1

2

3

4   Context-dependent choice and evaluation in real-world consumer behavior

5

6

7   A. Ross Otto[1*], Sean Devine[1], Eric Schulz[2], Aaron M. Bornstein[3†], & Kenway Louie[4,5†]

8

9   [1]Department of Psychology, McGill University, Montréal, Canada

10   [2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany

11   [3]Department of Cognitive Sciences and Center for the Neurobiology of Learning and Memory,

12   University of California, Irvine, United States

13   [4]Center for Neural Science, New York University, New York, United States

14   [5]Neuroscience Institute, New York University Grossman School of Medicine, New York, United

15   States

16

17   *Corresponding author: ross.otto@mcgill.ca

18   †These authors contributed equally to this submission

19

1   **Abstract**

2   A body of work spanning neuroscience, economics, and psychology indicates that decision-

3   making is context-dependent, which means that the value of an option depends not only on the

4   option in question, but also on the other options in the choice set—or the 'context'. While

5   context effects have been observed primarily in small-scale laboratory studies with tightly

6   constrained, artificially constructed choice sets, it remains to be determined whether these

7   context effects take hold in real-world choice problems, where choice sets are large and

8   decisions driven by rich histories of direct experience. Here, we investigate whether valuations

9   are context-dependent in real-world choice by analyzing a large restaurant rating dataset

10   (Yelp.com) as well as two independent replication datasets which provide complementary

11   operationalizations of restaurant choice. We find that users make fewer ratings-maximizing

12   choices in choice sets with higher-rated options—a hallmark of context-dependent choice— and

13   that post-choice restaurant ratings also varied systematically with the ratings of unchosen

14   restaurants. Furthermore, in a follow-up laboratory experiment using hypothetical choice sets

15   matched to the real-world data, we find further support for the idea that subjective valuations of

16   restaurants are scaled in accordance with the choice context, providing corroborating evidence

17   for a general mechanistic-level account of these effects. Taken together, our results provide a

18   potent demonstration of context-dependent choice in real-world choice settings, manifesting both

19   in decisions and subjective valuation of options.

20

21

2

Introduction

A central theme in the study of decision-making is the limited-capacity nature of human information processing (1). A classic example of choice behavior shaped jointly by presumed processing limitations and the demands of the decision environment is the context-dependent nature of preferences: the relative value of an option depends not only on the option in question but the other options in the choice set, or context (2–5). This sort of context-dependent valuation explains a number of interesting patterns of choice both in humans (6–10) and animals (11, 12).

Context-dependent—rather than stable—valuation famously challenges classical theories of rational choice (13) which posit that value is a static quantity independent of the context, and accordingly, preference between two options should be independent of contextual factors such as the other available options. At the same time, relative (versus absolute) valuation could be understood, computationally, as a rescaling of reward values, relative to the other values present in the context, allowing an individual's representation of value to adjust to the current value context (6). Indeed, recent theoretical treatments have argued that contextual—versus absolute— value representations permit a decision-maker to discriminate between two low-value options in one context, and at the same time, two high-valued options in another context (14, 15).

Context effects have been observed in both real-world choices—with respect to the temporal history of information viewed, or past choices faced by decision-makers (16, 17)—and in tightly constrained, small-sample demonstrations of "independence of irrelevant alternatives" (IIA) violations (18, 19). While a body of work has examined context effects arising from choice set composition—probing how the distribution of option values faced by the decision-maker systematically alter choices and subjective valuations of options—limited work has investigated whether these sorts of context effects hold in real-world settings (20). Unlike tightly-controlled

1 laboratory settings, the choice sets facing real-world decision-makers are arbitrarily large, and

2 composed of options which could be valued along multiple dimensions, each learned via rich

3 histories of direct experience. Understanding how these real-world context effects unfold is of

4 clear practical importance, as well as for understanding the basic mechanisms of context effects

5 more generally. For example, it is possible that these sorts of context effects do not scale to more

6 naturalistic settings, where individuals may resort to heuristics (e.g. 'taking the best').

7 Supporting this view, more naturalistic presentation of options has been shown to reduce

8 apparent suboptimalities in choice (21).

9   More generally, our understanding of context effects in choice might be inherently

10 limited by the artificial, but well-constrained nature of tasks employed in these laboratory choice

11 studies (22–25). Accordingly, we reasoned that we could leverage large-scale datasets that reflect

12 real-world choices in diverse settings in order to glean novel insights about how context effects

13 play out 'in the wild' (26, 27), thus directly evaluating whether these effects are meaningful and

14 general (28). Here we elucidate whether context effects could be observed in actual choices with

15 real-world consequences, for both the decision-maker and the restaurant, at a scale and breadth

16 sufficient to permit detection of context effects while averaging over potential option- and

17 subject-specific effects.

18   We demonstrate the generality of context-dependent valuation—both in choices and self-

19 reported valuation—by analyzing a massive real-world restaurant rating dataset. Here we

20 examine user behavior on Yelp, a web-based application which maintains crowd-sourced

21 reviews of local businesses which is particularly popular in North America. Our analysis

22 specifically focused on restaurants because, in most cases, people visit one restaurant per meal—

23 ensuring that options effectively compete with each other— and the categories of restaurants are

1   relatively well-defined in our dataset. With the presumption that each Yelp user's rating

2   evidences both a decision made to visit the rated restaurant over some competing set of options

3   (29) and the user's subjective valuation of their consummatory experience (30), this rating data

4   permits large-scale examination of contextual effects on both inferred choices and valuations. In

5   our analysis, we view each restaurant decision as one made from a "choice set"—defined as

6   similar, geographically proximate restaurants—with the goal of choosing the restaurant with the

7   highest utility (2, 31–33).

8   Our large-scale real-world analysis of ratings addresses two empirical questions. First,

9   what is the impact of the local value context upon choice efficiency—namely, do Yelp users

10   make fewer ratings-maximizing choices in sets with higher overall ratings? Based on laboratory-

11   based results indicating that choice contexts increasing the value of 'distractor' items (i.e.,

12   options lower in value than the higher-rated options in the set) decreased individuals' propensity

13   to choose the highest-valued item (22, 23) we hypothesized that we would see fewer choices of

14   the highest-rated option in higher-valued contexts (see (22, 24, 34) for discussion of other

15   possibly coexisting effects). One mechanism underpinning such a choice context effect would be

16   if options were selected on the basis of expected values, which are themselves context

17   dependent. To shed light on these intermediate computations we performed a second, laboratory-

18   based study, testing the hypothesis that if decision-makers' subjective valuations are context

19   dependent, then their ratings of the expected quality of a given restaurant should also exhibit

20   context-dependence in accordance with the value distribution of the choice set, even in the

21   absence of explicit choice. Taken together, we find evidence for context-dependent decision-

22   making and valuation in real-world behavior (manifesting both in real-world restaurant choices

5

1   and ratings) and furthermore, laboratory-based evidence suggesting that these context effects in

2   choice and subjective valuation are closely related.

3   **Results**

4   We examined signatures of context-dependent choice and subjective valuations using a

5   dataset from Yelp, a website and smartphone app popular in North America, on which users

6   review local businesses such as restaurants and retail stores. Each rating is made on a scale of

7   one to five stars, with higher star ratings indicating a better rating. For every Yelp user

8   considered, we analyzed each user-provided restaurant rating—which, following previous work

9   (29, 31), we took as a proxy for a user's choice to visit that restaurant at the time of the review—

10   with respect to the competing restaurant options in that neighborhood (i.e., the choice set).

11   This choice set was defined as all restaurants belonging to the chosen restaurant's

12   category within the same neighborhood (e.g. "Pizza" in a particular neighborhood in Pittsburgh).

13   These geographic neighborhoods were defined by a density-based spatial clustering algorithm

14   (35, 36), taking all user reviews as input (Figure 1A; see Methods). Thus, the 8 cities considered

15   contained multiple geographically-defined choice sets for each restaurant category which varied

16   considerably in their rating distributions (Figures 1B and C). Taking this approach, we can

17   analyze each user's choices (N=53,118 choices) in the context of similar restaurants that were

18   geographically proximate (37, 38). Critically, this analysis considered each option's average

19   rating at the time of the user's presumed choice—rounded to the nearest half-star, as is presented

20   to users on Yelp— in order to capture the decision-maker's information state at the time of

