RUNNING HEAD: The Fragile Nature of Contextual Preference Reversals

The Fragile Nature of Contextual Preference Reversals: Reply to Tsetsos, Chater, and Usher

Jennifer S. Trueblood

Vanderbilt University and University of California, Irvine

Scott D. Brown

University of Newcastle, Australia

Andrew Heathcote

University of Tasmania and University of Newcastle, Australia

Jennifer Trueblood
Department of Psychology
Vanderbilt University
PMB 407817
2301 Vanderbilt Place
Nashville, TN 37240-7817
phone: 949-824-1761
email: jennifer.s.trueblood@vanderbilt.edu

Abstract

Trueblood, Brown, and Heathcote (2014) developed a new model, called the Multi-attribute Linear Ballistic Accumulator (MLBA), to explain contextual preference reversals in multi-alternative choice. MLBA was shown to provide good accounts of human behavior through both qualitative analyses and quantitative fitting of choice data. Tsetsos, Chater, and Usher (in press) investigated the ability of MLBA to simultaneously capture three prominent context effects (attraction, compromise, and similarity). They concluded that MLBA must set a "fine balance" of competing forces to account for all three effects simultaneously, and that its predictions are sensitive to the position of the stimuli in the attribute space. Through a new experiment, we show that the three effects are very fragile, and that only a small subset of people shows all three simultaneously. Thus, the predictions that Tsetsos et al. generated from the MLBA model turn out to match closely real data in a new experiment. Support for these predictions provides strong evidence for the MLBA. A corollary is that a model that can "robustly" capture all three effects simultaneously is not necessarily a good model. Rather, a good model captures patterns found in human data, but cannot accommodate patterns that are not found.

Keywords: Decision-making; multi-alternative choice; preference reversal; context effects; dynamic models

THE FRAGILE NATURE OF CONTEXTUAL PREFERENCE REVERSALS: REPLY TO TSETSOS, CHATER, AND USHER

Everyday we make hundreds of choices. Some are seemingly trivial – what cereal should I eat for breakfast? Others have long lasting implications – what stock should I invest in? Despite their obvious differences, these two decisions have one important thing in common. Both are potentially sensitive to context. That is, our preferences for existing alternatives can be altered by the introduction of new alternatives. Context effects – preference changes depending on the availability of other options – have attracted a great deal of attention because they violate the property of simple scalability (Krantz, 1964; Tversky, 1972), a central property of most utility models. Trueblood, Brown, and Heathcote (2014) introduced a new model to explain three prominent context effects (attraction, similarity, and compromise) called the Multi-attribute Linear Ballistic Accumulator (MLBA). We showed that the model provides good quantitative fits to individual-subject level data as well as making new predictions about the influence of time pressure on the effects, which we confirmed experimentally. Further, MLBA is analytically tractable unlike many previous models of context preference reversals (Usher & McClelland, 2004; Roe, Busemeyer, & Townsend, 2001).

Tsetsos, Chater, and Usher (in press), hereafter TCU, question the ability of the MLBA to simultaneously capture the three context effects stating that it must set a "fine balance" of competing forces to account for all three effects, and that its predictions are sensitive to the position of the stimuli in the attribute space. In past research, models of context effects have been evaluated by their ability to simultaneously capture the three major context effects. TCU's arguments naturally lead to the question of how "robust" context effects are within participants. However, almost all past experiments of context effects have been between subjects (i.e., separate experiments for the three effects with different groups of subjects). There are very few studies examining the co-occurrence of the three effects within individuals. In Trueblood et al. (2014), we examined the co-occurrence of the three effects in a "combined" inference experiment (i.e., deciding which of three criminal suspects most likely committed a crime). In this experiment, we found that only a small minority (11%) of participants showed all three effects. In another study involving choices among consumer products, Berkowitsch, Scheibehenne, and Rieskamp (2014) found that only 19% of participants showed all three effects. They also examined correlations between the effects, finding that the attraction and compromise effects were positively correlated (r = .49), while the attraction and similarity effects (r = -.53), and the compromise and similarity effects (r=-.58), were negatively correlated.

