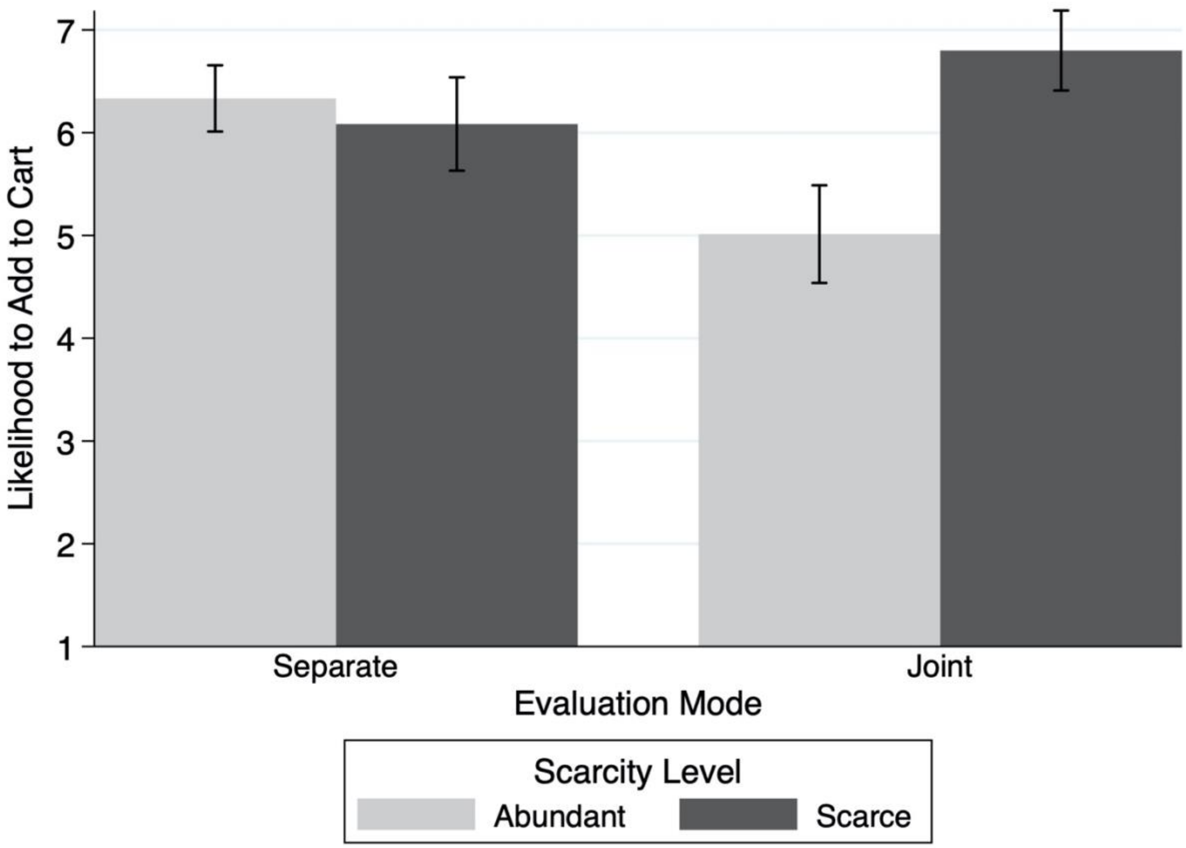
Worchel, Stephen, Jerry Lee, and Akanbi Adewole (1975), “Effects of Supply and Demand on Ratings of Object Value,” *Journal of Personality and Social Psychology*, 32 (5), 906-14.

Zhu, Meng and Rebecca K. Ratner (2015), “Scarcity Polarizes Preferences: The Impact on Choice Among Multiple Items in a Product Class,” *Journal of Marketing Research*, 52 (1), 13-26.