21   restaurant selection.
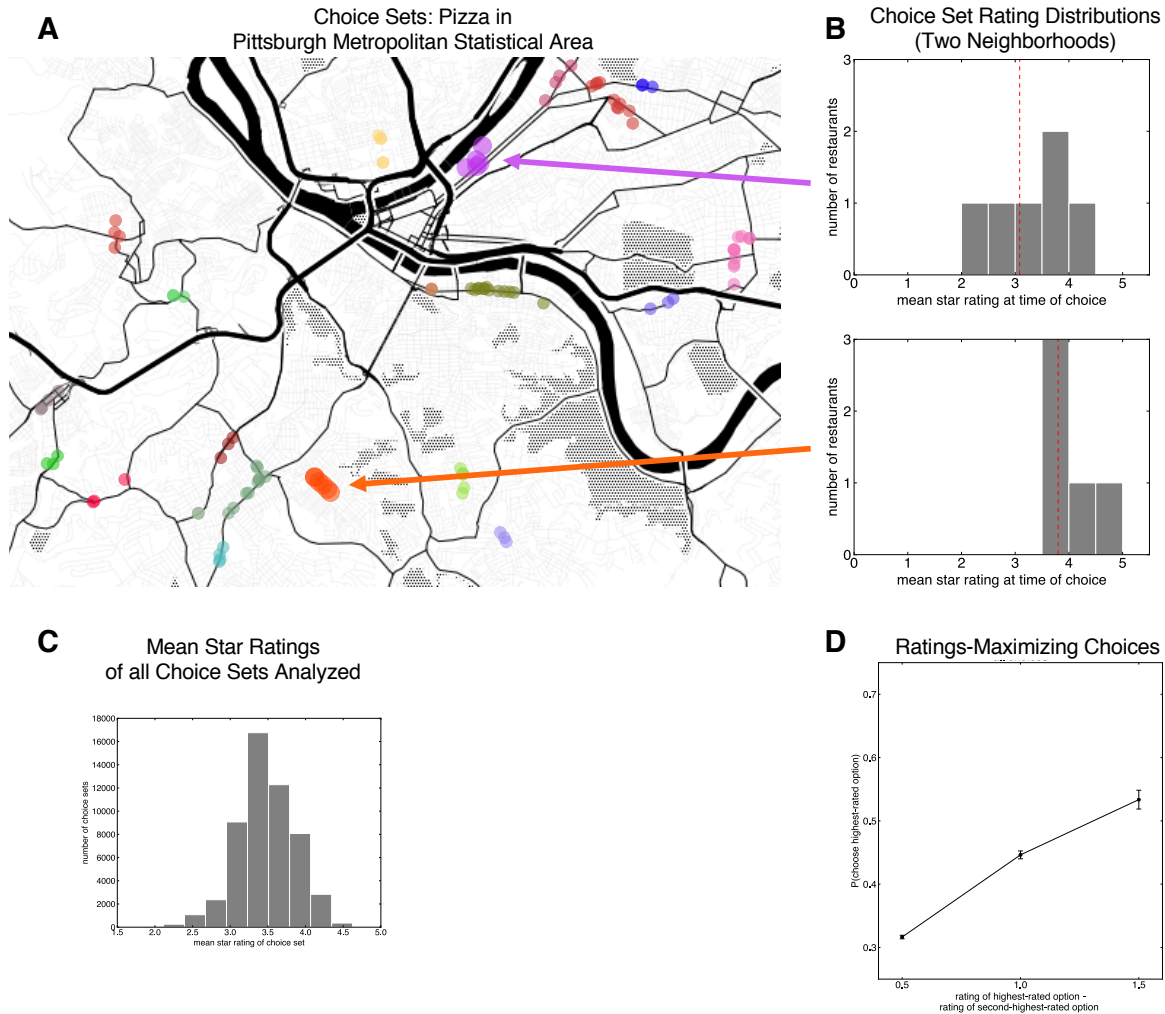
Figure 1. (A) Example of spatially-defined choice sets for Pizza restaurants in Pittsburgh, as computed by the density-based clustering algorithm. Each choice set is denoted by a unique color. (B) Distributions of option ratings for two example choice sets in the Yelp dataset, which correspond to the orange and purple clusters on the map. Notably, the mean star rating (red dashed line) is different across these two choices sets. (C) Distribution of the mean star rating of the choice set, across all choice sets analyzed (D) Proportion of ratings-maximizing (or 'target') choices as a function of the difference in ratings between the highest-rated and second-highest-rated option, demonstrating that users were sensitive to the sensitive to option' ratings.

We first sought to establish that users' choices were at all sensitive to the options' ratings.

Consistent with the idea that our operationalization of choice and definition of choice sets

capture a value-comparison process, the likelihood that a user chooses the highest-rated option

(e.g. a 4.5-star option) increased as a function of the difference between the highest-rated option

1   and the second-highest-rated option (e.g., a 4-star option) in the choice set (Figure 1D; logistic

2   regression β= 0.5182, 95% CI: 0.4145-0.6149, $p<0.0001$; Supplementary File 1B).

3   We next examined whether the value context—i.e. the star rating composition of the

4   choice set—affected individuals' propensity to choose the highest-rated restaurant. We found

5   that as the mean star rating of the choice set increased, choices of the highest-rated option (i.e.,

6   target) decreased markedly (Figure 2A, logistic regression β=-0.0846, 95% CI: -0.1425, -0.0181,

7    $p=0.002$; Table 1). In other words, the likelihood that a user chose the best-rated "target"

8   depended strongly upon the rating composition of the choice set, with the target choice becoming

9   less likely as the quality of the other options in the set increased, demonstrating a pronounced

10   context effect (5, 23). Intuitively, higher-rated targets were more likely to be chosen (29), and in

11   choice sets with more options, the choice of the highest-rated option was less likely (log number

12   of options β=-1.0066, 95% CI: -1.0776, -0.9454, $p<0.0001$), reflecting the well-documented set

13   size effect (34, 39). Importantly, our analysis controlled for the number of reviews (40) and the

14   price of the highest-rated option (31) by taking them as covariates in the regression. Furthermore,

15   this context effect—that, the negative effect of mean choice set star rating upon ratings-

16   maximizing choice—was statistically significant even when controlling for the variance of

17   choice set star ratings (which itself exerted a significant positive effect upon ratings-maximizing

18   choice; β= 0.5727, 95% CI: 0.4926, 0.6557, $p<0.0001$; Table 1).
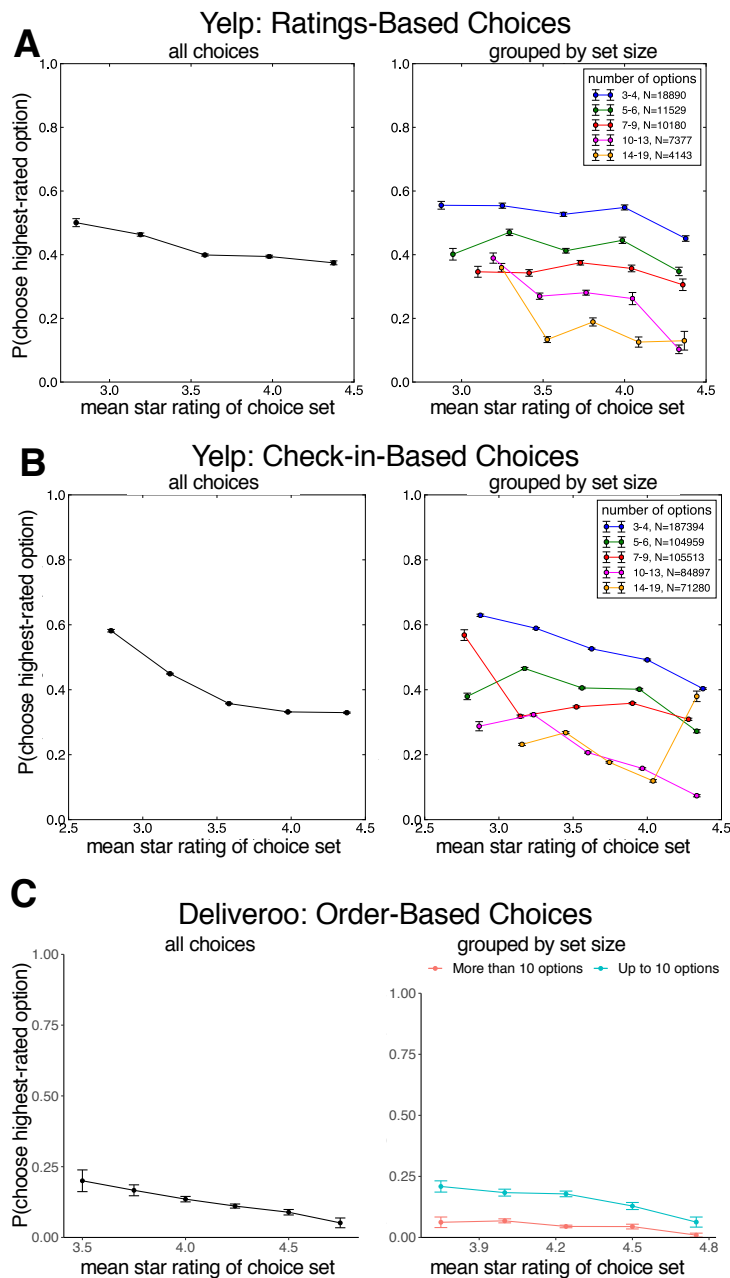
Figure 2. (A) Proportion of choices to the highest-rated option, inferred from Yelp ratings, as a function of the mean star rating of the choice set, for all choice sets (left panel) and broken down by choice set size (right panel). (B) Proportion of choices to the highest-rated option, inferred from Yelp Check-ins, as a function of the mean star rating of the choice set, for all choice sets (left panel) and broken down by choice set size (right panel). (C) Proportion of choices to the highest-rated option, determined from food orders on Deliveroo, as a function of the mean star rating of the choice set, for all choice sets (left panel) and broken down by choice set size (right panel). Across all three datasets, we observed a marked context effect such that participants became less likely to make ratings-maximizing choices as the mean rating of the choice set increased.

