**Reviewers' comments:  
  
Reviewer #1: With the revisions it is clear that this is an important paper. The attraction effect was important when it came out almost 40 years ago because it showed that human choice could not be simply characterized by a utility score for each object in a choice set, but depends on the relationships among the alternatives in the set. The attraction effect also violated the similarity effect, where similar items take disproportionate share from each other. The attraction idea was surprising because Duncan Luce said that regularity was one of the few choice axioms that has not been violated. Because attraction was multiply determined many future studies examined the sources and boundaries of the effect.  
  
The authors have clearly demonstrated that the attraction effect is negligible with complex, natural objects. The choices are among movies, a category in which it is a very difficult to test asymmetric dominance because the movie titles and images have many components whose values differ strongly across people. This study elegantly resolved that indeterminacy by measuring the preferences for movies at the individual level. Then, an elaborate sorting process defines individual asymmetrically dominating choice sets and shows that there is negligible evidence for the attraction effect. Notice that even if there was an attraction effect in some of the tested choice sets, the fact that it only worked for such a small fraction of possible choice sets indicates that the attraction effect is extremely rare. I suggest the following changes.  
  
It is important for the authors to not overstate their finding. Claiming a zero effect asks other researchers to find cases where the attraction effect is not binding. For that reason, it would be valuable to feature the findings currently in the appendix in the main paper, even if they were not specified in the pre-registration.**

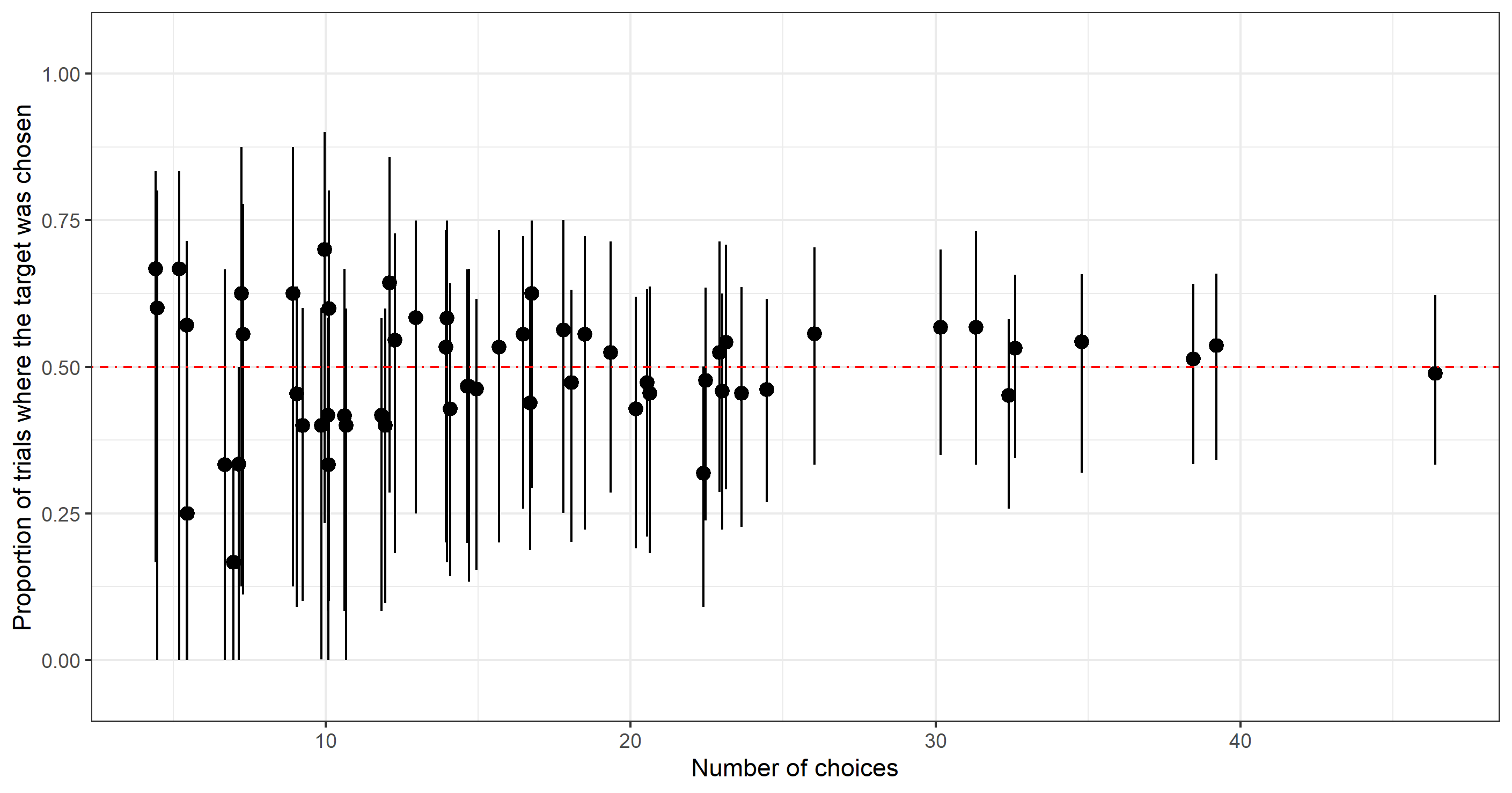
We now moved all tables and figures in the Appendix into the main body of the manuscript.  
 **Drop or revise Figure 5. This is a confusing figure. It appears that randomly choosing subsets of choices stratified by participant results in choices around 50% which have high variance with smaller samples but converge to a very narrow boundary. When 91% of observations have exactly zero attraction are merged with 9% that may not, it is not surprising that such an unreasonable confidence interval could arise. The bottom line is better given by Models 3-5 which show that the average positive and negative attraction effects are simply not significantly different from zero for the first choice.**

