

The appropriacy of averaging in the study of context effects

Shi Xian Liew¹  · Piers D. L. Howe¹ · Daniel R. Little¹

Published online: 29 March 2016
© Psychonomic Society, Inc. 2016

Abstract Models of human decision-making aim to simultaneously explain the similarity, attraction, and compromise effects. However, evidence that people show all three effects within the same paradigm has come from studies in which choices were averaged over participants. This averaging is only justified if those participants show qualitatively similar choice behaviors. To investigate whether this was the case, we repeated two experiments previously run by Trueblood (*Psychonomic Bulletin & Review*, 19(5), 962–968, 2012) and Berkowitsch, Scheibehenne, and Rieskamp (*Journal of Experimental Psychology*, 143(3), 1331–1348, 2014). We found that individuals displayed qualitative differences in their choice behavior. In general, people did not simultaneously display all three context effects. Instead, we found a tendency for some people to show either the similarity effect or the compromise effect but not both. More importantly, many individuals showed strong dimensional biases that were much larger than any effects of context. This research highlights the dangers of averaging indiscriminately and the necessity for accounting for individual differences and dimensional biases in decision-making.

Keywords Context effects · Averaging · Individual differences · Consumer products

Introduction

Tversky's (1972) finding that choices between two options are affected by the introduction of a third option challenged classical utility models of decision-making (e.g., Luce, 1959; von Neumann & Morgenstern, 1944), which assume that the relative choice probabilities of options in a choice set are not influenced by the presence of additional options. Context effects contradict this claim by showing that the relative preference for two options can depend not only on their relative utilities but also on the utilities of other options in the choice set.

Imagine you are undecided between two cell phones: a slim but structurally fragile phone (A) and a bulkier but robust phone (B). Utility models predict that you should still be indifferent to A and B even if a third phone (C) is introduced. However, if C is chosen so that it is similar to A and equally desirable (e.g., slightly bulkier but slightly more robust) then its presence can enhance the probability of choosing B relative to A. This is a violation of a core property of utility models known as *simple scalability* and is an example of the *similarity effect* (Tversky, 1972).

More generally, context effects occur when the relative choice probabilities of two options are influenced by their *context*, that is, their value relative to other options in the choice set. Aside from the similarity effect, two other context effects have gained prominence in the literature. If phone D (similar but inferior to A; e.g., Fig. 1b) is introduced to choice set {A, B}, observers are more likely to choose A over B (*attraction* effect; Huber, Payne, & Puto, 1982). If an equally desirable but more extreme phone E (e.g., Fig. 1c) is

Electronic supplementary material The online version of this article (doi:10.3758/s13423-016-1032-7) contains supplementary material, which is available to authorized users.