The results of Trueblood et al. (2014) and Berkowitsch et al. (2014) suggest that the effects are very fragile and only a small subset of people show all three simultaneously. Thus, developing models that can "robustly" produce all three context effects with a single set of parameters is perhaps a misguided exercise, possibly leading to models that fail to match empirical reality. Rather, the focus should be on developing models that can accurately capture patterns and correlations found in human data. To further investigate the fragile nature of context effects, we examined the co-occurrence of the three effects in a perceptual decision-making task (Trueblood et al., 2014; Trueblood, Brown, Heathcote, & Busemeyer, 2013). We then compared the results of the experiment to *a priori* predictions from MLBA using artificial stimuli and very general assumptions about parameter values.

Combined Perceptual Experiment

Trueblood et al. (2013) examined the three context effects using a simple perceptual decision-making task – deciding which of three rectangles had the largest area. The three effects were examined in three separate experiments with different participants. In the current experiment, all three effects were tested within participants during a single session.

Method

Seventy-five undergraduate students at the University of California, Irvine participated for course credit. Similar to Trueblood et al. (2013), participants were told they would see three rectangles on each trial and to select the one with the largest area. The height and width of the rectangles functioned as the two attribute values, analogous to attributes of price and quality in an experiment about consumer products. All rectangles were solid black in color and appeared on a white background. The rectangles were numbered from left to right and the location of different rectangles (i.e., target, competitor, and decoy) was randomized across trials. The vertical placement of the rectangles varied so that they were not all positioned on the same horizontal axis. Further details about the stimuli and experimental design can be found in Trueblood et al. (2013).

The experiment consisted of a total of 720 randomized trials with 160 testing the attraction effect, 160 testing the similarity effect, and 160 testing the compromise effect. The remaining 240 trials were catch trials containing one rectangle that was clearly larger than the other two. These catch trials were used to gauge accuracy and engagement throughout the experiment.

Results

Twenty participants answered more than one third of the catch trials incorrectly and were removed from the analyses. These participants were most likely not engaged in the task. For the remaining 55 participants, we removed trials with very short response times (less than 100 ms) and trials with very long response times (more than 8 s). On average, this procedure removed about 1% of trials for each participant (about five out of 480 trials for the three effects).

For each participant, we calculated the *relative choice share for the target* (RST; Berkowitsch et al., 2014; Trueblood, 2015), defined as the number of times the target was selected divided by the number of times the target plus the competitor were selected. For this analysis, we collapsed across two different types of choice sets differing in the orientation of the target option – sets where the target was oriented vertically and sets where the target was oriented horizontally. If the RST value is greater than 0.5, this indicates that the target is selected more often than the competitor. Values equal to 0.5 suggest that the target and competitor were preferred equally.

Using the RST values, we examined the how frequently multiple effects occurred within a single participant. Out of 55 participants, only 13 had RST values greater than 0.5 for all three effects. Of the remaining participants, 12 had RST values greater than 0.5 for the similarity and attraction effects, 4 had RST values greater than 0.5 for the similarity and compromise effects, 12 had RST values greater than 0.5 for the attraction and compromise effects, 13 had RST value greater than 0.5 for only one of the three effects, and one individual had RST values less than 0.5 for all three effects. Table 1 lists the percentage of participants showing the different combinations of effects.

INSERT TABLE 1 HERE

We also used a hierarchical Bayesian model (Trueblood, 2015) to test whether the RST values were on average greater than 0.5. For each effect, we assumed that the number of times the target is selected follows a binomial distribution where θ represents the probability of the target being selected and *n* is the total number of times the target plus the competitor are selected. We assumed that each individual has a different θ parameter for each of the three effects. We also assumed that these person-specific parameters are drawn from population-level beta distributions with hyperparameters defining the mean μ_j and concentration κ_j. A graphical model and the priors for μ_j and κ_j are shown in Figure 1 (Lee & Wagenmakers, 2014). The priors for these two parameters were determined from previous work (Trueblood, 2015; Lee & Wagenmakers, 2014) and set to be relatively vague. The prior for μ_j also slightly favors the null hypothesis that the RST values are equal to 0.5. Note that re-running the analysis with a uniform prior (which favors all values equally) did not change our results. Using JAGS, three MCMC chains were used to estimate the posterior distributions. All chains converged as measured by $\hat{R}$ values close to 1 (note that, $\hat{R}$ values close to 1 are ideal and values lower than 1.1 are considered satisfactory).