FIGURE 1

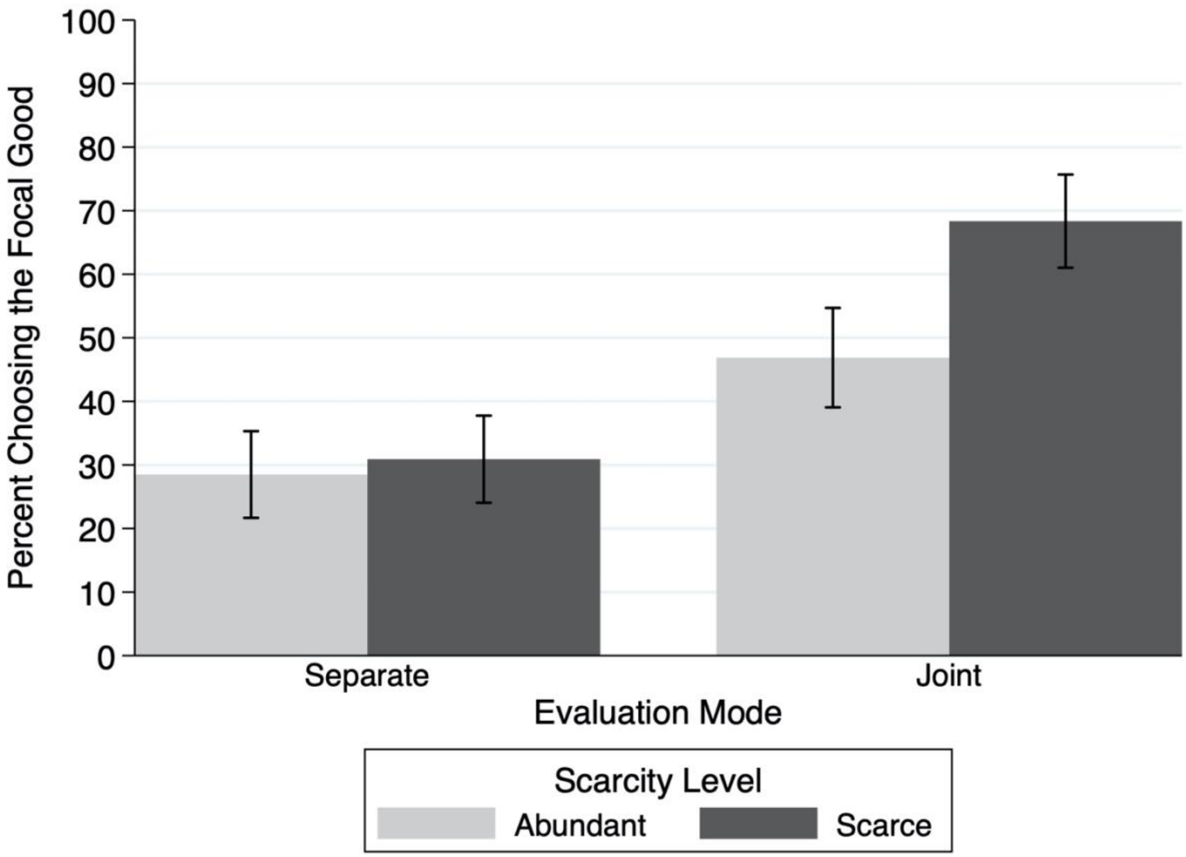
Note. Study 1 experimental stimuli. The top panel depicts the abundant good in joint evaluation mode, and the bottom panel depicts the abundant good in separate evaluation mode.