9

1   As we observed a modest correlation between the difference between the top two highest-

2   rated options and the mean star rating of the choice set (Spearman ρ=-0.1717, p<0.0001), we

3   sought to rule out the possibility that the apparent suboptimality in making ratings-maximizing

4   choices in sets with higher average option ratings (Figure 2A) simply arose from highly-rated

5   second-best options. To do this, we repeated the regression, taking the difference between the

6   ratings of the highest and second-highest options—demonstrated above to predict choice

7   accuracy (Figure 1D)—as a covariate, finding that the effect of mean star rating of choice set

8   (i.e., context) remained statistically significant (β= -0.1175, 95% CI: -0.186, -0.0611, p<0.001).

9   Further, to address the possibility that the apparent difficulty in making value-maximizing

10   choices in choice sets with higher average option ratings (e.g. Figure 2B) stems from confusion

11   in discriminating the best option, we repeated the analysis of choice as a function of mean choice

12   set rating, only considering choice sets where the second-best option is at least 1 star away from

13   the highest-rated option (24% of choice sets; for example, a set with a highest-rated option of 4.5

14   stars where the mean rating of the next-best option is 3.5 or less). We found that the contextual

15   effect of mean choice set rating upon ratings-maximizing "target" option selection rates was

16   even stronger in this restricted dataset (β= -0.9001, SE=0.065, p<0.0001) suggesting against the

17   possibility that these context effects arise from 'crowding' of options at the higher end of the

18   rating scale. Finally, to address the possibility that high-rated targets exert undue influence on

19   our computed mean choice set rating, we tested if target choices could be predicted as a function

20   of the mean choice set rating without the highest-rated (target) option, finding that the same

21   context effects take hold (β= -0.284, SE=0.0346, p<0.0001).

22   Importantly, this analysis takes a user's rating as a proxy for their choice to visit a

23   restaurant, but we sought to demonstrate that these choice context effects take hold in two

10

1 independent replication datasets which operationalize choice differently. First, we examined a

2 larger dataset of 'Check-ins' on Yelp, which are time-stamped, geolocated records of a user

3 having visited a particular restaurant, using the same choice set definition as the previous

4 analysis (N= 565,661 choices; see Methods). Examining users' propensity to 'Check into' the

5 highest-rated option as a function of the mean star rating of the choice set (Figure 2B), we found

6 a similar and pronounced context effect: users were significantly less likely to make choices to

7 the highest-rated 'target' option as the quality of the other options in the set increased (β=

8 0.2895, 95% CI:0.223, 0.3672, $p<0.000$; Supplementary File 1C). We also found that, unlike the

9 ratings-based choice examined above, the variance of the star ratings distribution exerted no

10 discernable effect upon ratings-maximizing choices (β=-0.0169, 95% CI: -0.0771, 0.0378,

11 p=0.29), suggesting that users' tendency to make poor choices in choice sets with high mean star

12 ratings purely from a preponderance of higher-valued options in these sets.

13 Following previous laboratory-based work (23, 24), we performed an additional analysis

14 of choices in the Yelp datasets (inferred from Ratings and Check-ins), examining the ratio of

15 choices made between the two top-rated options as a function of the quality of the "distractor"

16 items inferior to the two highest-rated options. This analysis provides a more direct examination

17 of possible violations of independence of irrelevant alternatives (IIA), according to which the

18 choice ratio between two options should not depend on the quality of other "irrelevant" options

19 (41). Importantly, this analysis only considered the subset of choices made to the highest- and

20 second-highest-rated options in each choice set (75% and 72% of choices in the ratings-based

21 and Check-ins-based choice datasets, respectively), probing whether the relative ratio of choices

22 made to the best option varied as a function of the average ratings of the distractor items rated

23 worse than the best two options. Conceptually replicating Louie et al. (23) and echoing the

11

1   results of our prior analysis, we observed that choice sets with more valuable inferior options led

2   to significant decreases in the relative ratio of ratings-maximizing choices in both ratings-based

3   choices (β= -0.482, 95% CI: -0.5566, -0.4, p<0.0001; Figure 2 - Figure supplement 1;

4   Supplementary File 1D) and in Check-in based choices (β= -0.497, 95% CI: -0.5131, -0.4802,

5   p<0.0001; Figure 2 - Figure supplement 2; Supplementary File 1E). In other words, we observed

6   that users' ability to choose between high-quality restaurants was systematically diminished by

7   the quality of putatively irrelevant, lower-quality options, a hallmark of context-dependent

8   decision-making.

9   Further, we examined whether context effects could be observed in actual purchase

10   decisions from the online food delivery service Deliveroo (N=1,613,968 orders; see Methods), a

11   dataset recently used to understand exploratory behavior in consumer choice (31). As in the two

12   other Yelp datasets, we again found a marked context effect (Figure 2C): users became less

13   likely to order from the highest-rated restaurant in a choice set as the average rating of the choice

14   set increased (β=-0.01777, 95% CI:-0.0256, -0.0099, p<0.001; Supplementary File 1F) but again,

15   variance of star ratings exhibited no significant effect upon choice (β= -0.0015, 95% CI: -

16   0.00920, 0.00627, p=0.711). Taken together, these two independent replication datasets support

17   the generality of these context effects of across different settings and operationalizations of

18   choice.

19   Interestingly, the Yelp ratings examined here—which reflect a user's consumption

20   experience with a particular restaurant—also afford the opportunity to probe how choice context

21   modulates a user's subjective valuation of an option following a presumed choice. Intuitively, a

22   user's restaurant ratings correlated with the average star ratings left by other users before the

23   current rating was made (r=0.175, p<0.0001), indicating that higher-rated restaurants were more

12

1   likely to receive higher star ratings after being chosen. Taking this dependency into

2   consideration, we focused on predicting the deviation between the user's rating and the average

3   restaurant rating at the presumed time of choice (Figure 3A). In other words, we asked whether

4   context had a reliable influence on how a user's ratings differed from the average ratings, which

5   were left prior that user's rating. We observed that higher average option ratings of the choice

6   set—i.e., higher-valued contexts—were associated with larger rating deviations (Figure 3B; $\beta=$

7   0.0576, 95% CI: 0.037, 0.0753, $p<0.0001$). Importantly, our regression model controlled for a

8   number of variables including the number of previous reviews, the price of the chosen option,

9   and average previous rating of the chosen restaurant (Table 2). At first blush, the directionality of

10   this context effect appears at odds with the context effects observed on choice (Figures 2A-C)

11   where more valuable (or, higher-rated) contexts reduce overall choice efficiency. Following

12   influential accounts of context effects on choice (3, 19, 22, 42), which propose that the subjective

13   value of an option is scaled relative to the values of set of options faced, one might expect a

14   decrease in rating deviation with more valuable (higher-rated) contexts.

15

16
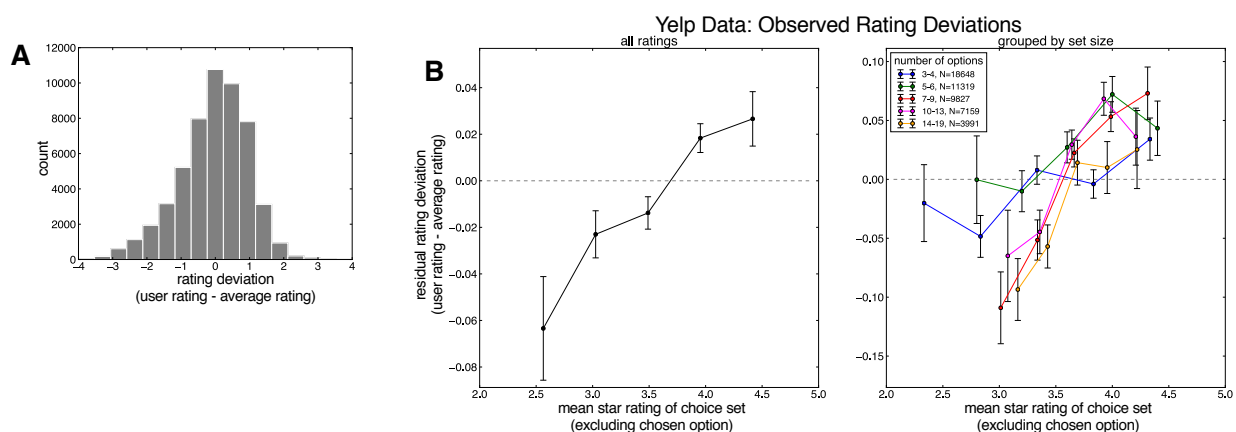
17
18   **Figure 3.** (A) Distribution of deviations between the user's rating and the average rating of the
19   chosen option at the time of choice for all ratings analyzed. (B) Deviation between the user's rating

1   and the average rating of the chosen option as a function of the mean star rating of the choice set

2   for all choice sets (left panel) and as a function of choice set size (right panel), after controlling for

3   a number of nuisance variables (see main text). We observed a marked context effect such that

4   higher-rated choice sets resulted in a more positive rating deviation.