INSERT FIGURE 1 HERE

Table 2 lists the means of the posteriors for the hyperparameters μ_j, the 95% highest posterior density intervals (HDIs) for μ_j, and the results of frequentist tests. The HDIs represent the most credible posterior RST values at the group level (Kruschke, 2011). If this range lies above 0.5, then one can infer that the target was on average selected more often than the competitor. Both the Bayesian and frequentist analyses suggests that μ_j was greater than 0.5 for both the attraction and similarity effects. The RST value for the compromise effect tended to be larger than 0.5 as well, but this result was not as strong as the attraction and similarity results. This finding is similar to Trueblood et al. (2013), which also found that the compromise effect was the weakest of the three effects.

We examined the correlation between the three effects using the individual parameters from the hierarchical Bayesian model. Correlations were calculated using the programs provided in Lee and Wagenmakers (2014). The posterior mean for the correlation between the similarity and attraction effects was -0.285 (95% HDI -0.527 to -0.044), the similarity and compromise effects was -0.271 (HDI -0.508 to -0.025), and the attraction and compromise effects was 0.666 (HDI 0.513 to 0.808). The data used in all of the analyses presented here are available in the Supplementary Information.

INSERT TABLE 2 HERE

Conclusions

Although there was evidence for the three context effects on average, very few individuals (13 out of 55) demonstrated all three effects simultaneously. These results confirm previous findings suggesting there are large individual differences in the manifestation of context effects (Trueblood et al., 2014; Berkowitsch et al., 2014). Further, our analyses revealed several interesting correlations between the effects. The attraction and compromise effects were both negatively correlated with the similarity effect, but positively correlated with each other, similar to Berkowitsch et al. (2014).

Model Predictions

To examine how well MLBA can account for the proportions of RST values and correlations found in the experiment, we conducted a prior predictive exercise where we calculated RST values predicted by the MLBA, by randomly sampling MLBA parameters. For this exercise, we examined three artificial choice sets from Trueblood et al. (2014) that were used for the qualitative analyses in that paper. These choice sets are listed in Table 3. Similar to the original paper, we fixed the start point parameter to $A = 1$, the threshold parameter to $\chi = 2$, the drift rate standard deviation to $s = 1$, and the baseline input parameter to $I_0 = 5$. The curvature parameter m determines the mapping from experimentally defined options (such as the height and width of rectangles in pixels) to subject values (see Trueblood et al., 2014, Figure 3). When $m > 1$, intermediate options are preferred to extremes and the opposite is true when $0 < m < 1$. When $m = 1$, subjective and objective values are equal. Because m must be positive, we assume m is log-normally distributed, i.e. $m \sim e^{N(0,1)}$. The attention weights in MLBA reflect the amount of attention given to pairwise comparisons of the options. The calculation of these weights involves free parameters λ_1 and λ_2. Because these parameters must also be positive, we assume that they are log-normally distributed as well, i.e. $\lambda_1, \lambda_2 \sim e^{N(0, 1)}$.

We simulated 200 MLBA "participants" and calculated how frequently multiple effects occurred within a single simulated participant (see Table 1). From the table, we see that MLBA closely matches the distribution of participants for the different combinations of effects from the experiment. Further, we calculated the correlation between the three effects using the RST values from the simulation. Similar to the experimental data, we found that the similarity and attraction effects were negatively correlated (r = -0.948) as well as the similarity and compromise effects (r = -0.657). However, the attraction and compromise effects were positively correlated (r = 0.651). Note that these results were not obtained by fitting data, but are *a priori* predictions using artificial stimuli and very general assumptions on parameter values.

INSERT TABLE 3 HERE

The results of the combined perceptual experiment provide further evidence of the fragile nature of context effects in human choice. That is, people do not “robustly” show all three context effects simultaneously. Further, the results of the prior predictive exercise suggests that the MLBA’s “fine balance” of opposing factors, identified by TCU, actually yields predictions that very closely match human data. Taken together, the modeling predictions provided here, along with the modeling work in Trueblood et al. (2014), show that MLBA provides a very good account of human behavior. Future research could examine how well alternative models such as Decision Field Theory (Roe, Busemeyer, & Townsend, 2001) and the Leaky Competing Accumulator model (Usher & McClelland, 2004) account for these results as well as empirically investigate similarities in context effects across different domains, such as perceptual and consumer choice tasks.