FIGURE 2



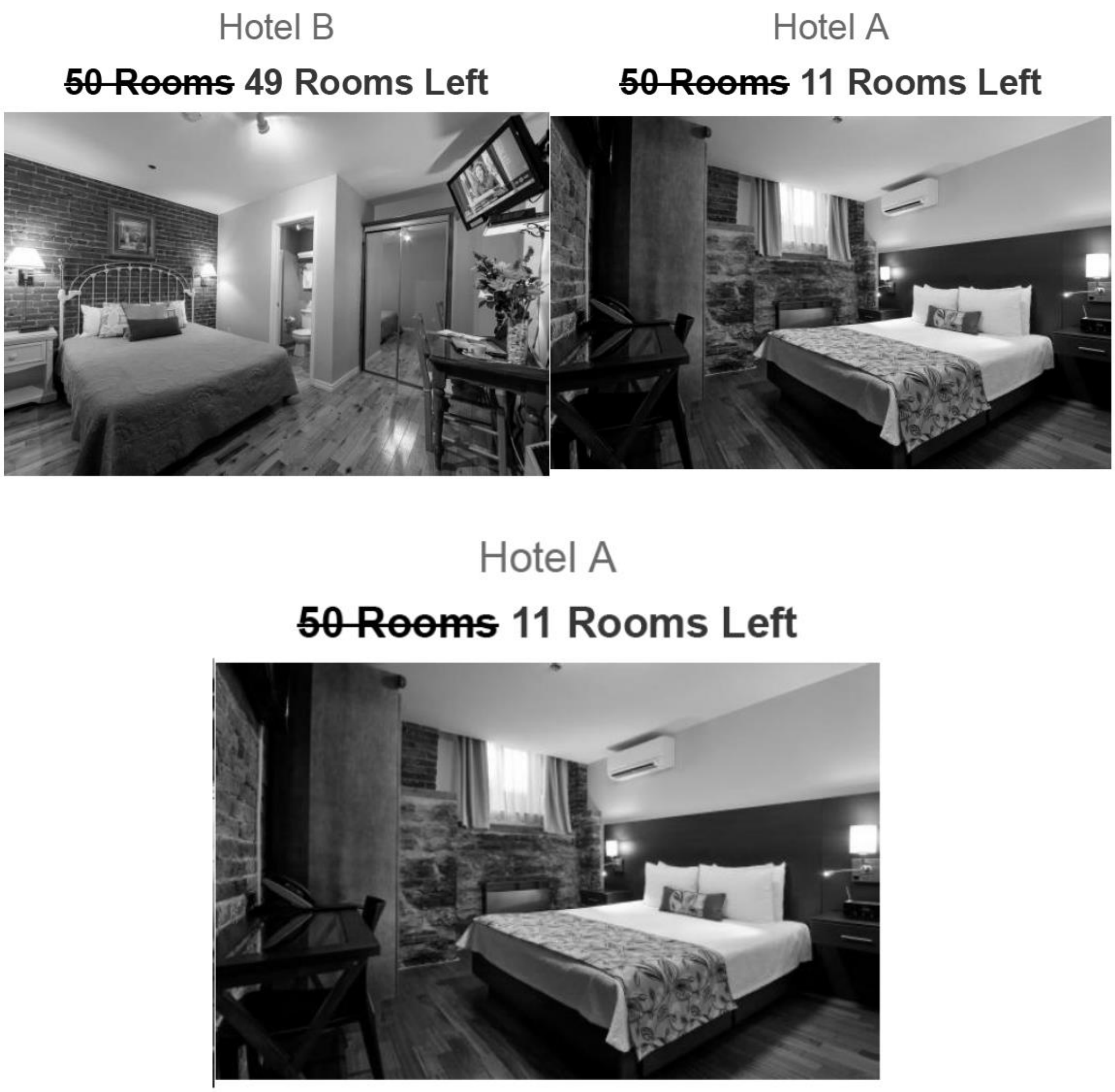
Note. Study 1 interaction of scarcity level and evaluation mode on likelihood to add the focal product to the cart. Error bars represent standard errors.