✉ Shi Xian Liew
sxliw@student.unimelb.edu.au

¹ Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

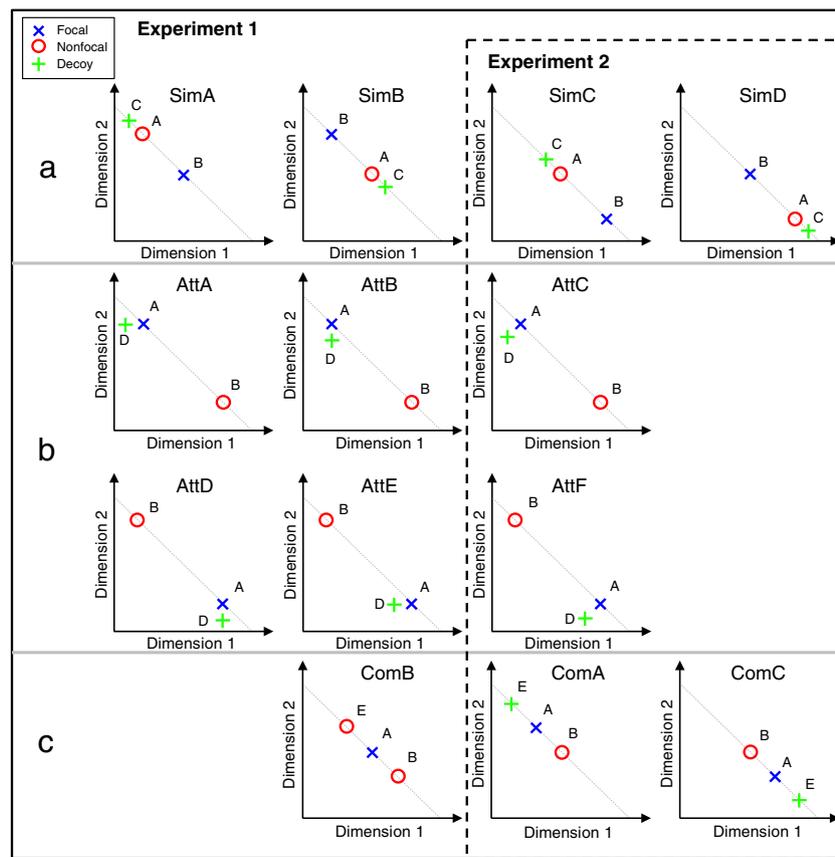


Fig. 1 Configurations of stimuli values used in Experiments 1 and 2 for the (a) similarity effect, (b) attraction effect, and (c) compromise effect. Discontinuous gray lines represent lines of indifference. Letters next to the options correspond to the examples detailed in the introduction. Experiment 1 used all 13 configurations, with the absolute positions of each option representing a particular mean eyewitness confidence value. Experiment 2

used only the six configurations within the discontinuous box. Unlike in Experiment 1, these stimuli were not dependent on specific mean values (i.e., only their relative differences were relevant). See Sections B and F of the Supplemental Material for more details on stimuli generation for both experiments

introduced to choice set $\{A, B\}$ such that A is an intermediate option between B and E, observers now choose A more than B (*compromise* effect; Simonson, 1989).

Typically, context effects have been studied by averaging choice probabilities across all participants (Trueblood, 2012; Trueblood, Brown, & Heathcote, 2014; Trueblood, Brown, Heathcote, & Busemeyer, 2013; Tversky, 1972). This is based on the (often implicit) assumption that all participants make similar choices and that any significant variation is due to random noise as opposed to systematic differences between participants. However, such an assumption is not necessarily valid (Estes & Maddox, 2005; Navarro, Griffiths, Steyvers, & Lee, 2006).

To our knowledge, only one context effect study has explicitly investigated individual differences (Kazinka, Enkavi, Vo, & Kable, 2014). In that study, a minority of participants showed a large attraction effect, whereas the remainder of participants showed no effect. Additional indications of individual differences are seen in Trueblood and colleagues

(2014), where only a minority of participants (11 %) showed all three effects, while 46 % of them demonstrated two out of three effects. A number of other context effect studies have also presented individual-level data (see Trueblood, 2012; Trueblood et al., 2013; Tversky, 1972). However, in all these studies inferences were ultimately made through averaging results across all participants. If there are systematic differences between individuals, this approach is potentially misleading.

A second issue is that context effect experiments typically average choices across multiple configurations for each effect. For a given context effect, a study might employ several different configurations where the relative locations of the options in a given attribute space remain the same but their absolute locations vary (Berkowitsch, Scheibehenne, & Rieskamp, 2014; Huber et al., 1982; Trueblood, 2012). These studies typically do not make inferences from the individual configurations separately but instead average across them, assuming that the

choice probabilities do not systematically differ between configurations.¹

The aim of the current paper is to test the appropriacy of averaging choice data across all participants as well as across different configurations of a single context effect. We repeated an experiment by Trueblood (2012) and another by Berkowitsch and colleagues (2014). We then used a hierarchical Bayesian clustering method to group the participants according to their behavior (Navarro et al., 2006). The presence of multiple groups would reflect systematic differences across participants, implying that it would be invalid to average across the whole sample (Estes & Maddox, 2005; Navarro et al., 2006).