General Discussion

TCU also raise concerns about MLBA’s ability to predict the frequency attraction and compromise effects as well as concerns about “back-end” predictions. They argue that MLBA cannot capture the frequency attraction effect when the decoy is very similar to the target. They also claim that MLBA provides a partial account of the compromise effect. The “back-end” process refers to the LBA model, which converts drift rates into choice probabilities and latencies. TCU argue that the LBA “back-end” will result in a decrease in response times as the number of options increases, contrary to Hick's Law (1952). We address each one of these concerns below.

Frequency Attraction Effect

TCU argue that the frequency attraction effect will reverse producing a "negative attraction" effect when the decoy is very close to the target option. In their analyses, they use artificial stimuli with the following values: target = (4,6), competitor = (6,4), and decoy = (4, 5.9). TCU searched across a large number of parameter values for MLBA and showed that the model rarely produces a frequency attraction effect when the decoy is at this particular location. With this specific choice set, the decoy option is very similar to the target and MLBA produces a similarity-type of effect instead (i.e., the competitor is preferred). We hypothesize that individuals will show a similarity-type of effect when the target and decoy are very close together – such as when the target is (4,6) and the decoy is (4, 5.9) – because they are difficult to discriminate. We suggest that future research could explore MLBA's prediction of a "negative attraction" effect for decoy options that are very similar and almost indistinguishable from target options. In particular, we expect this result to hold for perceptual stimuli such as the ones analyzed in Trueblood et al. (2014).

TCU also state, "assuming that each parameter set corresponds to a single participant, the prediction of MLBA is that it should be possible to find choice scenarios that generate negative attraction effects. This is a distinctive prediction of the model, which to the best of our knowledge, has no precedent in the literature" (p. 6). However, there are studies that report no significant effect of frequency decoys on target options (Huber et al., 1982; Dhar & Glazer, 1996). In fact, Huber et al. (1982, p. 94) concluded, "the weakness of the frequency strategy suggests that this strategy is not as successful in revising weights as had been expected; it also indicates that dominance per se may not be as critical as the particular placement of the decoy". Although these studies did not specifically investigate negative attraction effects, they suggest that frequency decoys do not necessarily lead to preference for the target. Further, Trueblood (2012; Figure 2b, p. 965) showed that there were a number of participants who demonstrated preference for the competitor over the target (i.e., a negative attraction effect). Thus it is possible to find choice scenarios and individuals that generate negative attraction effects.

Compromise Effect

TCU show that in order for MLBA to simultaneously predict the compromise and similarity effects, the value of the curvature parameter m must be greater than 1. When $m > 1$, MLBA predicts extremeness aversion (midrange options are preferred to extremes). TCU discuss two experiments (Tversky & Simonson, 1993; Pettibone, 2012) as evidence against extremeness aversion. However, these experiments did not test the similarity effect and the subjective inferiority of extreme options is only necessary when accounting for both compromise and similarity effects. Thus, the results of these experiments are not inconsistent with MLBA because we do not know if participants would have shown a similarity effect. In fact, our experiment above suggests that very few people show both compromise and similarity effects simultaneously and that the two effects are negatively correlated. When accounting for the compromise effect alone, the m parameter is not restricted to values greater than 1. Values of this parameter less than 1, corresponding to extremeness seeking, can also yield a compromise effect.

Further, Trueblood et al. (2014, p. 196) directly tested the subjective inferiority of extreme options in binary choice. In our task, we asked participants to choose between midrange options (i.e., options with moderate attribute values such as 50 on a scale of 0 to 100) and extreme options (i.e., options with extreme attribute values such as 20 and 80). The results of this experiment showed extremeness aversion where people preferred midrange options to extremes. Our results confirm previous findings of extremeness aversion in binary choice by Usher, Elhalal, and McClelland (2008). Note that in the two studies discussed by TCU (Tversky & Simonson, 1993; Pettibone, 2012), the subjective inferiority of extremes was not the main focus of the experiments and was never directly tested.