FIGURE 3



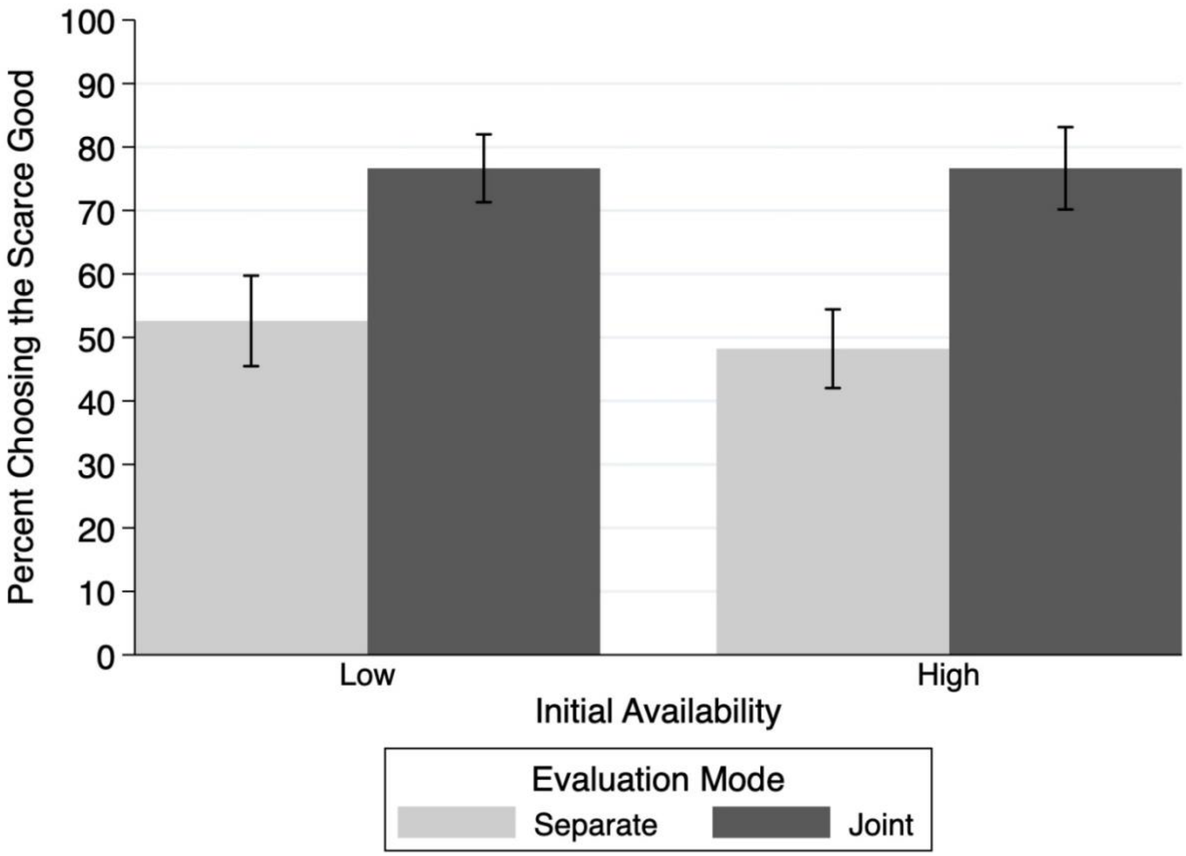
Note. Study 2a interaction of scarcity level and evaluation mode on the choice of the focal good. Error bars represent standard errors.

