

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.

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).

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).

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.

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 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.

Sincerely, Shane Frederick

First submission

AE Report

I have some difficulty seeing how each product class was really even a test of the attraction effect in the first place. For example, with apartments, the decoy only differed from the target in that it offered a view with a cloudy sky.

As you suggested, we now reproduce the stimuli from these studies in Appendix B; the first set of stimuli reproduces the choice set in question, which involved apartments varying in size (specified numerically in square feet) and view (depicted with photographs). One apartment was small (530 square feet), but had

a very attractive view, of what appeared to be the ocean shoreline. The target apartment was larger (910 square feet), but with a much less attractive view. The decoy was dominated by the target on both dimensions – it was slightly smaller (905 square feet) and the view appeared even worse (essentially the same view, through a dingier window). We referenced this as “cloudy sky” in the corresponding table. We understand and apologize for the confusion created by our odd choice of nomenclature.

Does this make the other option dominate this option? It’s really hard to say.

Relative to the target, the decoy apartment is worse on both of the specified dimensions: it is smaller and has an even worse view. So it seems rather clearly dominated. Moreover, respondents apparently recognize this, because only 2% of them choose it.

Would someone really take this into account in his or her decision?

48% of respondents chose the option we intended as target, compared to only 2% who choose the option we intended as the decoy. So it seems safe to conclude that respondents are accounting for this in their decision.

For fruit, what did the pictures look like? Were the black spots and dented skin apparent? Or, were they so significant that they resulted in feelings of disgust and possible contamination of the other orange?

The pieces of fruit were depicted using color photos, now reproduced in Appendix B. The decoy orange had two spots of discoloration. The decoy apple appeared to be bruised. So, yes, the imperfections of the decoy options were apparent. As intended (and as reported in Table 2), the decoys were essentially never chosen. We can only speculate what feelings the imperfect fruits evoked. To our surprise, their presence had no effect on the choice share between the two “normal” fruits – that is, we found neither attraction nor repulsion.

If the orange really looked disgusting, and you don’t find the attraction effect, this would be like telling a manager to use an inferior TV, but don’t bring in a TV that is smashed up with insects crawling out of it.

We now reproduce the materials in Appendix B. You can judge for yourself whether the orange was disgusting or merely inferior to the more perfect orange. With respect to your question regarding the sale of television sets, if a broken, insect-infested television can be regarded as an analog to a discolored orange, we would tell the manager that adding that option to the product line is unlikely to increase sales of the normal television – a message consistent with the one advanced in our paper.

For hotel rooms, the decoy was a room with "nice" décor and a polygon bath? Is a polygon bath really supposed to be inferior somehow?

Of course, our verbal summaries were impoverished versions of what respondents actually encountered – shortened so as to fit into a table. We now reproduce all of the visual stimuli in appendix B. With respect to the stimuli in this study the presence of a hot tub was one salient indicator of quality. The two expensive rooms (the target and the decoy) each possessed one, the cheaper room did not. Among the two rooms having one, other visual features were intended as cues for quality. We intended for the decoy to appear less fancy than the target option and, hence, to be dominated, since their price was the same (\$180 per night). Respondents largely interpreted the stimuli as we intended, since the target was chosen much more frequently (67%) than the decoy (13%).

The repeated objections to our stimuli appear to overlook two key features: (1) the decoy was self-evidently more similar to the target than to the competitor in every case; and (2) the decoy was dramatically less popular than the target in every case, as Table 2 reveals. Thus, a vast majority of respondents are, in fact, appreciating the intended differences: there is no other way to account for the data. Moreover, even when the experimental design was not fully successful (i.e., when the fraction preferring the decoy is non-negligible), this at most creates difficulty in interpreting the change in choice share, and thereby diminishes the evidential weight of that study. It in no way invalidates the general approach or the collective implications of the remaining studies.

I could go on like this for most of the stimuli used in Table 1. Thus, before just dismissing an effect, it is important to carefully do this, by making sure that the target really dominates the decoy,

As noted above, we don't think that our stimuli are so problematic, and we hope that this is clearer with our inclusion of Appendix B.

but the decoy isn't so bad that it results in disgust (and to show us exactly what you used as stimuli).