TCU also argue that when an extreme option objectively dominates two midrange options (i.e., choice set {A, B, E} in TCU Figure 2, p. 2), the compromise effect reverses in MLBA. However, this choice set does not constitute a compromise effect because E is superior to both A and B. To our knowledge, this type of choice set has never been tested experimentally. Thus, there is no evidence to confirm or disconfirm the predictions of MLBA. Further, the predictions discussed in TCU are dependent on very specific parameter ranges and may not hold more generally.

Back-End Predictions

The MLBA includes the LBA model of simple decision-making, which takes as inputs preference strengths for the different choice options (“drift rates”) and produces as outputs choices and choice latencies. The LBA is a simple independent race model, with the different choice options racing independently to determine a winner. All other things being equal, independent race models predict that decision times should become faster as more options are added to the choice set (because of "statistical facilitation"). This prediction does not match data, in which choices reliably slow down when more options are presented, as in the famous Hick's Law. However, the difficulty in applying such a test to consumer choice data is that "all else" is *not* equal. Decisions with different numbers of choice alternatives can be treated very differently by participants, and it is not reasonable to expect the parameters of the model to be constant across such conditions. For example, participants may make more cautious decisions when faced with more alternatives, and this would slow down their decision-making appropriately. In fact, this mechanism was invoked as part of Usher, Olami, and McClelland's (2002) explanation of Hick's Law in a competitive race model. There are also other mechanisms that might plausibly contribute to explaining Hick's Law in an independent race model, such as input normalization.

TCU also describe several interesting predictions of MLBA regarding three alternative choice sets for response times when an inferior option is shifted. They show that shifting an inferior option closer to target options can produce an increase in response times. This suggests that when an inferior option becomes "less" inferior and more similar to target options, people take longer to make a decision. Such a slow down could reflect increased competition as the inferior option becomes more desirable. Future research could test this interesting prediction as well as investigate context effects that manipulate choice-set size and measure RT.

Conclusion

In Trueblood et al. (2014), we developed the MLBA model and showed that it provides a very good account of behavior in multi-alternative, multi-attribute choice.

Despite the fact that models of context effects have existed for well over a decade (Usher & McClelland, 2004; Roe et al., 2001), we were one of the first to quantitatively fit models to context-effect choice data (also see Berkowitsch et al., 2014, for another recent example). Past modeling efforts have focused on qualitative questions alone; that is, does a model produce the expected ordering of some probability measures, and in particular, can the model account for all three major context effects with a single set of parameters? This model test assumes that human beings will also demonstrate all three major context effects simultaneously, but the evidence for this is missing. The arguments raised in TCU follow the past trend of expecting a model of context effects to account for qualitative patterns that have not be confirmed experimentally. In particular, TCU argue that MLBA must set a "fine balance" of opposing forces to account for the three context effects simultaneously. We showed that the effects are very fragile and only a small subset of people shows all three simultaneously. Further, MLBA provides a very good account of these experimental results. In the future, we encourage researchers to focus on developing models that can accurately capture patterns and correlations found in human data rather than on artificial modeling goals. Indeed, we believe it should count against a model if it has the flexibility to predict patterns that are not seen in empirical data (Roberts & Pashler, 2000).

References

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2013). Rigorously testing multialternative decision field theory against random utility models. Journal of Experimental Psychology: General. Advance online publication. doi:10.1037/a0035159

Dhar, R., & Glazer, R. (1996). Similarity in context: Cognitive representation and violation of preference and perceptual invariance in consumer choice. *Organizational Behavior and Human Decision Processes*, *67*(3), 280-293.

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. doi:10.1086/208899

Krantz, D. H. (1964). The scaling of small and large color differences (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 65-5777)

Kruschke, J. (2011). Doing Bayesian Data Analysis: A tutorial introduction With R. Burlington, MA: Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2014). Bayesian Cognitive Modeling: A Practical Course. Cambridge, UK: Cambridge University Press.

Pettibone, J. C. (2012). Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice. Judgment and Decision Making, 7, 513–523.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. Psychological Review, 108, 370–392. doi:10.1037/0033-295X.108.2.370

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Trueblood, J. S. (2015). Reference point effects in riskless choice without loss aversion. *Decision,* 2(1), 13-26.

Trueblood, J. S. (2012). Multi-alternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, 19 (5), 962-968 doi: 10.3758/s13423-012-0288-9

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, 179-205.