FIGURE 4



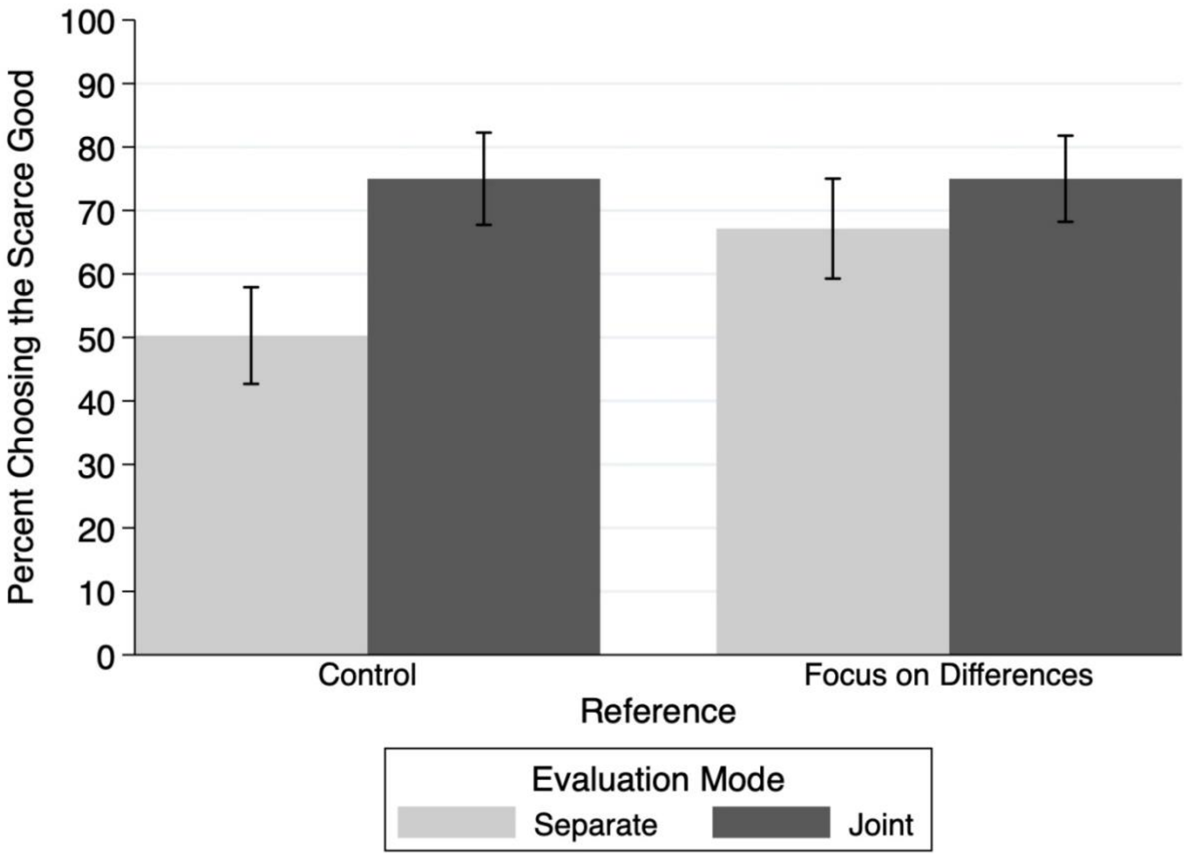
Note. Study 3 experimental stimuli. The top panel depicts the scarce good in joint evaluation mode, and the bottom panel depicts the scarce good in separate evaluation mode.

FIGURE 5



Note. Study 4 interaction of evaluation mode and initial availability on the percentage choosing the scarce good. Error bars represent standard errors.

FIGURE 6



Note. Study 5 interaction of reference and evaluation mode on the percentage choosing the scarce good. Error bars represent standard errors.

WEB APPENDIX:

"SCARCITY IS ALL RELATIVE: HOW THE VALUATION OF SCARCE GOODS DEPENDS ON THE PRESENCE OF ALTERNATIVES"

LINKS TO PRE-REGISTRATION FORMS

Study 1: https://aspredicted.org/blind.php?x=7qz4hh

Study 2a: https://aspredicted.org/blind.php?x=zb98xe

Study 2b: https://aspredicted.org/blind.php?x=55zw4h

Study 3: https://aspredicted.org/blind.php?x=wk6ak7

Study 4: https://aspredicted.org/blind.php?x=sy9ev8

Study 5: https://aspredicted.org/blind.php?x=8gc5k9

Study A1: https://aspredicted.org/blind.php?x=br3zn7

Study A2: https://aspredicted.org/blind.php?x=9f2g7a

ADDITIONAL ANALYSES

For completeness, we report the results for each study without using the full samples (i.e., without dropping participants who failed attention checks, as per our pre-registration). Additionally, we report the methods and results for two additional studies, Studies A1 and A2.