5

6   However, it is possible to view ratings not as a pure reflection of the user's consumption

7   utility of the restaurant—i.e., how good or bad the restaurant was, in absolute terms—but instead

8   as the outcome of a comparison with the user's contextually bound expectations about the

9   restaurant's quality (43, 44) before making a rating (and presumably before visiting the

10   restaurant). More formally, we propose that the subjective valuation of an option, reflected in a

11   user's rating of that option, is computed as the difference between the experienced utility of the

12   restaurant and the contextually-computed expectation of that option (akin to a "prediction error"

13   (45, 46)). On this view, choice contexts with higher average ratings drive should reduce the

14   user's expectation of an option, in proportion to the average rating of the choice set. Assuming

15   that the user's experienced utility for an option is a symmetric random variable centered around

16   that option's ground-truth utility (here, its average rating before the user's rating is made), an

17   option rated in a higher-valued context should receive higher ratings relative to the options'

18   ground-truth value. Indeed, this exact pattern manifests in the rating deviations (Figure 3B):

19   more valuable choice sets increase the deviation between a user's subjective valuation of an

20   option and the average rating of an option.

21   However, a limitation inherent to this real-world rating data is that only the user's final

22   rating is observable, but not the user's contextually computed expectation of a restaurant's

23   quality, which is a key component in this presumed comparison process. To demonstrate more

24   conclusively that the context effects observed in real-world ratings stem from contextually-

25   bound expectations, we conducted a laboratory experiment in which participants (N=62) viewed

26   choice sets drawn from the real-world choice sets designed to emulate aspects of the choice and

14

1   rating experience faced by Yelp users (Figure 4A; see Methods). In half of the 270 trials,

2   participants made hypothetical restaurant choices in a given choice set (e.g., pizza in a particular

3   neighborhood), and in the other half of trials, participants provided their expectation of the

4   quality of a randomly selected restaurant on the basis of its restaurant's rating and the choice set.
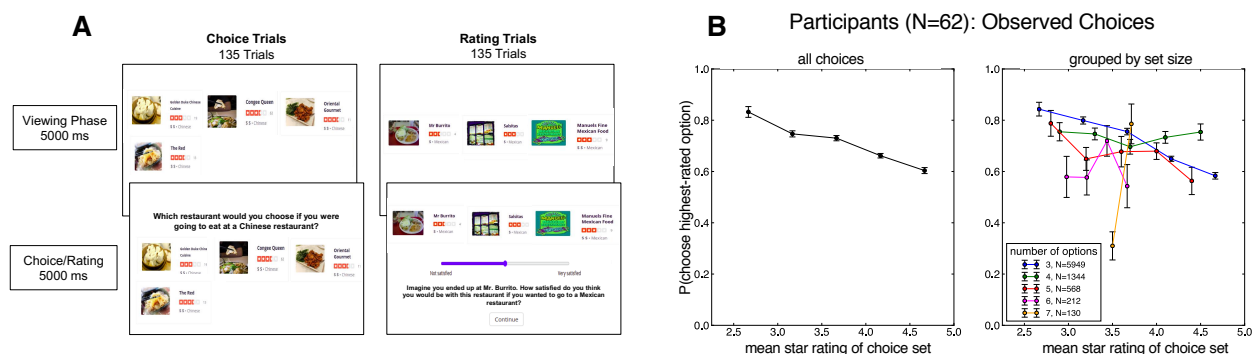

5
6
7   **Figure 4.** (A) Task screenshots for laboratory experiment. On each trial, participants views a
8   choice set consisting of 3-7 options from a single restaurant category, and either chooses a
9   restaurant (Choice trials; left panel) or indicates their expected satisfaction of a randomly selected
10   restaurant (Rating trials; right panel). (B) Proportion of choices of the highest-rated option as a
11   function of the mean star rating of the choice set, for all choice sets (left panel) and broken down
12   by choice set size (right panel). We observed a marked context effect in the Experiment, that
13   participants were less likely to make a ratings-maximizing choice in sets with higher mean star
14   ratings.
15

16   Analyzing choices, we observed participants were less significantly likely to choose the

17   highest-rated 'target' option in choice sets with higher mean ratings (Figure 4B; β= -0.8538, 95%

18   CI: -1.321, -0.4892, *p*<0.0001; Supplementary File 1G), but again, the variance of the star ratings

19   of the choice set exerted no appreciable influence on ratings-maximizing choices (β= -0.8538,

20   95% CI: -1.321, -0.4892, *p*<0.0001). In other words, our laboratory experiment replicated the

21   choice context effected observed in the real-world choice behavior—more valuable contexts

22   were associated with less likely choices of the highest-rated option. Following our analysis of

23   Yelp choices, we also probed for IIA violations in the 88% of choices made to the two best

24   options in each choice set in the experiment, finding that—as we observed in the Yelp datasets—

1   the average value of inferior distractor items significantly predicted participants' tendency to

2   choose the higher-valued of the two options (β=-0.6724, 95% CI: -0.8535, -0.5092, p<0.0001;

3   Figure 4 - Figure supplement 1; Supplementary File 1H). Turning to expected satisfaction

4   ratings, we observed a striking context effect on participants' expected satisfaction ratings

5   (Figure 5A), finding that more valuable contexts reduced participants' expectations (β= -2.7376,

6   95% CI: -3.3713, -2.0962, p<0.001). This analysis, as in the real-world ratings, controlled for the

7   influence of a number of variables that may have influenced participants' choices, including the

8   star rating and price of the chosen option (see Table 3).

9   In our account of the real-word ratings (Figure 3B), these (unobserved) contextually-

10   computed expectancies are the basis of the comparison with a decision-maker's true experience

11   with the rated option. In our experiment, which examined choice and rating behavior in many of

12   the same choice sets that Yelp users presumably faced in the real world, the observation that

13   these expectations decrease sharply in accordance with the average value of the context supports

14   our account of this comparison process: as contextually-bound expectations decrease, the user's

15   real-world rating—conceptualized as the true consumption utility minus expectations—

16   increases. Put another way, these experimentally elicited expectations—which appear subject to

17   contextual modulations of the surrounding choice set (Figure 5A)—provide a compelling

18   explanation for the contextual effects observed in real-world rating data.

16
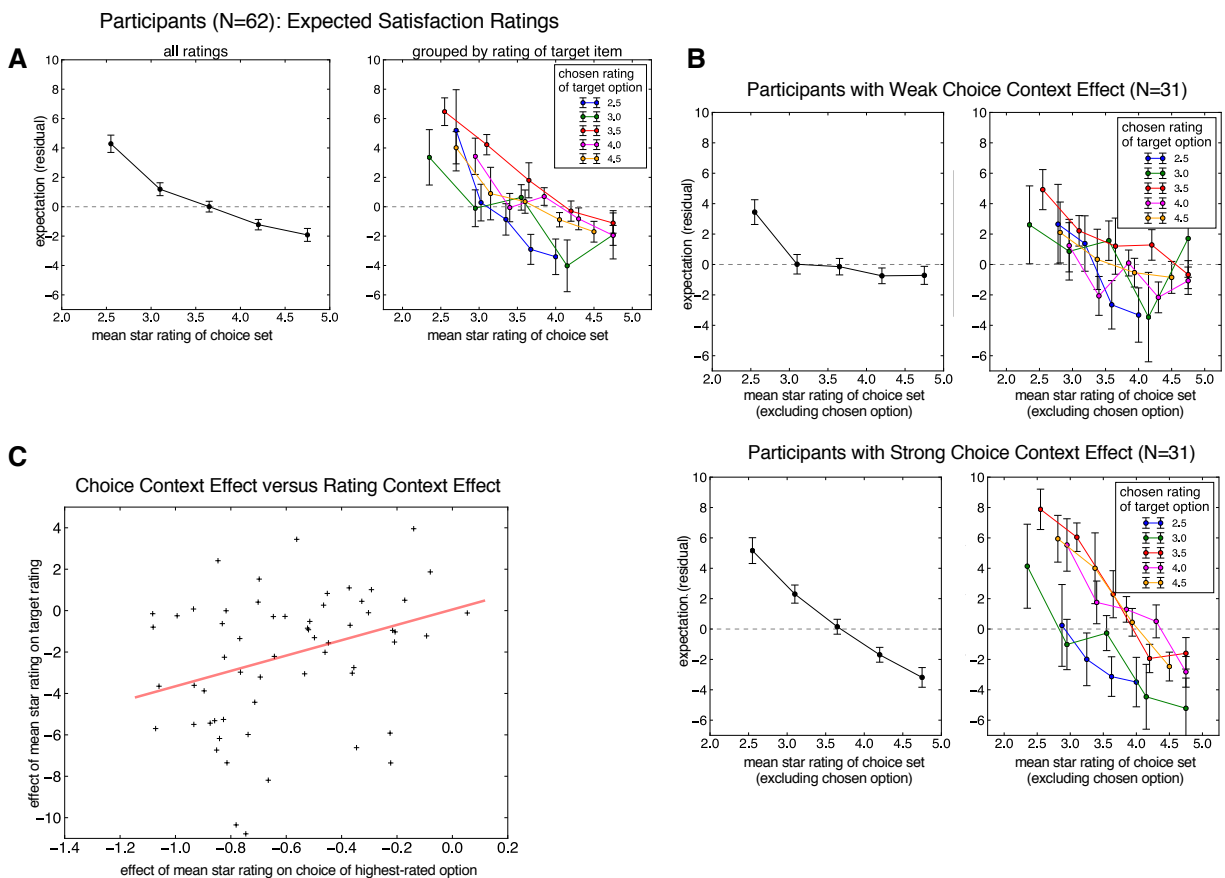
Figure 5. (A) Expected satisfaction ratings for the selected option, after controlling for a number of nuisance variables, as a function of the mean rating of the choice set, collapsed over all choice sets (left panel) and grouped by the rating of the selected option (right panel). We observed a marked context effect such that participants gave lower expected satisfaction ratings in higher-rated choice sets. (B) Context effects for participants whose choice context effect sizes were weak (top row) and strong (bottom row). We observed that these rating normalization effects were stronger in participants with steeper normalization in choices. (C) Correlation between subject-level choice context effects (estimated from the group-level mixed effects regression) and the participant-level effect of mean rating of choice set (i.e., context effect with respect to ratings; Participants who exhibited stronger (negative) context effects in choice also exhibited a stronger (negative) context effect in expected satisfaction ratings, statistically confirmed by robust regression.

