**Editor**

Thank you very much for submitting your manuscript "A zero attraction effect in naturalistic choice" for review and consideration for publication in *Decision*. I have received reviews from three highly qualified and experienced researchers in the field (two of them actually identify themselves) and I read the manuscript carefully. The reviewers think this is an interesting and important topic and they all appreciate the thought went into your design and the effort associated with its implementation.  Although none of them thinks this version is ready for publication, their evaluations are very positive and I am happy to invite you to re-submit a revised version that addresses their concerns and requests.

Their comments are very clear, straightforward and constructive.  Some of the reviewers were not sure they followed exactly the design (the nature of the tasks) and the analyses, so I urge you to make an effort to describe things more carefully and with more details. In addition, they all suggest additional analyses that can rule out alternative explanations or possible confounds.

For your guidance, reviewers' comments are appended below. I hope that you can address these points and I look forward to receiving your revised manuscript.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.

**Reviewer 1**

Reviewer #1: There are a number of novel and valuable characteristics in this study. It raises the interesting challenge of assessing whether the attraction effect for naturalistic stimuli occurs if the requirements in Huber et al (2017) are satisfied. The authors' selection process for appropriate groups of test movies is outstanding. The researchers are able to identify AB pairs with ratings more than 3 on a 7-point scale that are equally rated by each individual. The A', B' decoys are similar to their A and B targets but have ratings at least 3 units less. Finally, the authors also do a good job screening inappropriate subjects whose ratings were performed too quickly, with too much breadth, or having too extreme autocorrelation with previous choices. It would be interesting to test whether the results change given various levels of inappropriate responses.

We thank the reviewer for his comments. We conducted a one-sample t-test on the raw dataset (no exclusion criteria applied) to test if the proportion of trials where the target was chosen was significantly higher than 0.5, but we found no evidence for the attraction effect in this dataset either, t(151) = -.72, p = .764, 95% CI [.48; .51].

The results are very consistent, perhaps too much so. Overall, the choices between A and B are almost equal, despite having seen the same A B choice in the context of different A' or B' decoys. It is unfortunate that the tests of similarity and familiarity have no effect, given other researchers have found that these variables moderate the attraction effect. It is important that that less than 5% choose the decoy, showing that choices are largely consistent with large differences in ratings. However, for those who do choose the decoy, a similarity effect could draw shares from the target, thus limiting the attraction effect. Thus, it is important to demonstrate that the results do not differ if the authors drop the 5% of observations who chose the decoy. That test is unlikely to matter, but it is still needed.

In our tests of the attraction effect, we have already excluded trials where the decoy was chosen. We have now clarified this in the Exclusion criteria section.   
  
There remains a more substantial problem that could invalidate the general conclusions from the study. Fortunately, there are two simple tests that could determine if the problem exists. The choice design specifies that each respondent sees an A B B' and an A' A B choice set within the same set of tasks. If a respondent reasonably repeats the first choice between A and B despite the presence of a different undesired decoy, then the average attraction effect will be zero for that respondent, one choice counting for, and the other counting against the attraction effect. To test that possibility, two tests are necessary.  
  
1. Run the analysis only on the first choice made on the A B pair. If recalled repetition is substantial, then the attraction is likely to be positive on the first, but negative on the second choice.

2. Run an analysis on all those who switched A B across the two trials. Contrast the proportion of switches that are in the direction of the attraction effect against those which reverse it.

We thank the reviewer for suggesting these analyses. Our results show that participants were unlikely to switch from their first chosen movie. This pattern could either arise from participants having strong prior preferences over the target and competitor (although we think this is unlikely given equal preference ratings for the target and competitor), or committing to their first choice out of consistency. In either case, the analysis of the first trial for each A—B pair shows that the attraction effect is not being masked by participants not switching from their first choice.