STUDY 1

The full sample includes 397 participants. We conducted a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) ANOVA on the likelihood of adding the focal wine to the shopping cart. The results revealed the predicted significant interaction of scarcity level and evaluation mode ($F(1, 393) = 14.35$, $p < .001$). There was a simple effect of scarcity level ($F(1, 395) = 12.49$, $p < .001$) and not a simple effect of evaluation mode ($F(1, 395) = 1.58$, $p = .209$). Directionally consistent with our prediction, the scarce good was valued to a greater extent in the presence of the abundant good than by itself ($M_{joint} = 6.71$, $SD = 1.89$ vs. $M_{separate} = 6.23$, $SD = 1.86$; $t(198) = 1.79$, $p = .075$). Conversely, the valuation of the abundant good was lower in the presence of the abundant good than by itself ($M_{joint} = 5.33$, $SD = 2.09$ vs. $M_{separate} = 6.28$, $SD = 1.46$; $t(195) = 3.57$, $p < .001$). We did not find an effect of scarcity level within the separate evaluation condition ($M_{scarce} = 6.23$ vs. $M_{abundant} = 6.28$, $t(186) = .19$, $p = .849$).

STUDIES 2A AND 2B

The full sample for Study 2a includes 801 participants. For both studies, we ran a logistic regression of choosing the focal wine on evaluation mode (1 = joint, 0 = separate), scarcity level (1 = scarce, 0 = abundant), and their interaction, using robust standard errors. In Study 2a, the coefficient on the interaction was not significant ($\beta_{Interaction} = .46$, $SE = .3$, $Z = 1.53$, $p = .125$). The scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (67% vs. 34%, $\chi^2(1) = 42.59$, $p < .001$). The abundant good was also incidentally chosen more frequently in joint evaluation mode than in separate evaluation mode (49% vs. 28%, $\chi^2(1) = 12$, $p = .001$). We also found a simple effect of evaluation mode ($\beta_{joint} = .89$, $SE = .21$, $Z =$

4.20, $p < .001$), and no simple effect of scarcity level ($\beta_{scarce} = .27$, SE = .21, Z = 1.23, $p = .218$; within separate evaluation mode: 34% vs. 28%, $\chi^2(1) = 1.52$, $p = .217$) on choosing the focal wine.

The full sample for Study 2b includes 807 participants. The coefficient on the interaction was significant ($\beta_{Interaction} = .70$, SE = .3, Z = 2.36, $p = .018$). The scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (54% vs. 28%, $\chi^2(1) = 28.64$, $p < .001$). The abundant good was also incidentally chosen more frequently in joint evaluation mode than in separate evaluation mode (42% vs. 31%, $\chi^2(1) = 4.19$, $p = .041$), but to a lesser extent than was the scarce good. We also found a simple effect of evaluation mode ($\beta_{joint} = .42$, SE = .21, Z = 2.04, $p = .041$), and no simple effect of scarcity level ($\beta_{scarce} = -.17$, SE = .22, Z = -.77, $p = .443$; within separate evaluation mode: 31% vs. 28%, $\chi^2(1) = .59$, $p = .443$) on choosing the focal wine.

STUDY 3

The full sample includes 801 participants. We ran a logistic regression of choosing the focal hotel on evaluation mode (1 = joint, 0 = separate), scarcity level (1 = scarce, 0 = abundant), and their interaction, using robust standard errors. Importantly, the coefficient on the interaction was significant ($\beta_{Interaction} = .82$, SE = .32, Z = 2.6, $p = .009$). Thus, an asymmetric effect of evaluation mode on choice arose such that the scarce good was chosen more frequently in joint evaluation mode than separate evaluation mode (84% vs. 63%, $\chi^2(1) = 22.84$, $p < .001$); the abundant good was not chosen less (or more) frequently in joint evaluation mode than in separate evaluation mode (62% vs. 54%, $\chi^2(1) = 2.1$, $p = .147$). Additionally, the analysis revealed no simple effect of evaluation mode (β_{joint} = .30, SE = .21, Z = 1.45, p = .148) and a (marginal) simple effect of scarcity level (β_{scarce} = .36, SE = .20, Z = 1.75, p = .080). We found a (marginal) effect of scarcity level on choice within the separate evaluation mode condition (63% vs. 54%, $\chi^2(1) = 3.07$, p = .080).