Finally, we probed the relationship between individual differences in the extent of context-dependence manifested in choice— quantified by the subject-level effect of mean choice set rating upon the probability of ratings-maximizing choice—and context-dependence evident in separately-measured expectation ratings. To do this, we examined context-dependence in

17

1   expectation ratings separately for participants with small versus large context effects in choice

2   (Figure 5B; via median split), finding that context dependence in ratings was markedly stronger

3   in individuals whose choices also exhibited more context dependence. That is, the slope of the

4   relationship between the mean value of the choice set (i.e., context) and the rated expectation

5   was more pronounced for participants whose choices exhibited stronger contexts effects in

6   choice—though, critically, these contexts effects on choice and ratings were observable at the

7   aggregate, sample level. Further illustrating the concordance between context-dependence in

8   choice and expectancy ratings, we found that the strength of context dependence in choice

9   (quantified by the subject-level effect of mean rating of choice set upon ratings-maximizing

10   choices) predicted the degree to which ratings were dependent on the average value of the choice

11   set (Figure 5C; quantified by the subject-level random effect of mean rating of choice set; robust

12   regression; β= 3.7073, p= 0.0104). In other words, the degree of context-dependence in an

13   individual's choices strongly related to the degree of context-dependence in their ratings,

14   supporting the generality of these context effects to both choice and subjective valuation.

15   **Discussion**

16   Here we examined whether real-world choices, evidenced in a large dataset of online

17   restaurant reviews, are subject to context effects that have previously been observed in tightly-

18   constrained laboratory settings (22, 23). We observed that ratings-maximizing choice behavior

19   was systematically inflected by choice context, and importantly, this robust pattern of context-

20   dependent choice replicated in two separate datasets which operationalize choice differently. We

21   also found that subjective ratings were systematically biased by the rating distributions of the

22   choice sets faced by decision-makers. This observed context-dependence appeared to be driven

23   by the values (i.e., ratings) of the options alone, suggesting that this real-world rating information

                                                                                          18

1 was evaluated—and acted upon, insofar as informing restaurant choices—in a relative and

2 context-dependent, as opposed to absolute manner.

3 Further corroborating this account, in a laboratory experiment, we found evidence for

4 context dependence in decision-makers' choices and subjective expectations of options' values,

5 which, beyond replicating real-world choice context effects in a controlled setting, demonstrates

6 that the value composition of a choice set can systematically alter decision-makers' expectancy

7 of an option's value. Further, the concordance between the extent of context-dependence

8 observed in individuals' choices and context-dependence in ratings (Figure 6C) further supports

9 the idea that contextually-computed subjective valuations—observable in expectancy ratings—

10 serve as input to choices (23). Taken together, the marked context dependence we observe 'in the

11 wild' and experimentally further buttress the idea that an option's value is computed relative to

12 its context, an idea pervasive across neuroscience, psychology, and behavioral economics (3, 5,

13 48). Computationally, relative—versus absolute—value representation holds the benefit that

14 information processing resources are allocated efficiently to the encoding of subjective values

15 (14, 15).

16 Our results raise a number of questions about the nature of these observed context effects.

17 While our account postulates that the expectations informing the real-world ratings are computed

18 in a relative, context-dependent fashion—which finds more direct support in our laboratory

19 experiment (Figure 6A)—an open question concerns whether the subjective utility of

20 consumption is also computed contextually, relative to the value composition of the choice set

21 (or some other reference point). However, such effects could also arise from processes operating

22 at consumption (or feedback, e.g. (49, 50)). Future work should examine, more directly, the

23 extent to which subjective (consumption) utility might also exhibit context dependence.

19

1   While our analysis makes a core assumption that Yelp users' choices (and ratings) are

2   informed by ratings users observe on the Yelp website (or smartphone app) at the time of

3   presumed choice, it may be the case that these online ratings may act here as a proxy for users'

4   knowledge about the value of options in the environment, rather than directly informing the

5   observed choices. In other words, some users' (absolute) valuations of options could be informed

6   by direct experience with options or 'word of mouth' rather than star ratings. While user ratings

7   are often a valid—if noisy—indicator of "true" quality of the options (51) (but see (40)), we

8   remain agnostic about which source(s) of information are the basis for these choices. The

9   qualitative concordance between real-world choice patterns (Figure 2A) and laboratory choices

10   (5B), which were made in the absence of true knowledge about the options' 'true' values—and

11   instead, only on the basis of ratings information—hints that the rating information on Yelp might

12   plausibly have been the basis for real-world users' choices. Moreover, the low rates of ratings-

13   maximize choice in the real-world choice data (in contrast to the experimental choice data) likely

14   reflect other, possibly idiosyncratic considerations that decision-makers face in restaurant

15   choice—for example, personal experience or other contextual factors not represented in the data,

16   such as geographic, time, or social constraints.

17   It also is worth noting that our analysis approach—as in any real-world choice study (29,

18   31, 52)— relies critically on the operationalization of choice sets, which in our case were

19   spatially-defined clusters of restaurants belonging to a single category. This simplified our

20   analysis by guaranteeing that each restaurant appears in at most one choice set, and that

21   restaurant options only competed with each other within the context of a single category. Of

22   course, it is possible that an individual considers multiple restaurant categories when choosing

23   between restaurants (31), or that, in choosing across different categories, a restaurant could

20

1 present as an option in multiple choice sets (e.g., "pizza" and "Italian" restaurants). However,

2 any such construction would presumably be orthogonal to the context effects observed here,

3 absent any asymmetric relationship between restaurant rating and category membership. More

4 directly, it is unlikely that these simplifying assumptions systematically hold the consequence

5 that these robust context effects observed are the result of artifacts in light of the laboratory-

6 based replication of these choice patterns. Nevertheless, future work could endeavor to

7 understand how choice set rating distributions of different categories (and possibly even other

8 choice sets of the same category) might bear on context-dependent choices and ratings. Finally,

9 while a follow-up analysis revealed that these context effects were robust to the differing spatial

10 definition of choice sets across the Yelp and Deliveroo datasets, we should note it is possible that

11 individuals might, at times, strive to choose the a more globally "optimal" restaurant with respect

12 to an entire city (or large region of a city) rather a single neighborhood per se.

13   Relatedly, it is possible that the set size effect we observe on value-maximizing choice—

14 such that users were less likely to make an apparent ratings-maximizing choice in larger choice

15 sets (Figure 2A) could be an artifact of our post-hoc construction of choice sets. That is, adding

16 more options to the consideration set through any means will necessarily decrease the apparent

17 likelihood of a decision-maker choosing any one option. Suggesting against this possibility, the

18 same set size effects are observed in our laboratory data (Figure 5B) and have been previously

19 observed by others [17,37]. Nonetheless, the context effects of interest here are the distribution

20 of values (i.e., mean choice set rating) rather than the number of options in the set.

21

22   **Materials and Methods**

23   *Analysis of Yelp Dataset*

1   *Data sources and selection of users and restaurants*

2   Real-world data analysis focused on two data sources in the Yelp Open Dataset (obtained

3   from https://www.yelp.com/dataset): 1) reviews of business, which consists of a unique user ID,

4   a unique business ID, the date and time of review, the unique ID of the business reviewed, the

5   user's given star rating (ranging from 1 to 5 in whole numbers), and the text of the user's review,

6   and 2) data describing each business, which consists of unique business IDs, business names,

7   GPS coordinates (latitude and longitude) and restaurant categories. We note that the reviews in

8   this dataset reflects the output of Yelp's recommendation software, which filters out suspect or

9   biased reviews on the basis of quality, reliability, and user activity patterns.