**I am not sure how to incorporate Neil’s comment about the ABA’ and BAB’ design and not sound way too defensive – I am not sure if he was having a dig at our design in the first place?**

“*Our experimental design ensured that for each bespoke A B movie pair, participants were presented with both A, B, A' and B, A, B' triplets. Faced with two subsequent choices involving two equally highly rated A--B movie pairs and two different, but undesirable decoys, it is possible that the first choice is “sticky”, and will be repeated. If this is the case, then we can expect that the target and the competitor will be chosen exactly half of the time, resulting in a perfectly zero attraction effect.*

*Indeed, participants were overwhelmingly likely to stick with their first choice, as they only switched between A and B for 8.5% of A--B pairs (out of the 990 bespoke A--B pairs where the decoy was not chosen on either of the trials). Out of these 84 occasions when participants switched, 48 times they have chosen the target both times, which means that we found no evidence that the proportion of switches towards the attraction effect (i.e., choosing the target on both occasions as opposed to always choosing the competitor) was higher than 0.5, Χ2(1, N = 84)=1.44, p=.115, 95% CI [.46--.68].”*

*“Even though participants tended to stick with their first choice, the attraction effect might still be evident on the first choice for each bespoke A--B pair, and if this the case we can also expect a ``reverse'' attraction effect (higher likelihood of choosing the competitor) on the second choice. After excluding individual trials where the decoy was chosen, we found no evidence that the proportion of trials where the target was chosen was different from .5 on the first, t(134)=-.039, p=.969, 95% CI [.47--.53], or second choice, t(134)= .015, p=.988, 95\% CI [.46--.54].”*

If the attraction effect is still not significant given these two tests, then the paper is an important contribution provided those provided those tests are included. However, if a reliable attraction is found, then there is a valuable paper that identifies the magnitude of the attraction effect and contexts in which it is stronger or weaker. In particular, it may happen that attraction effect becomes diminishes where target-decoy similarity is less or where the respondent are more familiar with the stimuli.

To control for the effect of these potential confounds, we included these variables (familiarity with the movies/target-decoy similarity rating) in our main regression model (Model 2 in Table 1).  
It is also important to run the analysis across all the data while adjusting errors to account for within person association. Currently, it is not clear precisely how the authors performed their analysis.

To account for within-subject variability, we used a mixed effects regression framework with subject-specific intercepts. We have now made this more explicit in the text.

**Reviewer 2**

 In this paper the authors asked whether using naturalistic stimulus elicits the so-called attraction effect. The design of the study is motivated by mixed views on the robustness of the attraction effect and in particular a claim made by Frederick et al. (2014), according to which the attraction effect occurs only when options are described with numerical attributes. In response to that claim Huber et al. (2014) described another set of criteria that need to be met in order to test for the attraction effect. In the present study, the authors attempt to satisfy both the criteria of Frederick et al. and of Huber et al. and report a zero attraction effect.  
  
Overall the study is both timely and will interest a broad audience. Furthermore, the authors follow an extremely careful procedure in order to construct their choice sets. I believe that this paper will contribute to the ongoing debate regarding the robustness of the attraction effect. However, I can see a few limitations of the current study while I have some requests for further analyses. The latter are needed in order to see if the data consist of a mixture of "repulsion" and "attraction" effects, even within participants, that are determined by different factors (e.g. distance between target and decoy, preference for competitor and target).  
  
Limitations:  
  
1) The authors meticulously elicit similarity and preference ratings for different Netflix moving. These ratings are used in order to construct target-competitor-decoy triplets. Regarding preference ratings, the underlying (reasonable) assumption here is that if two movies receive equal ratings then participants should be indifferent between them. However, I am unclear if this works well in practice. In particular, if choosing between two dissimilar movies of equal ratings would yield 50%-50% choice probabilities. My concern is motivated by the oftentimes discrepancy encountered between judgment and choice experiments but also by the fact that evaluating or comparing dissimilar naturalistic stimuli, such as movies, may engender different cognitive processes. For instance, when asked to rate a movie participants may arrive at the rating by comparing the movie at hand with all other movies within the same genre (e.g. this is a very good action movie). But when asked to compare an action movie and a thriller, their choice could be guided by their overall preferences for one genre over the other (e.g. I strongly prefer a very good thriller over a very good action movie). Ensuring correspondence between equal ratings of dissimilar naturalistic stimuli and choice indifference is necessary in order to claim that the Frederick's criteria are met. This correspondence can only be assessed experimentally. Finally, asides the aforesaid criteria, if preferences elicited from ratings and from choices are decoupled this would undermine the appropriateness of the design in studying the attraction effect in the absence of binary baseline choices between target and competitor.

“Our experimental design relies on the assumption that participants should be indifferent between equally highly rated movies from different genres. However, if the cognitive processes underlying the evaluation and choice stages are different, then discrepancies between ratings and choices might arise. For example, it is possible that the rating reflects preference for the movie within its genre category, but the overall choice between two movies with different genres is driven by overall genre preferences.