STUDY 4

The full sample includes 1,001 participants. We ran a logistic regression of choosing the scarce wine on evaluation mode (1 = joint, 0 = separate), initial availability level (1 = high, 0 = low), and their interaction, using robust standard errors. As predicted, the interaction was not significant ($\beta_{interaction}$ = .30, SE = .28, Z = 1.10, p = .271). Indeed, we found an effect of a joint evaluation mode (vs. separate evaluation mode) on choosing the scarce good both when initial quantities were low and high, and these effects were remarkably similar (low initial availability: 74% vs. 54%, $\chi^2(1) = 19.8$, $p < .001$; high initial availability: 75% vs. 49%, $\chi^2(1) = 34.55$, $p < .001$). We thus found a simple effect of evaluation mode (β_{joint} = .84, SE = .19, Z = 4.41, $p < .001$), though no simple effect of initial availability (β_{high} = -.23, SE = .18, Z = -1.26, p = .209).

STUDY 5

The full sample includes 701 participants. We ran a logistic regression of choosing the focal wine on evaluation mode (1 = joint, 0 = separate), reference (1 = difference, 0 = control), and their interaction, using robust standard errors. We found no significant interaction ($\beta_{interaction}$ = -.47, SE = .33, Z = -1.44, p = .150). As predicted, in the control, we find that the scarce good is chosen more frequently in joint evaluation mode than in separate evaluation mode (71% vs. 50%, $\chi^2(1) = 16$, $p < .001$). However, when participants focused on the differences in scarcity levels between the two goods, the effect is (marginally) attenuated (75% vs. 67%, $\chi^2(1) = 3$, $p = .084$). Additionally, the analysis revealed simple effects of evaluation mode ($\beta_{joint} = .89$, SE = .22, Z = 3.96, $p < .001$) and of reference ($\beta_{difference} = .68$, SE = .22, Z = 3.08, $p = .002$).

STUDY A1

Study A1 extended the results of Study 1 with the inclusion of the initial availability as a reference point.

Method

We recruited 398 adults (54% female, $M_{age} = 37.3$, $SD = 12.1$) via Amazon Mechanical Turk to complete the survey in exchange for $.20. As pre-registered, we dropped 45 participants who failed the attention check and 4 participants who took the study more than once, leaving 349 remaining participants for subsequent analyses.

This experiment followed a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) between-subjects design. The methods were identical to Study 1, with two exceptions. First, to control for inferences participants might make about retailers' strategic stocking decisions, they were also informed that the wine website stocks 100 bottles of each wine at the beginning of each month. This additionally served as the initial reference point and created the implicit, evaluable scale from 0 to 100 bottles available. Second, scarcity levels were 2 and 75 (both out of 100).

Results

We conducted a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) ANOVA on the likelihood of adding the focal wine to the shopping cart. The results revealed the predicted significant interaction of scarcity level and evaluation mode ($F(1, 345) = 8.95$, $p = .003$). There was a simple effect of scarcity level ($F(1, 347) = 36.35$, $p < .001$) and not a simple effect of evaluation mode ($F(1, 347) = 2.54$, $p = .112$). Consistent with our prediction, the scarce good was valued to a greater extent in the presence of the abundant good than by itself ($M_{joint} = 6.59$, SD = 1.95 vs. $M_{separate} = 5.64$, SD = 1.80; $t(176) = 3.36$, $p < .001$). Conversely, the valuation of the abundant good was not lower in the presence of the scarcer good than by itself ($M_{joint} = 4.72$, SD = 1.96 vs. $M_{separate} = 5.01$, SD = 2.00; $t(169) = 0.95$, $p = .342$). We did find an effect of scarcity level within the separate evaluation condition ($M_{scarce} = 5.64$ vs. $M_{abundant} = 5.01$, $t(177) = 2.24$, $p = .026$).