10   We examined only restaurants in predominantly English-speaking US and Canada

11   Metropolitan Statistical Areas (MSAs) with at least 1000 restaurants, leaving six US MSAs in

12   the analysis: Charlotte-Concord-Gastonia NC-SC, Cleveland-Elyria OH, Las Vegas-Henderson-

13   Paradise NV, Madison WI, Phoenix-Mesa-Scottsdale AZ, and Pittsburgh PA (US), and two

14   Canadian MSAs: Calgary AB, and Toronto ON. We only examined choices pertaining to

15   restaurant businesses—that is, all business entries containing the category "Restaurants" in the

16   Yelp data, which totaled 59,315 restaurants in the US and Canada. Each business was mapped to

17   an MSA with the US Census Bureau's defined Core-Based Statistical Areas (in the US) or

18   Census Metropolitan Areas (in Canada). We examined user reviews that were posted between

19   January 1 2012, extending to November 14 2018 (the temporal end of the Yelp dataset).

20   *Calculation of restaurant categories and neighborhood clusters*

21   As business in the Yelp dataset are assigned multiple, unordered categories (e.g.,

22   "Restaurants, Italian, Pizza"), we inferred the single category that best described each restaurant.

23   For each restaurant, we tabulated the occurrence of each Yelp-assigned category term in the text

22

1   of the user reviews associated with that restaurant, and used the most frequently-occurring

2   category term (e.g. "Pizza"). This ensured that each restaurant occurred in at most one possible

3   consideration set, defined by restaurant category and neighborhood, preventing a restaurant from

4   occurring in multiple consideration sets (e.g. both "Pizza" and "Italian").

5   To define spatial neighborhoods within each MSA—which, alongside the restaurant

6   categories, jointly defined the restaurant consideration sets—we employed the DBSCAN

7   algorithm (35), a density-based spatial clustering algorithm that permitted us to define "clusters"

8   based on users' restaurant visit patterns, allowing neighborhoods of any arbitrary shape. The

9   DBSCAN algorithm took the GPS coordinates of each restaurant review as input (e.g., 1,191,427

10   data points in the Phoenix-Mesa-Scottsdale, AZ MSA) and two parameters: a search radius (in

11   decimal degrees), and a minimum number of data points within the search radius. Because the

12   restaurant density and overall size of the MSAs considered vary considerably, the search radius

13   and the minimum number of points were separately tuned for each MSA by perming nearest-

14   neighbor analysis on the restaurants' GPS coordinates and the total number of restaurants in each

15   MSA, following previous work (36). Across all MSAs, the average number of restaurants,

16   irrespective of category, per neighborhood (i.e., cluster) was 11.389 (SD= 1.346; see

17   Supplementary File 1A for statistics describing resultant category and neighborhood sizes).

18   *Selection of User Data*

19   We only examined the choice and review behavior of users for whom there were at least

20   100 reviews in the Yelp dataset, yielding 4,591 unique users. Within each user's series of

21   reviews, we only considered restaurants in the most frequently-visited (i.e., rated) MSA for each

22   user, for which we assume users likely have the most knowledge of the restaurant options. The

23   resultant distribution of users per MSA was as follows: Phoenix-Mesa-Scottsdale AZ: 29.84%,

23

1   Las Vegas-Henderson-Paradise NV: 28.33%, Toronto ON: 21.26%, Charlotte-Concord-Gastonia

2   NC-SC: 7.382%, Pittsburgh PA: 4.849%, Cleveland-Elyria OH: 4.324%, Calgary AB: 2.078%,

3   and Madison WI: 1.935%.

4   *Calculation of Choice sets and Choice and Rating Measures*

5   For every restaurant that a user reviewed, we analyzed that restaurant in the context of the

6   restaurant's category and its inferred geographic cluster —which, together, define a choice set.

7   Our analysis only considered choices made within sets composed of 3 or more restaurants, with

8   at least 3 unique ratings (the context effects of interest hold when we consider all possible choice

9   sets), resulting in 53,118 choices examined in the final analysis. For each option in the choice

10   consideration set, we computed the mean star rating for each option on the date of the user's

11   review—that is, considering reviews up to the point of that review but not including the review

12   to capture the user's possible informational state at the time of choice—rounding to the nearest

13   half-star, as is seen by Yelp users (51). We computed the probability of choosing the (possibly

14   non-unique) highest-rated restaurant in the choice set, following previous work (23, 24)—as a

15   function of the mean star rating of the choice set in question computed over these rounded star

16   ratings.

17   In our analyses of post-choice ratings, the Rating Deviation measure was computed as the

18   difference between the mean rating of the chosen restaurant before that user's rating and the

19   user's rating itself. To ensure user ratings measured pure context effects, the average star rating

20   of the choice sets (the abscissa in Figure 4B) analyses omitted the chosen (i.e., rated) option. The

21   residuals plotted in Figures 4B and Figures 6A-C are estimated from the mixed-effects

22   regression models reported in Tables 2 and 3, omitting the mean star rating of the choice set as a

23   predictor variable.

1   *Replication Dataset: Yelp Check-ins*

2   We analyzed the patterns of check-ins—time-stamped, geolocated record of a user having

3   been at a particular business—as a proxy for visitations to restaurants in the Yelp dataset. Using

4   the Yelp mobile app, a user can only "check into" a business, which will be shared with their

5   Yelp social network (i.e. "friends") if their current GPS coordinates are proximate to the

6   coordinates of the business they are checking into. Unlike ratings, the aggregate number of

7   check-ins to a business are not visible to other users, which mitigates the possibility that a check-

8   in could be used to bias the rating distribution or popularity of a particular restaurant.

9   The Yelp dataset contains a list of timestamped check-ins for each business (check-ins

10   per restaurant M=8,503; note these are not linked to specific users), which we analyzed similarly

11   to the ratings-inferred choice analysis. For the 565,661 check-ins considered in the analysis,

12   which spanned the years 2014-2016, we determined the surrounding choice set for that check-in

13   using the same neighborhood and category definitions as in the ratings-based choice analysis and

14   tabulated whether that check-in was made to the highest-rated option in that choice set. We then

15   examined whether the mean star rating of the choice set influenced the probability of checking

16   into the highest-rated option.

17   *Replication Dataset: Deliveroo Orders*

18   We analyzed choice patterns of customers ordering food on Deliveroo, an online food

19   delivery service popular in several countries outside of North America (31). Our dataset

20   consisted of 1,613,968 food orders, placed by 195,333 users ordering from 30,552 restaurants of

21   180 cuisine types across 197 cities, spanning February 2018 to March 2018. Similar to Yelp,

22   each restaurant's star rating (1-5 stars), the number of previous ratings for the restaurant at the

23   time of the order, and the price of the restaurant. Here, we defined choice sets as unique

25

1 combinations of cities and cuisine types (e.g., "Mexican" in Coventry, UK). We filtered out

2 choice sets with less than 5 restaurants, leaving to 821 unique choice sets in total. We calculated

3 the probability of a user ordering from the highest-rated restaurant, for each choice set separately

4 over the course of the two months. Mirroring our analysis of the Yelp datasets, we then

5 examined whether the mean star rating of each choice set influenced the user's probability of

6 ordering from the highest-rated restaurant in the set.

7 *Inferential Statistics*

8   In both the real-world and laboratory data, we estimated mixed-effects regression models,

9 were using Markov chain Monte Carlo via the MCMCglmm package (version 2.32) for R (53),

10 using logistic regression in models predicting choices of the highest-rated option with

11 uninformative priors. Fixed effects, which are indicated in each regression table, were all taken

12 as random effects over users (or participants). In all applicable models, we log-transformed the

13 length of the text review, the number of reviews for the chosen option, and the number of options

14 in the current choice set to mitigate skew apparent in these variables. In the experiment, subject-

15 level effect sizes in choice and rating context effects (Figure 6) were computed as per-subject

16 regression coefficients from the group analysis (conditioned on the group level estimates),

17 superimposed on the estimated group-level effect.

18   *Choice Experiment*

19 *Participants*

20   We collected data from 100 participants on Amazon Mechanical Turk (Crump et al.,

21 2013), using the psiTurk package (54). This sample size, which we have adopted as a standard

22 for all online studies employing within-subject designs, ensures adequate statistical power to

23 detect meaningful differences across conditions, while at the same time protecting against false

26

1   positives results or inflated effect sizes which can result from small samples. Further, this large

2   sample size anticipates the participant exclusion rates, resulting from similar criteria, which we

3   have observed in prior online investigations of this sort. Participants provided informed consent

4   in accordance with the McGill University Research Ethics Board and were paid $4.00 USD for

5   their participation. Thirty-eight participants were excluded from the sample due to failing to

6   correctly respond to at least 60% of catch trials (see below), leaving a final sample size of 62

7   (age M= 41.88, SD= 12.00, 32% female). Note that the exclusion of these participants does not

8   affect the significance of results reported above.

9   *Design*

10   Our experiment utilized 275 choice sets extracted from the real-world analysis described

11   above. Each choice set was presented to participants in a 4x3 grid of restaurants depicting an

12   image of each restaurant (from the yelp website), the average star rating of each restaurant (1 to 5

13   stars in increments of half stars, following the Yelp interface), the number of reviews used to

14   compute the average star rating, the price of each restaurant represented as dollar signs ranging

15   from one to four, and the category (i.e, cuisine the restaurant) (see Figure 5A). The size of each

16   choice set varied from 3 (71% of the choice sets) to 7 (1%) options. Following our real-world

17   analysis, restaurants in each choice were always drawn from the same category (e.g., 'pizza').