Experiment 1

Following Trueblood (2012), participants had to choose which of three suspects they felt was most likely to be guilty of an unspecified crime based on the numerical strengths of two eyewitness testimonies. Each context effect was tested with a different group of participants.

Method

Participants Using Amazon's Mechanical Turk, we recruited 87 participants (47 females, $M_{age} = 39.44$ years) for the similarity effect, 184 participants (86 females, $M_{age} = 37.83$ years) for the attraction effect, and 90 participants (56 females, $M_{age} = 40.23$ years) for the compromise effect. These numbers were based on a power analysis of Trueblood (2012) with a replication probability of .95. Upon successful completion of the task, participants were reimbursed US\$4.00.

Materials and procedure On each trial, the participant was shown two eyewitness testimony strengths (ranging from 0–100) for each of three suspects (see Section A, Fig. S1 of Supplemental Material) where a higher number represented greater confidence in a suspect's guilt. There were 240 trials for each of the three context effects. Half of these trials were catch trials (i.e., one option was objectively superior to the others). The remaining trials tested a particular context effect. The testimony strengths in these trials were generated to fit all the configurations displayed in Fig. 1 (see Section B of Supplemental Material for details on stimulus generation). These configurations varied on the placement of the third, or *decoy*, option, relative to the two competitive options. The *focal* option was expected to experience an increase in choice probability if a context effect is present;

the other competitive option is termed the *nonfocal* option. The order of suspects was randomized on each trial.

Analysis Similar to Trueblood (2012), we excluded participants with catch trial accuracy less than two standard deviations below the mean. To assess the presence of a context effect, we used JASP (Love et al., 2015) to conduct a Bayes Factor t-test (Morey & Rouder, 2015; Rouder, Speckman, Sun, Morey, & Iverson, 2009) comparing the hypothesis that the focal choice probability was greater than the nonfocal choice probability (H_1) to the null hypothesis (H_0). The Bayes Factor is a ratio of probabilities comparing the probability of the data under H_1 , assuming a Cauchy prior (width = $\sqrt{0.50}$) on the effect size, to the probability of the data under H_0 , assuming a point prior at 0 on the effect size:

$$BF_{10} = \frac{p(\text{Data}|H_1)}{p(\text{Data}|H_0)}. \quad (1)$$

We also provide the 95 % highest density interval (HDI_δ) of the posterior distribution over effect size, δ .

Following this initial analysis, we clustered the choice probabilities of options across all six configurations for each participant using a hierarchical Bayesian clustering method (see Navarro et al., 2006; additional details are given in Section C of the Supplemental Material) and then re-examined the resulting choice probabilities in each cluster.

Results and discussion

Similarity effect Twelve participants (13.79 %) were removed due to low catch trial accuracy ($M_{catch\ trial} = 0.81$, $SD_{catch\ trial} = 0.20$). Averaging across all participants, the data showed very strong evidence² for the similarity effect ($t(74) = 14.09$, $BF_{10} = 3.62 \times 10^{19}$, 95 % HDI_δ [1.26, 1.95]; Fig. 2a).

The cluster analysis revealed four clusters.³ Averaging across all configurations, the three largest clusters (with sample sizes of $n = 43$, $n = 23$, and $n = 6$) showed very strong evidence for the similarity effect while the smallest cluster ($n = 3$) revealed weak evidence for no similarity effect (Table 1; Fig. 2b; posterior distributions over the number of clusters are presented in Section C of the Supplemental Material). Replotting the choice probabilities on the individual-configuration level (see Section E, Fig. S10 of the Supplemental Material) revealed interesting patterns. Specifically, the largest cluster contained members who have a preference towards the extreme values (i.e., the option that had the highest value regardless of dimension tended to be

¹ The exceptions to this are attraction effect studies that distinguished between the range, frequency, and range-frequency type attraction effects (see Huber et al., 1982; Trueblood, 2012).