Although not pre-registered, we find similar results when including those who failed the attention checks (N = 394). We conducted a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) ANOVA on the likelihood of adding the focal wine to the shopping cart. The results revealed the predicted significant interaction of scarcity level and evaluation mode ($F(1, 390) = 11.49$, $p < .001$). There was a simple effect of scarcity level ($F(1, 392) = 26.99$, $p < .001$) and not a simple effect of evaluation mode ($F(1, 392) = 1.56$, $p = .213$). Consistent with our prediction, the scarce good was valued to a greater extent in the presence of the abundant good than by itself ($M_{joint} = 6.52$, SD = 1.97 vs. $M_{separate} = 5.61$, SD = 1.84; $t(195) = 3.35$, $p = .001$). Conversely, the valuation of the abundant good was not lower in the presence of the scarcer good than by itself ($M_{joint} = 4.83$, SD = 1.97 vs. $M_{separate} = 5.25$, SD = 2.03; $t(195) = 1.48$, $p = .140$). We did not find an effect of scarcity level within the separate evaluation condition ($M_{scarce} = 5.61$ vs. $M_{abundant} = 5.25$, $t(193) = 1.28$, $p = .202$).

STUDY A2

Study A2 extended the results of Study A1 to a choice-based paradigm.

Method

We recruited 701 adults (52.6% female, M_{age} = 36.8, SD = 11.8) via Amazon Mechanical Turk to complete the survey in exchange for $.20. As pre-registered, we dropped 95 participants who failed the attention check, leaving 606 remaining participants for subsequent analyses.

This experiment followed a 2 (scarcity level of focal good: scarce, abundant) x 2 (evaluation mode: separate, joint) between-subjects design. The methods were identical to that of Study A1 with two exceptions. First, instead of using likelihood to add to an online shopping cart, we asked about the choice of the focal product. Specifically, participants were asked if they would purchase the focal product or keep searching for a different bottle. Second, scarcity levels were 5 and 85 (both out of 100).

Results

We ran a logistic regression of choosing the focal wine on evaluation mode (1 = joint, 0 = separate), scarcity level (1 = scarce, 0 = abundant), and their interaction, using robust standard errors. The coefficient on the interaction was significant ($\beta_{interaction}$ = 1.13, SE = .35, Z = 3.25, p = .001). The scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (78% vs. 57%, $\chi^2(1) = 15.04$, $p < .001$). However, the abundant good was not chosen less (or more) frequently in joint evaluation mode than in separate evaluation mode (37% vs. 41%, $\chi^2(1) = 0.49$, $p = .486$). We also found no simple effect of evaluation mode (β_{joint} = -.17, SE = .24, Z = -0.70, p = .487), and a simple effect of scarcity level (β_{scarce} = .65, SE = .23, Z =

2.89, p = .004; within separate evaluation mode: 57% vs. 41%, $\chi^2(1) = 8.42$, p = .004) on choosing the focal wine.

Although not pre-registered, we find similar results when including those who failed the attention checks (N = 701). We ran a logistic regression of choosing the focal wine on evaluation mode (1 = joint, 0 = separate), scarcity level (1 = scarce, 0 = abundant), and their interaction, using robust standard errors. The coefficient on the interaction was significant ($\beta_{interaction}$ = .74, SE = .32, Z = 2.33, p = .020). The scarce good was chosen more frequently in joint evaluation mode than in separate evaluation mode (74% vs. 59%, $\chi^2(1) = 9.67$, p = .002). However, the abundant good was not chosen less (or more) frequently in joint evaluation mode than in separate evaluation mode (43% vs. 43%, $\chi^2(1) = 0.02$, p = .889). We also found no simple effect of evaluation mode (β_{joint} = -.03, SE = .22, Z = -0.14, p = .889), and a simple effect of scarcity level (β_{scarce} = .62, SE = .22, Z = 2.87, p = .004; within separate evaluation mode: 59% vs. 43%, $\chi^2(1)$ = 8.31, p = .002) on choosing the focal wine.