We added a sentence to clarify the funnel pattern seen on Figure 5.

*“The proportion of trials where the target was chosen shows substantial variation for participants who were presented with relatively fewer choice trials, whereas it is more narrowly concentrated around .5 for participants with a higher number of trials.” (page 12)*

**The analysis in tables 3-5 has fixed effects for each individual. What proportion are significantly different from zero? Indeed a histogram of the individual effects would be a good replacement for Figure 5. Alternatively, within a random effects model for individuals, consider examining the distribution of the individual tests of attraction to provide an estimate of how many have significant positive or negative attraction effects.**

Our original analysis plan did not aim to estimate the attraction effect on the individual level, and therefore our experimental design does not have enough statistical power for a precise estimation of these effects. Nevertheless, to give a sense of the variation in the attraction effect on the individual level, here we include a version of Figure 5, where we calculated individual bootstrapped confidence intervals (CIs) for the proportion of trials where the target was chosen. We excluded participants for whom this proportion was exactly 0.5 (and therefore could not have had a significant negative or positive attraction effect; about 57% of participants). From the remaining 58 participants, only one had CIs with the upper limit below 0.5 and no participant had CIs with a lower limit higher than 0.5.

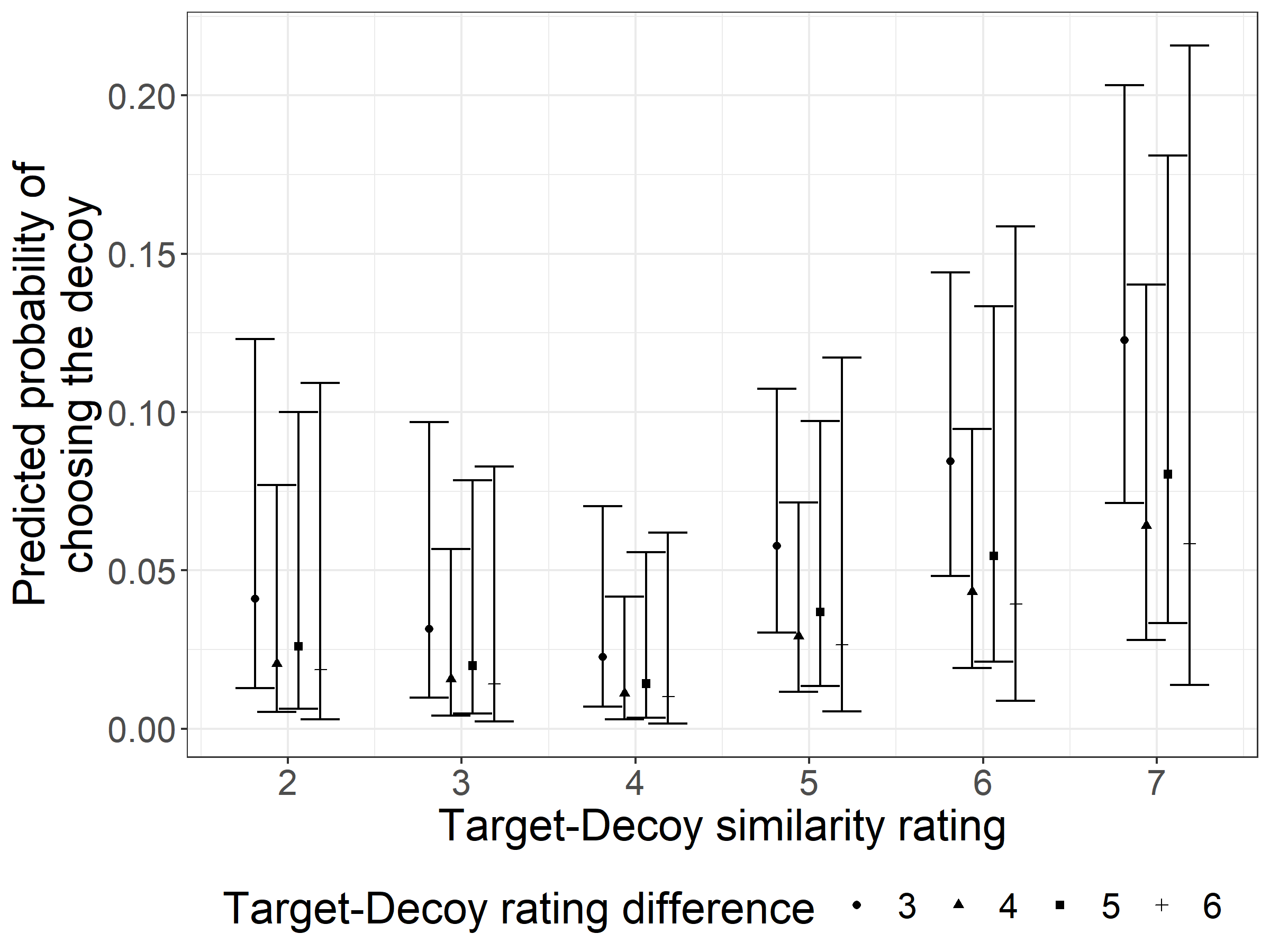


**Additionally, report the significance of the covariance factors. For example, test to see there is a significant positive association between TD rating difference and the measure of the attraction effect.**

This is exactly what we have done in Models 2, 4, 5, 8, and 9 to explore the factors influencing the strength of the attraction effect. We found no evidence for an association between TD rating difference and our measure of the attraction effect.