² Interpretations of Bayes factors were guided by Jeffreys (1961).

³ When plotting out the individual data, cluster memberships clearly reflect qualitative differences. See Section D of the Supplemental Material for scatterplots of data from both Experiment 1 and Experiment 2.

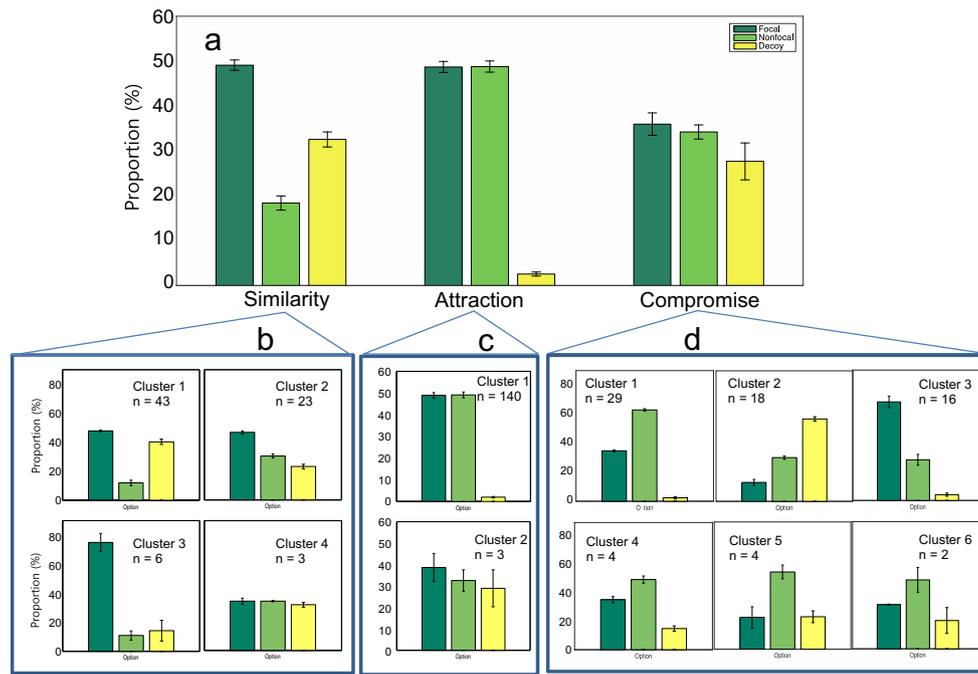


Fig. 2 Panel (a) shows choice probabilities of each option averaged across all participants and configurations for each context effect condition in Experiment 1. Lower panels show choice probabilities

averaged within each cluster from (b) similarity effect, (c) attraction effect, and (d) compromise effect conditions. Error bars represent one standard error of the mean

chosen). In contrast, the second-largest cluster contained participants who have a preference toward the options that were located at the center of the attribute space.

Attraction effect Forty-one participants (22.28 %) were removed due to low catch trial accuracy ($M_{catch\ trials} = 0.82$,

$SD_{catch\ trials} = 0.21$). Overall, the data revealed strong evidence for no attraction effect ($t(142) = -0.04$, $BF_{10} = 0.09$, 95 % $HDI_5 [-0.17, 0.16]$; Fig. 2a). The cluster analysis revealed two clusters (see Fig. 2c and Supplemental Material Section E, Fig. S11). The largest cluster comprised most of the sample (97.90 %; $n = 140$). Unsurprisingly, it showed the same result compared to averaging across the whole sample, that is, strong evidence for no attraction effect (Table 1).

There were clear configuration effects observed in cluster 1 (Fig. S11). Participants in this group demonstrated a bias towards eyewitness one’s testimony strengths (x-dimension in the figures). Although averaging across participants seemed warranted by the size of the cluster, averaging across configurations of the attraction effect was not appropriate as it obscured a distinctive choice pattern.