To address this concern, we first tested whether overall genre preferences have an influence on choices over and above the information reflected in individual movie ratings, by including the difference between the average target–competitor genre ratings for each participant in our regression model described in Table 1 (Model 2), assuming that average genre ratings serve as a suitable proxy for overall genre preferences. Model 3 in Table 2 in the Appendix shows that a stronger genre preference for the target over the competitor slightly increases the probability of choosing the target: a one standard deviation increase in the difference of the average target-competitor rating (corresponding to a 0.9 unit increase in average ratings on the original rating scale) increases the odds of choosing the target by about 10%.

Second, to determine whether overall genre preferences influence the strength of the attraction effect, we conducted two tests. First, we added a binary explanatory variable in Model 3, capturing whether the option with the higher overall genre rating was chosen on that particular trial (this is the case in 53% of trials). The results from Model 3 in Table 1 show that choosing the option with a higher overall genre preference did not affect the probability that the target will be chosen. In addition, a one-sample t-test on a subset of trials where overall genre preferences for the target and competitor are roughly equal (where the absolute difference between the average genre ratings was less than 0.25 – about 23% of all trials) shows that the proportion of trials where the target was chosen is not significantly higher than .5, t(98) = 􀀀:142, p = :556.

Overall, while there is some evidence that overall genre preferences slightly influenced choices between the target and competitor, our results suggest that this had no effect on the strength of the attraction effect.”

2) The claims made by Frederick and Huber et al. may not be up to date given more recent experimental results. For instance, Spektor et al. (2018, Psych Science) show that an attraction effect with non-numerical stimuli is obtained only when the alternatives are horizontally aligned (Fig. 5). This speaks to the possibility that when attribute-wise processing is facilitated then the attraction effect ensues and strongly contradicts Fredericks's claim. Interestingly, in all other experiments in Spektor et al., in which the rectangles are not aligned, a repulsion (negative attraction) effect is obtained. It seems, thus, that it matters for the attraction effect whether people engage in attribute- or alternative-wise processing. With naturalistic stimuli, different people may have different strategies which overall gives a zero attraction effect. The paper should reflect the current state-of-the-art (e.g. Spektor et al) beyond the claims made by Frederick et al. and Huber et al.

“The first multiattribute choice experiments demonstrating the attraction effect (e.g., Huber, Payne, & Puto, 1982; Simonson & Tversky, 1992) almost exclusively used stimuli presented as a set of numerical attributes (e.g., cars presented as numerical values for gas mileage and ride quality). Trueblood, Brown, Heathcote, and Busemeyer (2013) have also found evidence for the attraction effect in a perceptual choice experiment, where participants were asked to select the largest from three rectangles with varying widths and heights. However, recent research suggests that the attraction effect might only occur under very specific conditions. In particular, it had been shown that the effect is much more likely to occur when an attribute-wise comparison strategy is employed in the choice process as opposed to an alternative-wise strategy (Noguchi & Stewart, 2014). One implication of this mechanism is that the strength of the attraction effect seems to be highly dependent on the exact presentation format of the numerical or perceptual choice options (e.g., Spektor, Kellen, & Hotaling, 2018; Cataldo & Cohen, 2019).”

Further analyses:  
  
1) The probability of choosing the decoy is low, perhaps too low in comparison to other studies. It is thus an open question whether the decoy was placed way too far from the target, rendering the manipulation ineffective. I recommended plotting the magnitude of the attraction effect against the probability of choosing the decoy (and perhaps the similarity of the two based on the ratings) in order to see if there is any regularity there. The possibility that the decoy was too inferior to generate a preference reversal should be discussed. The mixed effect model ran with similarity ratings shows a lack of effect, but I recommend showing this relationship also descriptively (there might be non-monotonic patterns).

? the attraction effect against the probability of choosing the decoy – any suggestion what he/she means?

“The decoy was chosen very rarely, in less than 5% of trials. Previously, it had been shown that a decoy that is placed too far from the target can result in a reverse attraction effect (repulsion effect; Spektor et al., 2018). On a 1-7 preference rating scale, we allowed for a minimum distance of 3 and a maximum of 6 between the target and decoy. While we have not find any evidence that the target-decoy rating difference influenced the strength of the attraction effect (see Model 2 in Table 1), a non-linear association between target-decoy preference and the attraction effect might still exist. To examine this possibility whilst controlling for the perceived similarity of the target-decoy pair, we ran a logistic regression model with target chosen as the outcome variable, and target-decoy rating difference and their perceived similarity as explanatory variables, and estimated the effect for each level of these explanatory variables. Our results do not offer support for the hypothesis that the attraction effect is sensitive to the preference difference between the target and decoy. Figure 6 in the Appendix shows the predicted probabilities from this analysis.”