With respect to the capacity of our decoy options to elicit disgust, we suspect very few met this threshold for most people. None of us are disgusted by a nice hotel room that is slightly less nice than an even nicer one, or by larger apartments with unappealing views, or by weakly flavored fruit drinks, or critically panned movies. Two of us would probably even prefer a flawed orange to a perfect apple. Across the 36 studies we conducted, the decoys elicited disgust in probably just one case – the jellybean study, in which one of the two conditions actually consumed *Bertie Botts* jellybeans flavored to resemble pepper, dirt, grass, and ear wax, respectively.

But, in any case, it is not obvious why the putative presence of disgust is necessarily invalidating or uninteresting. For instance, one could imagine making

the image of the decoy orange sequentially less attractive up to, and into, various degrees of disgusting. It is quite possible that the size and valence of contextual effects would respond to such manipulations. That is an empirical question, and a somewhat interesting one, though our studies were not designed to address it. (We should note that in the one study involving unambiguously disgusting stimuli – the jellybean study – the decoys actually had little effect on the choice share of the flavors whose appearance they most closely resembled.)

I think the basic point that you are really trying to make is that the attraction effect will only hold when it is obvious that one option dominates the other. But, again, is that news? I'm not so sure.

That is *not* the basic point we are trying to make. As noted above, dominance is quite clear in most cases. We are alleging that the attraction effect appears to be restricted to highly stylized stimuli that involve 2 X 2 numeric matrices.

We base this strong claim on our repeated failure to find such effects outside of these contexts using studies sufficiently powered to detect even moderate effects. In all we tested for the effect 27 times and failed to find it 27 times. We think this yields a conclusion.

At the same time, I do understand your point that if a manager tries to implement the attraction effect, it is hard to say what will happen in the real world. Will everyone really see the target as dominating the decoy? Did you do any field experiments?

We suppose it depends what you mean by “field.” We did several studies involving actual consumption of actual goods (and are one of the few to have done so). In any case, we see no basis for assuming that effects we fail to find in the lab would be present in the field.

For study 2, I guess again I'm wondering how this is different from Ratneshwar et al.

There are two main differences: (1) we are not *supplementing* an ambiguous numeric specification with a somewhat less ambiguous verbal description, but are rather *replacing* a numeric specification (probability) with its perceptual counterpart (the shaded fraction of a probability wheel). (2) We find no significant attraction effect, whereas Ratneshwar et al. continue to find significant effects.

For study 3A, I looked at Figure 2, and at first glance I could not tell the difference between Picture B and Picture C. Thus, how can we expect the attraction effect to hold if someone doesn't even notice that the target dominates the decoy (whereas this is obvious with the quality ratings)?

Though we admit the visual differences are subtle – especially if you do not use a color printer – respondents *can*, apparently, discriminate the two stimuli. Once

again, of the 60 respondents in that condition of the study, note that only 1 of them chose the decoy, compared to 21 who chose the target. If respondents truly regarded the picture qualities of the TVs as identical, you'd expect the *opposite* pattern, since the decoy is slightly less expensive (\$339 vs. \$350).

And, for study 3B, I'm afraid that you have insignificant results (no attraction effect, no repulsion effect). Thus, I'm not sure how you can really use Study 3B to build your paper.

We contend that the attraction effect does not occur if you move away from highly abstract "schematic" alternatives that amount to 2 X 2 numeric matrices. Correspondingly, insignificant results are at the very heart of our case, and, thus, finding another one here adds evidentiary weight to this claim. Going beyond the epistemic value of the null effect reported for the "choose first" condition, the contrast between this null effect and the (now significant) effect in the "rate first" condition suggests that contextual effects can be influenced by supplementing perceptual representations with (self-generated) numeric representations. However, this study was only intended to explore the contours of the boundary conditions for the effect, the results of this "rate first" condition do not bear strongly on our claims.

*We note that in the revised manuscript, the study in question is now 3C rather than 3B.