Table 1 Results of within-participant Bayesian t-test between focal and nonfocal choice probabilities for each cluster of Experiment 1

Context effect	Cluster (size)	t (df)	BF_{10}	95 % HDI_5
Similarity	1 ($n = 43$)	17.44 (42)	5.80×10^{17}	[1.97, 3.25]
	2 ($n = 23$)	8.43 (22)	5.56×10^5	[0.98, 2.34]
	3 ($n = 6$)	9.59 (5)	135	[0.94, 5.94]
	4 ($n = 3$)	<0.01 (2)	0.47	[-0.90, 0.89]
Attraction	1 ($n = 140$)	-0.09 (139)	0.09	[-0.18, 0.16]
	2 ($n = 3$)	-0.94 (2)	0.64	[-0.56, 1.44]
Compromise	1 ($n = 29$)	-21.47 (28)	6.85×10^{15}	[-5.00, -2.82]
	2 ($n = 18$)	-5.91 (17)	1369	[-1.95, -0.64]
	3 ($n = 16$)	5.22 (15)	280	[0.53, 1.88]
	4 ($n = 4$)	-3.15 (3)	2.34	[-2.71, 0.11]
	5 ($n = 4$)	-2.70 (3)	1.83	[-2.36, 0.16]
	6 ($n = 2$)	-2.10 (1)	1.07	[-2.96, 0.52]

Clusters are ordered from largest to smallest (membership sizes represented in parentheses in the second column)

Compromise effect Seventeen participants (18.89 %) were excluded for low catch trial accuracy ($M_{catch\ trial} = 0.79$, $SD_{catch\ trial} = 0.21$). Overall, there was moderate evidence for no compromise effect ($t(72) = -0.14$, $BF_{10} = 0.13$, 95 % $HDI_5 [-0.24, 0.21]$; Fig. 2a).

The cluster analysis revealed six clusters (see Fig. 2d and Supplemental Material Section E, Fig. S12). In the largest two clusters ($n = 29$ and $n = 18$, respectively), there was very strong evidence for a negative compromise effect where the nonfocal option was chosen more

than the focal option (see Table 1 for statistics). The third largest cluster ($n = 16$) showed very strong evidence for the compromise effect.

Configuration effects were observed across the two largest clusters. The largest cluster showed a bias where the option closest to the centre of attribute space tended to be chosen more than the other options. Conversely, there appeared to be an extremeness bias in the second largest cluster, where extreme options were chosen more than those located near the centre of the attribute space. The third cluster showed a consistent compromise effect across the three different configurations.

Experiment 2

In this experiment, we repeated the methods of Berkowitsch and colleagues (2014) whereby each participant first completed an *indifference-pairing* session, in which participants completed a missing value to equate the preferences between two options. They then completed a *preferential-choice* session as in Experiment 1.

Method

Participants We recruited 134 participants (44 females, $M_{age} = 34.10$ years) through Amazon's Mechanical Turk. This number was chosen as it has a statistical power of .95 for detecting the effect sizes reported by Berkowitsch and colleagues (2014) while taking into consideration the expected attrition rate between the two sessions. Participants were reimbursed with US\$2.00 and US\$4.00 on completion of the first and second sessions (completed about a week apart), respectively.

Materials Six different consumer products were used in this study (color printer, digital camera, notebook computer, racing bike, vacuum cleaner, and washing machine). These products varied on two attribute values specific to each product (e.g., digital cameras varied on their picture quality and memory capacity). On each trial, two options of the same product category were displayed with attribute values randomly selected from a uniform distribution with a range that depended on the product category (see Table S2 in the Supplemental Material, Section F, for the full list of products along with the range of attribute values). One attribute value of one option was randomly removed and the observer entered a value for this attribute so that the two options were equally desirable. There were a total of 108 trials (18 for each product category). The product categories were presented one at a time as separate blocks (i.e., all 18 examples of one product category were presented before the next product category was presented),

and the order in which the product categories were presented was randomized.