2) How did the preference ratings of the target-competitor (4-4 vs. 7-7) influence the attraction effect? Even if this is part of the logistic model (not sure if it is) I would also recommend to plot this relationship descriptively.  
  
“While not part of our original regression model, it is possible that the strength of the attraction effect is influenced by the overall preference for the target and competitor (this is at least 4 and at most 7 in our experiment). To explore this possibility, we calculated the proportion of trials where the target was chosen for each level of target-competitor preference rating. Figure 7 in the Appendix shows these proportions, suggesting that target-competitor preference ratings had no effect on the strength of the attraction effect.”

**Reviewer 3**

Unsurprisingly, I believe the conclusions of this paper, though I have some reservations. Further, since it bears so heavily on *this* work, I decided to share (almost in full) exchanges I had with reviewers, associate editors, and editors in the long road to getting *our* paper [Frederick, Lee, & Baskin, 2014] published. (In the first submission, the *objections* are in **red** and our *responses* are in black. In our second submission, the *objections* are in black and our *responses* are in **red**.)

p. 2) Regarding our experiments, you write: “*participants could even sample the choice options (e.g. squash, mints, popcorn*).” Although the statement is correct, none of the *examples* are correct, and I’m not sure how you came up with them. (Indeed, I’m not even sure whether squash refers to the vegetable or the sport.) We had people actually taste kool-aid and actually eat jellybeans and actually feel paper towels and tissue. The remaining stimuli were realistic, but not real.

We thank the reviewer for pointing this out. In part this results from linguistic differences, e.g. kool-aid is not sold in the UK, but instead we have a a similar category of concentrated drink flavouring products called squash. We have now corrected this section of the paper to accurately and more globally represent the stimuli used.

“These experiments included choice options with numerical attributes, as well as complex, real-world stimuli (e.g., fruits, bottled water, apartments, etc.), and in some of these experiments, participants could even sample the choice options (e.g. Kool-Aid, facial tissue, jelly beans).”

p. 2) I thought the two published objections to our work were weak and I bristle at your genuflection to them here (though I do understand the rhetorical strategy of saying “I’ve heard you,” as well as your desire to differentiate what you did from what we did). I don’t think our studies *were* flawed in the ways these commentators suggested (see my defense in the exchanges below). Accordingly, I object to sentences like “*results derived from flawed experiments will not advance our scientific understanding*.” and attendant suggestions that our work wasn’t “rigorous” or “stringent.” Similarly, I object to the place in the General Discussion in which you write: “*Our experiment is the first investigation to rigorously test the attraction effect with naturalistic stimuli whilst also addressing all of the criticisms raised in connection with Frederick et al.’s experiments*.” Again, though I acknowledge our stimulus selection/development was more casual than your laudably sophisticated procedures, I think our experiments were mostly just fine and the Huber/Simonson objections mostly just silly. Thus, I don’t think it is accurate to say you are the first to do this. Indeed, we didn’t even claim to be the first: five other papers before us also used naturalistic stimuli (although we could not replicate those results).

*Removed all references to primacy, deleted “flawed” part and toned down the rest.*

*My attempt at objectively describing his work:*

*I think we should keep the bit about us being first. We are not saying we are the first to use naturalistic stimuli, we are saying we are the first to meet Huber et al’s specifications.*

“These two studies sparked considerable interest amongst decision making researchers, and led to the re-examination of the boundary conditions of the attraction effect. While the results from these studies are consistent in showing no evidence for the attraction effect across a wide variety of naturalistic choice options, the degree to which the individual experiments presented in these studies invoked an attraction effect-type choice scenario, and thus constitute a stringent test of the attraction effect has been subsequently questioned.”

“Our experimental design was carefully developed to test the attraction effect with naturalistic

stimuli whilst avoiding the five critical conditions set out by Huber et al. (2014).”

p. 4) It is awkward to write “movies that are part of the same sequel” (a sequel is single movie). You might say “part of a ***series***” p. 4) You write “it is unclear how Frederick et al. decided which movie should be the target.” Let me clarify, then. The customary nomenclature in tests of AD involve the *target* and *competitor* as the two “legitimate” options that one might sensibly choose and the *decoy* as the option which most closely resembles, but is inferior to, the target. Hence the identity of the decoy essentially defines the identity of the target. The assignation is quite clear for nearly all our stimuli. Further, while we did *assume* that the sequel functioned as the decoy in our stimuli, it is important to note that the data support the assumption (it is rarely chosen).