Reviewer: 1

Not original

We think our work is good ordinary science and will not try to exaggerate its originality. We are certainly not the only researchers to have used more realistic stimuli when investigating contextual effects (though across three decades of writing on the topic there are remarkably *few* such studies; see appendix A). However, two things are worth noting.

(1) We agree that Ratneshwar, Shocker, and Stewart (1987) were one of the first to examine an issue we also examine here – boundary conditions for the attraction effect. However, our manipulations, results, and claims differ from theirs in the ways we note above, in our response to the Associate Editor regarding study 2. Furthermore, if experimental stimuli were arrayed on the continuum ranging from extremely abstract/highly stylized to fully realistic, theirs would lie almost as far to the left as the typical study on this topic, whereas most of ours lie pretty far to the right.

(2) By drawing upon much more data and a much wider variety of designs, we're able to make a much bolder claim – that the attraction effect is nonexistent in many or most realistic contexts. In addition to the study designs we conceived,

we reexamine the small literature which reports significant effects in less stylized settings. In each case, we failed to replicate the reported results.

Although it is true that complex stimuli where the relations among options are hard to detect are less conducive to the attraction effect, the basic effect is still there.

(1) As noted earlier in our responses to the Associate Editor, we disagree that the relations are difficult to detect. We have little doubt that respondents can and do make such judgments; indeed, we suspect that they could often recognize the dominance structure *more* quickly in several of our studies than with the more typical *stylized* stimuli (e.g., that after tasting three samples of Kool-aid, participants likely recognize that dilute grape tastes a lot like regular grape, except less flavorful).

(2) Regarding the assertion that the “basic effect” is “still there,” we respectfully disagree. Of course, we can never prove that there are no black swans outside the lab, but across 27 studies we fail to find a single one – not even when we look in the few places in which previous researchers have claimed to find them. We think that is enough data to draw a practical conclusion, to initiate a dialogue about the practical reality of the effect, and to promote further discussion about its boundary conditions.

Of course, even those who are prepared to concede that black swans are absent in the wild might reasonably remain interested in how swans can be made black in the laboratory. Our studies speak to these people as well. Importantly, since the effect appears to require that at least two attributes be numerically denominated, our results support *tradeoff contrast* (Simonson & Tversky, 1992) as the mechanism causing the blackness of swans in the lab. Correspondingly, our results weigh strongly against other interpretations of the effect that others have championed, because none of these other mechanisms would explain the striking disparity between perceptual and numeric representations of product attributes.

Reviewer: 2

You define an important boundary condition for the attraction effect. Any researcher who has tested such context effects has found such reversals, but too often not mentioned them. The broad empirics of your paper are therefore refreshing and believable, and the demonstration of a repulsion effect could be quite important. Below are a number of suggestions for your paper.

Thanks for the kind words about our paper. We hope you find the revision even more compelling (even if we didn’t follow all of the promising leads you proposed). We address some of your comments below.

1. Be clearer on your sampling scheme for Study 1. One of the problems with the choice-context effects literature is that choices across product classes are easily run. It is thus easy to ignore those that

do not fit the expected effect, leading to results that seem reliable but may be spurious. Accordingly in your table 1 it is important to be clear on the set of product categories tested. Did you have, say, 100 different tests of attraction, most of which used numbers, but the rest did not? Alternatively, Table 2 could reflect a biased sample of the non-numerical tests performed. I assume that was not what you did, but clarity is needed.

We apologize for the confusion on this point and hope it is clarified in the revised manuscript. Table 1 is, in fact, exhaustive. It lists every study we’ve ever run that is not listed elsewhere in the paper. In other words, there was no selection. Essentially, the existing literature was the control group for studies 1a-1s. The existing literature shows many examples of sizable attraction effects using stylized stimuli (2 X 2 numeric matrices). Heeding earlier calls for more research on contextual effects outside of this paradigm, we constructed choice sets that would retain the essential elements of asymmetric dominance (i.e., a core set compared to an expanded set that included a decoy which was similar, but inferior to one of those options). Our intent was to test under which situations (if any) the attraction effect would replicate. We found no evidence of the effect in any of the studies. As noted above with the black swan metaphor, this does not of course mean that no effect *could* be found, but it is certainly very informative regarding the boundary conditions of the effect. Our other studies (2a, 2b, 3a, 3b) and our newly added Appendix C (which outlines our failures to replicate several published results) substantially strengthen the collective implication of the (non) results from Studies 1a-1s

2. Focus on the cases of significant repulsion effects. Null effects are OK, but less meaningful theoretically. Additionally, anything that makes dominance hard to immediately recognize should reduce the attraction effect, but that is far less interesting than the repulsion case. Thus emphasize repulsion more.