For the preferential-choice session, a decoy was added to each choice pair. The decoy attribute values depended on the context effect being tested and were designed to fit the six configurations as presented in Fig. 1 (with specific details in the Supplemental Materials, Section F). Each of the 18 trials per product category were equally split among the three context effects (i.e., there were six trials showing the similarity, attraction, and compromise effects each).⁴

The stimuli were presented to the participants as a table in both the indifference-pairing session and in the preferential-choice session (see Supplemental Materials, Section A, Fig. S2). The column (option) order was randomized on each trial, although the row (attribute) order was kept constant for a given product category.

Procedure Indifference-pairing session: Six blocks of 18 trials each were presented to the participant. Prior to the presentation of each block, participants were given a brief description of the product category and its associated attributes. The direction of preference for each attribute was also specified (e.g., for printing speed, the participant was told that a higher value is more preferable). On each trial, participants provided a number for each missing attribute value to make both options appear equally attractive.

Preferential-choice session: The instructions were similar to Experiment 1, except that participants had to choose the product they found the most attractive based on the attribute values displayed.

Analysis We used the same analyses as in Experiment 1.

Results and discussion

Twenty-three participants did not return for the preferential-choice session. One participant was excluded for failing to complete all experimental trials. Figure 3a shows the choice probabilities for each option. Overall, we found moderate evidence favoring no similarity effect ($BF_{10} = 0.12$, 95 % HDI₈ [-0.13, 0.23]), very strong evidence for a strong attraction effect ($BF_{10} = 5.19 \times 10^7$, 95 % HDI₈ [0.45, 0.87]), and weak evidence for a moderately-sized compromise effect ($BF_{10} = 1.26$, 95 % HDI₈ [0.03, 0.40]).

Three clusters were extracted (see Fig. 3b), with the first cluster ($n = 53$) showing very strong evidence for the attraction and compromise effects and very strong evidence for a negative similarity effect (i.e., the nonfocal was chosen more than the focal). The second cluster ($n = 30$) revealed strong evidence for the

⁴ Of the total trials, 23.26 % were excluded for having one competitive option dominating the other.

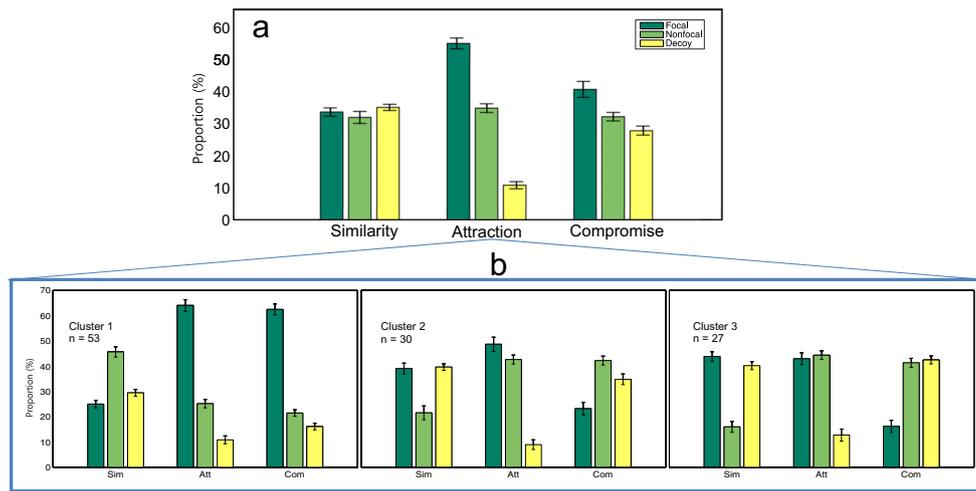


Fig. 3 Panel (a) shows choice probabilities of each option averaged across all participants and configurations in Experiment 2. Panel (b) shows choice probabilities averaged within each cluster. Error bars represent one standard error of the mean

similarity effect, weak evidence for the no attraction effect, and very strong evidence for a negative compromise effect. The third cluster ($n = 27$) revealed very strong evidence for the similarity effect, moderate evidence for the no attraction effect, and very strong evidence for a negative compromise effect. Statistical results are presented in Table 2.