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, 901–908. doi:10.1177/0956797612464241

Tsetsos, K. Chater, N., and Usher, M. (in press). Examing the mechanism underlying contextual preference reversal: Comment of MLBA (Trueblood, Brown and Heathcote, 2014). *Psychological Review*.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79, 281–299. doi:10.1037/h0032955

Tversky, A., & Simonson, I. (1993). Context dependent preferences. Management Science, 39, 1179–1189. doi:10.1287/mnsc.39.10.1179

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. Psychological Review, 111, 757–769. doi:10.1037/0033-295X.111.3.757

Usher, M., Elhalal, A., & McClelland, J. L. (2008). The neurodynamics of choice, value-based decisions, and preference reversal. In N. Chater & M. Oaksford (Eds.), The probabilistic mind: Prospects for Bayesian cognitive science (pp. 277–300). Oxford, England: Oxford University Press. doi:10.1093/acprof:oso/9780199216093.003.0013

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed-accuracy tradeoff. *Mathematical Psychology, 46*, 704-715.

Table 1. The percentage of participants showing different combinations of effects compared to simulations using MLBA

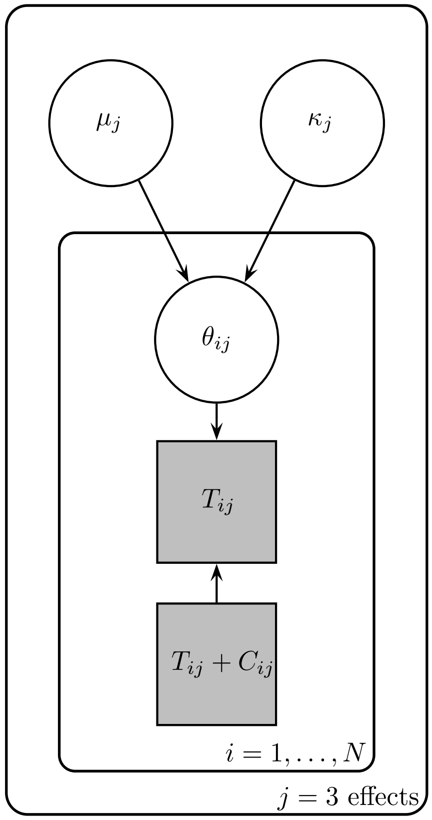
Combinations of effects	Experiment (%)	MLBA model (%)
Three effects	23.6	20.0
Similarity and Attraction	21.8	13.5
Similarity and Compromise	7.3	3.0
Attraction and Compromise	21.8	29.5
One effect	23.6	34.0

Table 2. Bayesian and frequentist analyses for the three context effects

Effect	Posterior Mean	HDI	Frequentist test
Attraction	0.531	0.514-0.548	$t(54) = 3.74$, $p < 0.001$
Similarity	0.555	0.532-0.578	$t(54) = 4.67$, $p < 0.001$
Compromise	0.518	0.498-0.538	$t(54) = 1.77$, $p = 0.08$

Table 3. Artificial Stimuli from Trueblood et al. (2014) used for the prior predictive exercise

Effect	Target	Competitor	Decoy
Attraction	(4,6)	(6,4)	(3.7, 5.7)
Similarity	(4,6)	(6,4)	(6.2, 3.8)
Compromise	(4,6)	(6,4)	(2, 8)



$$\mu_j \sim \mathrm{Beta}(2, 2)$$
$$\kappa_j \sim \mathrm{Gamma}(0.001, 0.001)$$
$$a_j = \mu_j \kappa_j$$
$$b_j = (1 - \mu_j)\kappa_j$$
$$\theta_{ij} \sim \mathrm{Beta}(a_j, b_j)$$
$$T_{ij} \sim \mathrm{Binomial}(\theta_{ij}, T_{ij} + C_{ij})$$

Figure 1. Hierarchical Bayesian graphical model testing whether the RST values for three context effects were greater than 0.5. The bounding rectangles represent independent replications of the graphical structure. The inner rectangle shows replications for individuals i = 1 to N and the outer rectangle shows replications for the three different effects. Tij denotes the number of times the target was selected and Cij denotes the number of times the competitor was selected.