Using stimuli with greater realism, which involve perceptual representations of one or more attributes, we repeatedly find that the decoy either fails to affect (or even diminishes) the choice share of the target option. These results certainly have theoretical meaning, however they are labeled. When viewed under the assumptions of early choice models (e.g. those by Luce), they might rightly be called *null effects*, contrasted with the attraction *effects*. However, when viewed against the backdrop of studies finding attraction effects with stylized stimuli, our results might be considered *effects*, since perceptual representations could be seen as *affecting* choice processes: whether by facilitating some process that is absent with stylized stimuli or by impeding some process that is otherwise present (such as tradeoff contrast).

The common inclination to dismiss so-called “null” effects contributes to the problem you cite in point 1, in which the published results create a distorted representation of reality. Existence proofs of a psychologically interesting phenomenon were a sufficient justification for many of the first publications on the

attraction effect, but the message that academics and practitioners now draw from their reading of the literature has been grossly deformed by a combination of file drawer effects, false positive results, and an inadequate attention to boundary conditions. Our paper attempts to correct this. We regard this as a substantial and sufficient contribution.

3. Build up a theory of repulsion. It seems most likely when the decoy’s attributes reinforce an association (e.g., dilution with either grape or cherry flavor, duck farts with spring water) and that association generates an aversion to similar stimuli. A way to start would be to rank the size of the aversion effect in Table 2 and build up hypotheses about what is driving the results found.

We set out to delineate the boundary conditions of the attraction effect, and sort of stumbled across occasional instances of this opposing phenomenon, which we term the *repulsion effect*. These are intriguing, and a paper exploring the repulsion effect may be very interesting. However, though we understand that this is the paper that you (and the AE and the editor) want us to write, it is not the paper we want to write.

Though we admit that a sub-goal of our paper *is* to stimulate further research on the repulsion effect, for the purposes of this paper, we are mostly content to simply treat significant repulsion effects as especially strong evidence against attraction effects. We are not yet prepared to say much more and further attempts to do so would likely distract from our central message regarding the practical reality of the attraction effect.

Regarding repulsion effects, the account you offer above is essentially identical to our own thoughts on the issue. However, we will note that a casual version of such an analysis does not yield the patterns one might expect. Namely, we fail to find repulsion effects in contexts in which one might expect the aversive decoy to “taint” the most similar members of the choice set (i.e., in studies involving the consumption of disgusting jelly beans and pictures of damaged fruit).

4. You argue appropriately that both relative tradeoff rates and range and frequency effects will be less apparent when numbers are not included. However, as Wedell and Pettibone demonstrate, the major driver of attraction is quickly seeing that the target dominates the decoy. You need to test for the ease by which that dominance is perceived.

It is not so much that perceptual stimuli make tradeoff rates less *apparent*, but rather that tradeoff rates will be impossible to compute unless both dimensions are numeric. Regarding the detection of dominance, it seems clear that respondents *are* detecting dominance for the reasons already outlined in our response to the AE. However, a central message of the paper would actually remain intact if failure to detect dominance *was* the reason we fail to find evidence for the attraction effect. Namely, if dominance were always obscured in realistic settings, then the attraction effect would never occur in realistic settings.

5. Be clearer on the boundaries of the attraction effect. Attraction clearly occurs with quasi-numeric quality ratings such as star ratings. However, does it occur with clearly ordered phrases such as 'top rated' vs. 'consistently good quality?' You give an example of a fruit with blemishes as the decoy. The problem there is that one might infer the target had similar unseen blemishes. Would attraction also work fruits of different sizes? Put differently, visual images of volume or length may act like numbers. Can you test whether they do?