“However, with naturalistic stimuli, this task is significantly more complicated. Frederick et al. (2014) have also used movie stimuli in two of their experiments: they chose pairs of movies that are part of the same series or are starring the same actor (but have distinctly different genres) to create target-decoy pairs. In these experiments, the role of each of the three movies (target, competitor, decoy) was always the same for all participants, and based upon population average ratings rather than individual ratings.”

p. 6) It took me a very long time to comprehend what you are trying to do in stimuli creation. The description in the Experimental procedure question seems much clearer than the description that precedes it. By the end, I liked what you did, but I think/hope it can be expressed more simply.

*Added summary para before long, detailed description he did not like:*

“To determine the choice sets used in the experiment, the first step was to create a set of quadruplets, each consisting of two movies that are very similar (for the two target-decoy pairs; i.e., A–A’, B–B’), whilst making sure that the two pairs are overall sufficiently different from each other (for the target-competitor pair; i.e., A–B).

The creation of these quadruplets can be described briefly as follows. Given a set of 400 movies, we used their genre information to determine their pairwise similarities, and only kept the 3,011 most similar pairs. We first manually reduced this list to only include movie pairs we perceived as similar (1,242), and then further reduced the number of pairs to the most similar pairs (253 movie pairs comprising of 231 individual movies), based on independent similarity ratings. To create the final quadruplets from these target-decoy pairs (and determine the target-competitor pairs), we simply paired up movie pairs that had very little genre overlap – resulting in 20,022 unique quadruplets.

The details of the construction of these quadruplets are somewhat arbitrary—a different recipe could have been used. However, the main point is that the choice triplets created from these quadruplets pass Huber et al.’s (2014) criteria, as we detail below.”

p. 8) “*We did not collect any data about the demographics of our sample, as we did not expect it to affect our results*.” Can I just say how refreshing it is to read this?! I commend you on having the common sense and courage to just say this. Good for you. I hope it catches on.

p. 12) Can you clarify what you mean when you write that “*we hypothesize that spatial separability of the attribute dimensions might be a key role in the comparison process*.” Also, although you’ve obviously read our paper, it seems like you haven’t when you write: “*Future research could test this hypothesis this* [sic] *by exploring how the strength of the attraction effect varies with different representations of the same choice options (e.g. numerical, visual).*” This is literally *exactly* what we did in in our paper in studies 2a through 2c and studies 3a through 3c.

“In conclusion, our results are in line with that of Frederick et al. (2014), and provide strong evidence that the attraction effect does not extend to choice between naturalistic options. While we did not aim to investigate the exact reason behind why the attraction effect is robust in choices involving numerical attribute dimensions but is absent from choices involving naturalistic options, we hypothesize that the separability of the attribute dimensions is an important factor. More specifically, while there is strong evidence that attribute-wise comparison strategies are key to the attraction effect in numerical or perceptual choices, such comparison strategies are less likely to occur with complex, naturalistic objects. Future research could test this hypothesis by exploring how the strength of the attraction effect varies with different representations of the same choice options (separate attributes versus naturalistic representation). Results from such experiments could provide us with important insights about the boundary conditions of the attraction effect, and the cognitive process underlying this choice bias.”

In general, despite my predisposition to like this paper, and my admiration of some of the design features, I confess I have a hard time following exactly what was done. Maybe that requires a diagram or something. If I understand your procedure correctly (and I’m not sure I do), respondents first encountered the {A,B,b} triplet and later encountered the {A,B,a} triplet and you examined whether the nature of the decoy (a or b) affected the choice between the core options (A or B). Even with the many interceding “filler” choices, this seems like a weak test of the potential power of the decoy option, in much the same way that it would be a weird test to ask respondents: Which do you prefer: A or B? and, then immediately after, Which do you prefer: A or B or b? That said, I think there is an opportunity here for *both* a within-subjects analysis (seeing whether a person changes their choices with different decoys) *and* a between-subjects analysis (seeing whether people, *on average*, make different choices with different decoys). Assuming item order was randomized, I think both types of analysis are possible, and I would like to see both.

I think we’ve done the within subjects analysis, but not sure what the means by the between subjects analysis? Any suggestions?