Configuration-level choice probabilities are provided in Section G of the Supplemental Material. Participants in cluster 2 tended to choose the option that was highest on the second attribute (i.e., the y-dimension) for all three context effects. In cluster 3, participants preferred choosing the option highest on the first attribute (i.e., the x-dimension). Participants in cluster 1 revealed only small configuration effects. Instead, their

preferred choices tended to be the competitive option that had a similar decoy.

Note that, like Experiment 1, the result obtained by averaging over all participants did not reflect the qualitative differences present in the data. The overall lack of the similarity effect was the consequence of averaging across different clusters that separately showed positive and negative similarity effects. Additionally, the appearance of a weak overall compromise effect was the result of combining one major cluster showing a strong compromise effect and two smaller clusters showing a negative compromise effect.

Interestingly, a negative correlation between the similarity and compromise effects within participants was found (i.e., participants who showed the similarity effect also tended to show a negative compromise, and vice versa). This negative correlation was also observed in a number of other studies (see Berkowitsch et al., 2014; Trueblood, Brown, & Heathcote, 2015), and is predicted by at least three prominent models of context effects: Multialternative Decision Field Theory (MDFT; Roe, Busemeyer, & Townsend, 2001), Leaky Competing Accumulator (LCA; Usher & McClelland, 2004), and Multiattribute Linear Ballistic Accumulator (MLBA; Trueblood et al., 2014) models.

Table 2 Results of within-participant Bayesian t-tests between focal and nonfocal choice probabilities for each context effect in each cluster of Experiment 2

Cluster (size)	Context effect	$t(df)$	BF_{10}	95 % HDI $_{\delta}$
1 ($n = 53$)	Similarity	-6.25 (52)	17.72×10^4	[-1.14, -0.51]
	Attraction	10.60 (52)	5.39×10^{11}	[1.03, 1.81]
	Compromise	12.31 (52)	1.20×10^{14}	[1.24, 2.08]
2 ($n = 30$)	Similarity	3.87 (29)	54.95	[0.26, 1.06]
	Attraction	1.50 (29)	0.53	[-0.10, 0.60]
	Compromise	-5.45 (29)	2,721	[-1.38, -0.50]
3 ($n = 27$)	Similarity	7.22 (26)	13.24×10^4	[0.79, 1.86]
	Attraction	-0.39 (26)	0.22	[-0.43, 0.29]
	Compromise	-6.07 (26)	9,107	[-1.60, -0.60]

Clusters are ordered from largest to smallest (membership sizes represented in parentheses in the first column)

General discussion

Across two experiments, we found the presence of individual differences and configuration effects that indicate different decision-making patterns for different participants. These patterns are qualitatively distinct (e.g., choosing the most extreme values or choosing options with values highest on a particular dimension). Consequently, making inferences by averaging

the effects of these patterns across all participants and configurations is not valid.

This study also shows that analyses of context effects must account for decoy choice probabilities. Some studies have measured context effects by calculating the *relative choice share of target* (RST; Berkowitsch et al., 2014; Trueblood et al., 2014, Trueblood et al., 2015), which is the frequency of choosing a focal option divided by the sum of choosing both focal and nonfocal options. This is useful in representing the strength of a context effect on a single dimension. However, without considering decoy choice probabilities, what may appear to be a context effect can instead be the result of a dimensional bias. For example, in Experiment 2, cluster 2 appears to show the similarity effect (i.e., focal choice probabilities were on average higher than those of the nonfocal's). However, this observation was mainly the consequence of participants choosing options highest on the second attribute, resulting in a preference for the focal and decoy options over the nonfocal.