**Negative attraction effects are certainly possible if generated by a negative similarity effect, as the authors note in their reference to the Spektor et al (2018). A reading of the Figure A1 suggests that greater target-decoy similarity and smaller rating differences may result in fewer choices for the target over the competitor This could arise if the TD are close in ratings or similarity so that the decoy chosen takes disproportionate share from the target. It is important to refute that possibility by testing whether decoy choice increases for TD pairs that are more similar or have more equal ratings.**

To explore this, using the subset of first choices, we ran a logistic model where the outcome variable was the decoy being chosen and the explanatory variables were the target-decoy similarity ratings and rating difference. Here we include a plot of the predicted probability of choosing the decoy for each level of the explanatory variables. Altogether the decoy was chosen in less than 5% of trials. However, the decoy was more likely to be chosen when the rating difference between the target and the decoy was low, and when the decoy and target were perceived as more similar. Nonetheless, this pattern could not have affected our measure of the attraction effect, since we measured the attraction effect by calculating the proportion of trials where the target was chosen after excluding all trials where the decoy was chosen.

****

**Finally, Table A2 is confusing. Given the 2000 observations it clearly includes both the first and second choices. Thus except for including a genre rating difference, it has the same problems as Table 1 where that 91% of the observations must be exactly zero. Further, it is not clear from the text, but the results in figure A1 certainly should come from the first choice in each AB dyad.**

We reran the regressions in Table A2 (now Models 6 and 8 in Table 3) using only the first choices (see Models 7 and 9 in Table 3). We also updated Figures A1 and A2 (now Figures 7 and 8) so that these visualisations only include the first choices.

**The paper importantly shows that the attraction effect in negligible for complex and natural objects. It does so using a laborious method that is both clever and innovative. However, the authors need to be careful not to overstate their case by building a statistical argument around a questionable methodology.  
  