18   Each participant viewed 275 choice sets, which were randomly divided into 135 choice trials,

19   135 rating trials, and 5 catch trials.

20   *Procedure*

21   Choice trials required participants to answer the prompt "Which restaurant would you

22   choose if you were going to eat at a [category] restaurant?" among the available options.

23   Participants indicated their choice by clicking on a cell in the grid with their mouse. On rating

1   trials, a randomly chosen restaurant from the choice set was highlighted, and participants

2   provided an expectation in response to the prompt "Imagine you ended up at [highlighted

3   restaurant]. How satisfied do you think you would be with this restaurant if you were going to eat

4   at a [cuisine type] restaurant?" Participants responded to this question using a slider bar that

5   ranged from "Not satisfied" to "Very Satisfied" (the output of which was not visible to

6   participants, but ranged from 0 to 100). Finally, each participant completed 5 catch trials (used to

7   identify participants who did not follow instructions), in which they were required to choose the

8   restaurant with the star highest rating. Importantly, no specific directions were provided

9   concerning the information participants should use to make choices and ratings.

10   For all trial types, the available options were presented on the screen for 5 seconds during

11   which time participants could not make a response, and following this viewing period,

12   participants had 5 seconds to make a response. Following each choice and rating, participants

13   rated their confidence in the previous response using a that ranged from 1 ("Not at all confident")

14   to 7 ("Very confident"). Participants had 5 seconds to make a confidence response. Confidence

15   data were not examined in the present analyses.

16

Data Availability

Data from Yelp.com, anonymized experiment data, and analysis code can be accessed via the Open Science Framework at https://osf.io/ec5dx/ (DOI: 10.17605/OSF.IO/EC5DX). Due to access restrictions owing to the commercial and proprietary nature of the data, the Deliveroo data analyzed here was only accessible to one of the current study's authors, and cannot be made publicly available. We are able to provide summary-level data (i.e. means depicted in Figure 2C) in tabular format upon request or upon publication. Researchers seeking access to this, or related raw data are encouraged to contact Deliveroo corporate.

References

1. H. Simon, *Models of bounded rationality* (MIT Press, 1982).

2. J. R. Bettman, M. F. Luce, J. W. Payne, Constructive Consumer Choice Processes. *J. Consum. Res.* **25**, 187–217 (1998).

3. A. Rangel, J. A. Clithero, Value normalization in decision making: theory and evidence. *Curr. Opin. Neurobiol.* **22**, 970–981 (2012).

4. J. S. Trueblood, S. D. Brown, A. Heathcote, J. R. Busemeyer, Not Just for Consumers: Context Effects Are Fundamental to Decision Making. *Psychol. Sci.* **24**, 901–908 (2013).

5. A. Tversky, I. Simonson, Context-Dependent Preferences. *Manag. Sci.* **39**, 1179–1189 (1993).

6. S. Bavard, M. Lebreton, M. Khamassi, G. Coricelli, S. Palminteri, Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nat. Commun.* **9**, 4503 (2018).

7. M. W. Khaw, P. W. Glimcher, K. Louie, Normalized value coding explains dynamic adaptation in the human valuation process. *Proc. Natl. Acad. Sci.* **114**, 12696–12701 (2017).

8. A. R. Otto, E. Vassena, It's all relative: Reward-induced cognitive control modulation depends on context. *J. Exp. Psychol. Gen.* **150**, 306–313 (2021).

9. N. Stewart, N. Chater, H. P. Stott, S. Reimers, Prospect relativity: how choice options influence decision under risk. *J. Exp. Psychol. Gen.* **132**, 23–46 (2003).

10. I. Vlaev, B. Seymour, R. J. Dolan, N. Chater, The Price of Pain and the Value of Suffering. *Psychol. Sci.* **20**, 309–317 (2009).

11. M. Bateson, S. D. Healy, T. A. Hurly, Context–dependent foraging decisions in rufous hummingbirds. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 1271–1276 (2003).

12. T. A. Waite, Background context and decision making in hoarding gray jays. *Behav. Ecol.* **12**, 318–324 (2001).

13. J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, 1944).

14. R. Polanía, M. Woodford, C. C. Ruff, Efficient coding of subjective value. *Nat. Neurosci.* **22**, 134 (2019).

15. M. Woodford, Prospect Theory as Efficient Perceptual Distortion. *Am. Econ. Rev.* **102**, 41–46 (2012).

16. U. M. Dholakia, I. Simonson, The Effect of Explicit Reference Points on Consumer Choice and Online Bidding Behavior. *Mark. Sci.* **24**, 206–217 (2005).

17. U. Simonsohn, G. Loewenstein, Mistake: The Effect of Previously Encountered Prices on Current Housing Demand. *Econ. J.* **116**, 175–199 (2006).

18. J. R. Doyle, D. J. O'Connor, G. M. Reynolds, P. A. Bottomley, The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychol. Mark.* **16**, 225–243 (1999).

19. T. Dumbalska, V. Li, K. Tsetsos, C. Summerfield, A map of decoy influence in human multialternative choice. *Proc. Natl. Acad. Sci.* **117**, 25169–25178 (2020).

20. C. Wu, K. Cosguner, Profiting from the Decoy Effect: A Case Study of an Online Diamond Retailer. *Mark. Sci.* **39**, 974–995 (2020).

21. A. Jarvstad, U. Hahn, S. K. Rushton, P. A. Warren, Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proc. Natl. Acad. Sci.* **110**, 16271–16276 (2013).

22. B. K. Chau, C.-K. Law, A. Lopez-Persem, M. C. Klein-Flügge, M. F. Rushworth, Consistent patterns of distractor effects during decision making. *eLife* **9**, e53850 (2020).

23. K. Louie, M. W. Khaw, P. W. Glimcher, Normalization is a general neural mechanism for context-dependent decision making. *Proc. Natl. Acad. Sci.* **110**, 6139–6144 (2013).

24. S. Gluth, N. Kern, M. Kortmann, C. L. Vitali, Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nat. Hum. Behav.* **4**, 634–645 (2020).

25. M. S. Spektor, S. Bhatia, S. Gluth, The elusiveness of context effects in decision making. *Trends Cogn. Sci.* **25**, 843–854 (2021).

26. T. L. Griffiths, Manifesto for a new (computational) cognitive revolution. *Cognition* **135**, 21–23 (2015).

27. A. R. Otto, S. M. Fleming, P. W. Glimcher, Unexpected but Incidental Positive Outcomes Predict Real-World Gambling. *Psychol. Sci.* **27**, 299–311 (2016).

28. D. M. Bartels, E. J. Johnson, Connecting cognition and consumer choice. *Cognition* **135**, 47–51 (2015).

29. D. R. Davis, J. I. Dingel, J. Monras, E. Morales, How Segregated Is Urban Consumption? *J. Polit. Econ.* **127**, 1684–1738 (2019).

30. D. W. Vinson, R. Dale, M. N. Jones, Decision contamination in the wild: Sequential dependencies in online review ratings. *Behav. Res. Methods* **51**, 1477–1484 (2019).

31. E. Schulz, *et al.*, Structured, uncertainty-driven exploration in real-world consumer choice. *Proc. Natl. Acad. Sci.* **116**, 13903–13908 (2019).

32. A. D. Shocker, M. Ben-Akiva, B. Boccara, P. Nedungadi, Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Mark. Lett.* **2**, 181–197 (1991).

33. C. Wu, H. Che, T. Y. Chan, X. Lu, The Economic Value of Online Reviews. *Mark. Sci.* **34**, 739–754 (2015).

34. R. Webb, P. W. Glimcher, K. Louie, The Normalization of Consumer Valuations: Context-Dependent Preferences from Neurobiological Constraints. *Manag. Sci.* **67**, 93–125 (2020).

35. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96., (AAAI Press, 1996), pp. 226–231.

36. S. Gao, *et al.*, A data-synthesis-driven method for detecting and extracting vague cognitive regions. *Int. J. Geogr. Inf. Sci.* **31**, 1245–1271 (2017).

37. M. Luca, G. Zervas, Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Manag. Sci.* **62**, 3412–3427 (2016).

1   38. D. Mayzlin, Y. Dover, J. Chevalier, Promotional Reviews: An Empirical Investigation of
2   Online Review Manipulation. *Am. Econ. Rev.* **104**, 2421–2455 (2014).

3   39. E. Reutskaja, R. Nagel, C. F. Camerer, A. Rangel, Search Dynamics in Consumer Choice
4   under Time Pressure: An Eye-Tracking Study. *Am. Econ. Rev.* **101**, 900–926 (2011).

5   40. B. de Langhe, P. M. Fernbach, D. R. Lichtenstein, Navigating by the Stars: Investigating
6   the Actual and Perceived Validity of Online User Ratings. *J. Consum. Res.* **42**, 817–833
7   (2016).

8   41. R. D. Luce, The choice axiom after twenty years. *J. Math. Psychol.* **15**, 215–233 (1977).