Though we think we've drawn a clear distinction between situations in which attraction effects are moderately common (choice options represented as 2X2 numeric matrices) from situations in which they are not *at all* common (pretty much anywhere else), further exploration along the boundary might well be warranted, and you draw some interesting test cases of precisely where that boundary falls. Note that in our response to Reviewer 1, we noted a distinction between *supplementing* numbers with verbal descriptions (e.g. Ratneshwar, Shocker and Stewart, 1987), *replacing* numbers with verbal descriptions (e.g., Sen, 1998), and replacing numbers with perceptual representations (as in several of our studies).

Similarly, using your example, you can imagine cases in which the numeric vs. perceptual distinction may be blurred. Consider a choice between two packs of postage stamps: one in which the number of postage stamps is specified and one in which the stamps are "merely" shown, but readily countable. Would these two representations yield identical choices and identical contextual effects? We are unsure. In studies 2a and 2b, we *do* find considerable differences when we compare numeric and visual representations of probability, though we do not know whether this would extend to some of the test cases you suggest. We'd conjecture that the visual representation would not yield an attraction effect, unless the stamps were in fact, counted and overtly recorded, akin to our study 3c. We'd hope to avoid having to submit an exhaustive analysis of the precise contours of the boundary we emphasize, though we acknowledge that further work along these lines is very much in the spirit of the paper we want to submit.

6. Study 2a is really a fine study, with many respondents and 2b provides a really clear case where dominance is readily apparent.

Thank you. One of the strengths of our manuscript is the large number of respondents we use to obtain a more reliable estimate of the size of the effect. We typically fail to find attraction effects despite these comparatively high powered studies.

We agree that dominance is readily apparent in 2b – and arguably even more salient in 2a, in which the decoy is dominated on *both* dimensions.

7. Study 3b is intriguing, implying that asking for numerical rating of attributes provided by the chooser will resurrect the attraction effect. How important is a numerical response? Would the

same reversal occur if respondents merely were asked to express their attribute valuations on a non-numerical line?

That is a very interesting question. Indeed, we expanded this study (now study 3c) so as to test a variant of your idea. In that study, ratings on a visual analog scale acted similarly to numeric ratings, and, thus, we report only the pooled results (see endnote #8). However, in that study there were no large contextual effects in *any* condition, so a conclusive answer to your interesting question remains out of reach for now.

Our second attempt

Dear Author(s):

Your submission to the Journal of Marketing Research, JMR-12-0061.R1, entitled "The Limits of Attraction," has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

The revised version of your paper was read by the same two reviewers and AE who were involved in the last round, and, being new to the paper, I read both this new version and the earlier one as well as both sets of reviewer comments. Let me cut to chase by saying that the call on this one is not easy, as the reactions by the review team could not be more varied, ranging from conditional accept (R2) to reject (AE). Here is the landscape as I see it. I think everyone would agree that no other "effects" paper in the BDT area has attracted more interest, controversy, and used up more lab subjects than Joel and Chris' original 1982 attraction effect paper. Unlike the experimental results leading to Prospect Theory, which one could view it as a not-totally unfriendly refinement of utility theory, the attraction effect was a result that, if true, threw a fundamental monkey's wrench in random utility theory, which was (and still is) the analytic backbone of modern IO economics and a large chunk of the marketing research field. As a result—and I remember this quite vividly—it was a result that many suspected was a lab parlor trick that would not replicate when put under close inspection. The most common argument was that since standard RUMs did quite well (thank you) in predicting choices from real-world sets, it must be something about the artificiality of the lab choice sets that was doing it, where, for example, "quality" was defined by an ambiguous numeric score. I am pretty old, and recall talking to Allan Shocker at a conference around 1984 or so telling me that he didn't buy it, and was poised to write a paper showing that all one needed to do was swap out the abstract options for "realistic" ones and sure they effect would do away. Problem was when they did that study it did NOT go away (Ratneshwar, Shocker, and Stewart '87). My sense of the current status of the effect is that most accept it as very real but also very limited, a result that plays a role in choice theory akin to empirical demonstrations of bending light in physics: it is helpful in showing the Newtonian laws (here, RUT) do not explain the universe, but the domain is sufficiently limited that for most practical applications standard theory works fine (hence Sawtooth is not ripping clients off).