Despite performing a power analysis and repeating the methods of the original experiments by Trueblood (2012) and Berkowitsch and colleagues (2014), the present study did not fully replicate the original findings. This further questions the robustness of these context effects, which in recent years has been challenged by a number of other studies. For example, Frederick, Lee, and Baskin (2014) and Yang and Lynn (2014) found that the attraction effect was much stronger in abstract, stylized stimuli (e.g., numerical representations of attributes) as opposed to natural stimuli (e.g., pictorial stimuli). In addition, Trueblood et al. (2015) found (using perceptual stimuli) only a minority of participants (13 out of 55) demonstrated all three context effects. The present study extends the scepticism raised by these experiments on the robustness of context effects.

In support of the generalizability of these results, we found similar effects in a third experiment (not reported here) that combined the procedures of the two experiments presented here. Specifically, we observed multiple clusters of choice patterns, one of which revealed strong biases towards extreme values (not unlike cluster 2 of the compromise effect condition in Experiment 1). In addition, the negative correlation between similarity and compromise effects was also observed, and no cluster showed the presence of all three context effects.

Ultimately, our study is among the few that have explicitly investigated and found individual differences in context effect data. It contends that it is not appropriate to average choice probabilities across participants without first verifying that all participants are responding in a qualitatively similar manner. One way to address this is to adopt an appropriate cluster analysis (as was performed in this study) to identify systematically different groups of responses. It will also be worth investigating if current decision-making models – such as the MLBA (Trueblood et al., 2014), LCA (Usher &

McClelland, 2004), and MDFT (Roe et al., 2001) – are able to predict the choice patterns of these groups. It is possible that qualitative differences at the data level might be explained by quantitative variation in the parameters of a single model. Clustering the data first makes these differences abundantly clear (cf. Navarro et al., 2006, Fig. 14).

A few questions regarding the nature of context effects are consequently raised. For instance, it is not yet clear whether context effects are the result of similarity comparisons (e.g., more similar options do in fact strongly compete for choice share) or if they are artifacts from more general decision processes (e.g., choosing the option highest on one dimension). It is also not yet well understood what motivates different people to choose in different ways when faced with equally attractive choices. Additional work in these directions would likely yield interesting findings.

Author note This work was supported by a Melbourne Research Grant Support Scheme Grant to Piers Howe, as well as ARC Discovery Project Grant DP120103120 and DP160102360 and a Melbourne Research Grant Support Scheme Grant to Daniel Little. We would also like to thank Jennifer Trueblood and Nicolas Berkowitsch for very generously providing information and data from their studies. In addition, we are grateful to Scott Brown, Jennifer Trueblood, and Sudeep Bhatia for their invaluable comments and suggestions.

References

- Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology*, *143*(3), 1331–1348.
- Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review*, *12*(3), 403–408.
- Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, *51*(4), 487–507.
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.
- Kazinka, R., Enkavi, A. Z., Vo, K., & Kable, J. W. (2014, September). Individual differences in the asymmetric dominance effect. Poster session presented at the Society for Neuroeconomics 2014 Annual Meeting, Miami, Florida, USA.
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Wagenmakers, E.-J. (2015). JASP (Version 0.7)[Computer software]. Retrieved from <http://jasp-stats.org>.
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.
- Morey, R. D. & Rouder, J. N. (2015). BayesFactor (Version 0.9.10-2)[Computer software]. Retrieved from <http://bayesfactorpcl.r-forge.r-project.org>.
- Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, *50*(2), 101–122.
- Roe, R. M., Bussemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*(2), 370–392.

- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174.
- Trueblood, J. S. (2012). Multialternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, *19*(5), 962–968.
- Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, *24*(6), 901–908.
- Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multi-attribute linear ballistic accumulator model of context effects in multi-alternative choice. *Psychological Review*, *121*(2), 179–205.
- Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review*, *122*(4), 848–853.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299.
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757–769.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Yang, S., & Lynn, M. (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research*, *51*(4), 508–513.