  
Reviewer #2: I like the changes that have been made. My comments are now very minor.**

**page 2). I think it is very odd not to cite Frederick et al. here, along with Specktor et al & Cataldo and Cohen. Showing that the attraction effect is "highly dependent on the exact presentation format" was the central point of our paper, as you correctly note on page 19.**

We agree and added this reference. **page 3). I may be wrong here, but I don't think it is quite right to say that Yang and Lynn reported difficulties replicating the attraction effect in experiments where the stimuli were pictorial. That is true in their data, but I don't think they "reported difficulties." I actually don't think they made that distinction at all.**

We changed the sentence to “*who demonstrated difficulties replicating the attraction effect in experiments where the stimuli had qualitative-verbal or pictorial depictions*”.

**page 4) I actually think the attraction effect is limited to choice tasks in which options are FULLY represented with numerical attributes; where the only presentation format is numeric. I think that is a more accurate statement than (at least partly).**

We removed the word “partly”. **page 4) The paragraph preceding the "Testing the attraction effect" section is somewhat awkwardly written to me. You have a counterfactual result that actually obtains (no effect) and then a counterfactual result that does not (an effect) and then a factual result (no effect). There is nothing "wrong" exactly, but I find the structure confusing and I think someone reading fast might miss the key point: no effect with naturalistic stimuli.**

We shortened the paragraph as follows:

*“In this article, we describe a rigorous test of the attraction effect with complex, naturalistic choice options, using a carefully developed experimental methodology that addresses all of the critical conditions discussed by Huber et al. (2014). Our results show a precisely zero attraction effect. Taken together with earlier results from Frederick et al. (2014) and Yang and Lynn (2014), we see this as strong evidence for the claim that the attraction effect is limited to choice tasks where options are represented in a stylised format with objectively defined attributes.”* **page 12) I find this sentence confusing "This indicates that participants were able to identify the lowest rated option in the choice stage" The word "identify" indicates that that was their GOAL, as opposed to simply indicating their preference from which one can INFER that one or more of the options was dispreferred.**

Thank you, we changed the sentence to:*“The fact that participants only chose the lowest-rated option very rarely in the choice stage suggests that they reported their preferences truthfully in the rating stage.”* **page 16) Typo at bottom of page. "We we have not find any" Replace find with found.**

Thank you, we corrected this. **page 17) I recognize your geographical provenance, but I think most readers will prefer while to whilst. ; )**

Thank you, we updated the manuscript. **Nice paper!  
  
  
Reviewer #3: I recommended acceptance of the first revision.  
I maintain this recommendation after reading second revision and the replies to the other two reviewers. I am, however, less enthusiastic about the authors placing emphasis on "numeric vs. non-numeric" stimuli. The reason is that Trueblood et al (2013) have demonstrated the attraction effect using rectangles. This finding has been replicated by Trueblood et al in subsequent publications. I think it is more appropriate to draw a distinction between stylised vs. non-stylised stimuli (or stimuli with externally and objectively defined attributes vs. naturalistic stimuli).**

We agree with this, and changed the wording to make this distinction clear.

**A few more remarks that might be helpful to the authors:  
  
1) Huber et al., described 5 properties that inhibit the occurrence of the attraction effect. Logically, avoiding designs that have these properties does not guarantee the occurrence of the attraction effect. There is a myriad of other reasons in an experimental design that can lead to a null attraction effect result (e.g. lack of motivation from the participants, lack of understanding of the task etc). It is great that the authors took these properties into account but I do not think that this is the central take away from this paper.**

We agree that there are other reasons why the experimental design can lead to a null effect. While the five criteria served as an important reference point for our experimental design, we also aimed to minimise the effect of other potential confounds (e.g., we applied a stringent exclusion criteria to exclude participants who did not take the task sufficiently seriously, and while we cannot explicitly exclude the possibility that participants did not fully understand the task, their responses were remarkably consistent).

**2) The within-subjects design the authors employed together with the type of stimuli is, in my opinion, the important contribution of this paper.Previous replication failures in the literature have employed between-subjects designs, which are by definition less powerful (see also the work by Michael Regenweter and colleagues). This is also why I believe that the within-subjects experiments by Trueblood et al (using non-numerical stimuli) should weigh more heavily in any debate pertaining to the robustness of the attraction effect.**

Experimental power was a major consideration in choosing a within-subjects design. However, we did not anticipate that presenting participants with triplets including the same two equally highly rated options twice within a short period of time would result in a “sticky first choice”. In Trueblood’s rectangle experiment this was not an issue due to the high number of choice trials and the less memorable nature of the chosen option. For our experiment, a between-subjects design perhaps would actually have worked better.