9   42. B. Seymour, S. M. McClure, Anchors, scales and the relative coding of value in the brain.
10   *Curr. Opin. Neurobiol.* **18**, 173–178 (2008).

11 43. R. L. Oliver, *Satisfaction: A Behavioral Perspective on the Consumer: A Behavioral
12   Perspective on the Consumer* (Routledge, 2014).

13 44. A. Parasuraman, V. A. Zeithaml, L. L. Berry, A Conceptual Model of Service Quality and
14   Its Implications for Future Research. *J. Mark.* **49**, 41–50 (1985).

15 45. R. B. Rutledge, N. Skandali, P. Dayan, R. J. Dolan, A computational and neural model of
16   momentary subjective well-being. *Proc. Natl. Acad. Sci.* **111**, 12252–12257 (2014).

17 46. W. Schultz, P. Dayan, P. R. Montague, A Neural Substrate of Prediction and Reward.
18   *Science* **275**, 1593–1599 (1997).

19 47. D. Kahneman, A. Tversky, Prospect Theory: An Analysis of Decision under Risk.
20   *Econometrica* **47**, 263–291 (1979).

21 48. T. A. Klein, M. Ullsperger, G. Jocham, Learning relative values in the striatum induces
22   violations of normative decision making. *Nat. Commun.* **8**, 16033 (2017).

23 49. L. Fontanesi, S. Palminteri, M. Lebreton, Decomposing the effects of context valence and
24   feedback information on speed and accuracy during reinforcement learning: a meta-
25   analytical approach using diffusion decision modeling. *Cogn. Affect. Behav. Neurosci.* **19**,
26   490–502 (2019).

27 50. S. Palminteri, M. Khamassi, M. Joffily, G. Coricelli, Contextual modulation of value
28   signals in reward and punishment learning. *Nat. Commun.* **6**, 8096 (2015).

29 51. M. Luca, "Reviews, Reputation, and Revenue: The Case of Yelp.Com" (Social Science
30   Research Network, 2016) https:/doi.org/10.2139/ssrn.1928601 (April 17, 2020).

31 52. P. S. Riefer, R. Prior, N. Blair, G. Pavey, B. C. Love, Coherency-maximizing exploration in
32   the supermarket. *Nat. Hum. Behav.* **1**, 0017 (2017).

53. J. D. Hadfield, MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *J. Stat. Softw.* **33**, 1–22 (2010).

54. T. M. Gureckis, *et al.*, psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behav. Res. Methods*, 1–14 (2015).

1 **Tables**

2

Coefficient	Estimate	95% CI	p-value
(Intercept)	-0.0768	-0.3256, 0.1662	0.261
mean star rating of choice set	**-0.0846**	**-0.1425, -0.0181**	**0.002***
number of options (log)	-1.0066	-1.0776, -0.9454	<0.0001*
number of reviews for highest-rated option	0.3687	0.3476, 0.3896	<0.0001*
price of highest-rated option	0.0995	0.069, 0.1343	<0.0001*
variance of star ratings of choice set	0.5727	0.4926, 0.6557	<0.0001*

3

4 **Table 1.** Coefficient estimates for mixed-effects logistic regression predicting ratings-
5 maximizing choices as a function of mean star rating of choice set (i.e., the context effect) and
6 log-transformed choice set size, controlling for the price of the highest-rating option, and the log-
7 transformed number of reviews of the highest-rated option in the Yelp choice dataset.

Coefficient	Estimate	95% CI	p-value
(Intercept)	2.2507	2.1255, 2.3526	<0.0001*
mean star rating of chosen option	-0.4338	-0.448, -0.4197	<0.0001*
length of user review	-0.1759	-0.1903, -0.1637	<0.0001*
number of reviews for chosen option	0.027	0.0208, 0.0333	<0.0001*
price of chosen option	0.0843	0.0711, 0.099	<0.0001*
mean star rating of choice set	**0.0576**	**0.037, 0.0753**	**<0.0001***
number of options (log)	0.003	-0.0136, 0.0211	0.368

Table 2. Coefficient estimates for mixed-effects regression predicting the difference between user rating and the mean star rating of the chosen option at the time of choice (rating deviation in Figure 3) in the Yelp dataset, as a function of mean star rating of choice set excluding the chosen option (i.e., the context effect), controlling for the rating of the chosen option at the time of choice log-transformed choice set size, the price of the chosen option, and the log-transformed number of reviews of the chosen option.

Coefficient	Estimate	95% CI	p-value
(Intercept)	29.3453	24.5856, 33.2665	<0.0001*
mean star rating of chosen option	12.078	11.5269, 12.5713	<0.0001*
number of reviews for chosen option	2.4053	2.0711, 2.6817	<0.0001*
price of chosen option	0.1306	-0.4522, 0.703	0.336
mean star rating of choice set	**-2.7376**	**-3.3713, -2.0962**	**<0.0001***
number of options (log)	-1.4998	-1.9922, -0.9704	<0.0001*

1
2 **Table 3.** Coefficient estimates for mixed-effects regression predicting expected satisfaction
3 ratings as a function of the mean star rating of choice set excluding the selected option (i.e., the
4 context effect), controlling for the rating of the selected option at the time of choice, log-
5 transformed choice set size, the price of the selected option, and the log-transformed number of
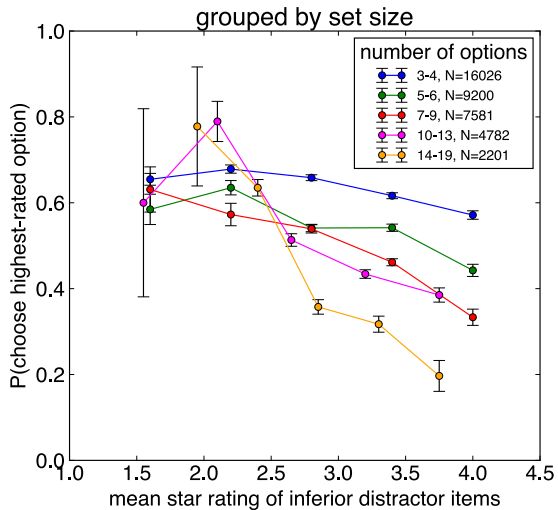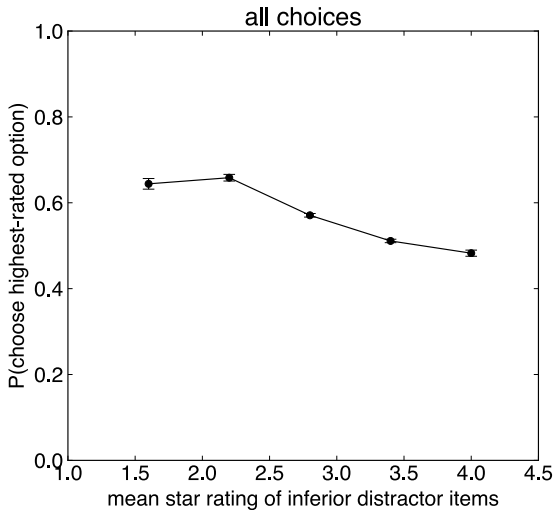6 reviews of the selected option.
7

Supplementary Figure Legends

Figure 2 – Figure Supplement 1. Proportion of choices to the highest-rated option, inferred from Yelp ratings conditioned on users choosing one of the two highest-rated options in the set, as a function of the mean star rating of the inferior "distractor" items (the sub-top-two-highest rated options), for all choice sets (left panel) and broken down by choice set size (right panel).
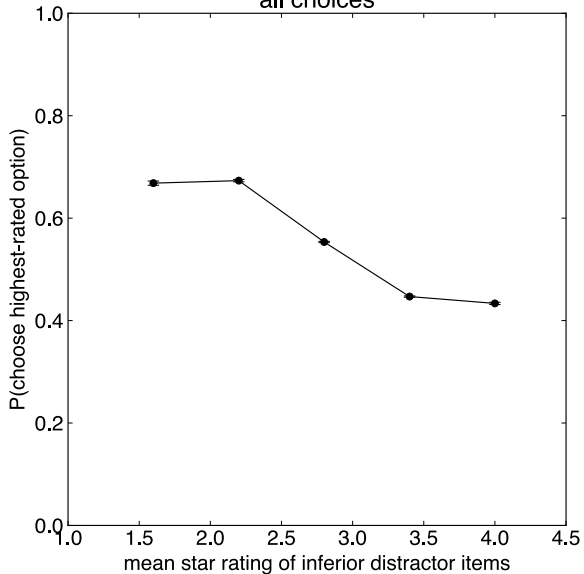
Figure 2 – Figure Supplement 2. Proportion of choices to the highest-rated option, inferred from Yelp Check-ins conditioned on users choosing one of the two highest-rated options in the set, as a function of the mean star rating of the inferior "distractor" items (the sub-top-two-highest rated options), for all choice sets (left panel) and broken down by choice set size (right panel).

Figure 4 – Figure Supplement 1. Proportion of choices to the highest-rated option in the Experiment, conditioned on users choosing one of the two highest-rated options in the set, as a function of the mean star rating of the inferior "distractor" items (the sub-top-two-highest rated options), for all choice sets (left panel) and broken down by choice set size (right panel).