I agree with your characterization of the effect, though not your characterization of the popular perception of the effect. Understandably, when best-selling books within our field speak of these massive effects in realistic contexts, prior suggestions regarding the breadth and depth of the effect are prone to be substantially exaggerated.

Against this background, like the AE, I found this paper to be a puzzle. It's quite well done, and written with a populist, revisionist, tone that underlies many recent attempts to reconsider classic social psych results (like old-age priming). But unlike those efforts the reexamination here is not a new one; it reads more as a throw-back that resurrects many of the debates that occurred in the 80s on the robustness of the attraction effect, but without really acknowledging them, or explicitly building on them. What's also odd is that the AE raised this same point in the last round, but you did nothing with it, preferring to plow on as if you were the first to seriously challenge the empirical status of the attraction effect, and the first to conclusively debunk it.

I don't think this is accurate. I believe there are five papers that can be considered serious attempts to examine the boundary conditions of the effect: Kivetz, Netzer, and Srinivasan, 2004; Ratneshwar, Shocker and Stewart, 1987; Sen, 1998; Simonson and Tversky, 1992 and Trueblood, 2013. RSS antedates the others, but the other papers are just as much a test of the boundary conditions of the phenomenon – indeed, the stimuli used by RSS represent a comparatively modest departure from the typical studies with highly stylized stimuli. Those other studies claim that the attraction effects persist in contexts departing *further* from the historical benchmark than RSS itself.

In this sense, I find it odd to dismiss the evidentiary value of our paper by saying that “RSS already did this.” Consider the following analogy. Study 1 shows that mixing in some breast milk with formula reduces autism. Studies 2-5 show that breastfeeding has no effect on autism. Study 6 shows that breastfeeding eliminates autism. In the context of studies 2-5, study 6 retains evidentiary value, even if study 1 had earlier found results that are qualitatively consistent with (though weaker than) study 6. Moreover, to the extent that design flaws in Study 1 limit its evidentiary value, study 6 is even more important. We did not dwell upon the flaws of study 1, because we think it coincidentally arrives at a qualitatively correct conclusion, though I don't think it goes far enough, nor does it emphasize what we see as the essential distinction. More on this later.

If you may permit a grandiose analogy, I feel like the AE repeatedly upholding RSS as grounds for rejecting this paper is a bit like rejecting Prospect Theory on the grounds that Markowitz already mentioned something about reference dependence in a prior paper. Surely, there is room for more than one paper on a topic. We succinctly summarize a massive amount of data using a broad array of

stimuli, which represent a much more aggressive departure from the status quo than RSS.

Now, what you do here is much more comprehensive than what anyone has done before, but rather than use this impressive arsenal of work to offer a more nuanced (or clarified) view of when the attraction effect works and why, you opt for the more blunt—and less helpful conclusion that it simply is “not true”.

Though inaccurate, that would not be an altogether inappropriate pragmatic takeaway. That said, this is not the conclusion we draw or report. With stimuli involving 2X2 numeric matrices, we report significant effects in 4 of 5 cases. Though the distinction we emphasize (perceptual vs. numeric) tends to correlate with “realism,” it is distinct. For instance, unlike most choice options, gambles can be represented with perfect *realism* using only numbers. We find strong attraction effects here. The effect is eliminated when one of the dimensions (probability) is represented perceptually (as the visual area of a probability wheel). However, in this case, at least, we don’t believe that is because the stimuli are more “realistic” – indeed, I think they are less so, as the probability wheel representation is rarer in the real world than the numbers themselves. Of course, in most of the remaining stimuli, perceptual representations correspond with greater realism.

As point of comparison, have a look at what Scheibehenne, et al. (JCR 2010) did with their reconsideration of Iyengar’s choice overload effect, which also does not always replicate. What they did was rather than just saying, “it’s not true”, they tried to reconcile the work that had been done via a meta analysis, pointing out that the strength of her effect was extremely rare. I thought you would be doing something like that here---trying to use the lab studies to resolve just what it is about some set ups that allow the effect to be observed and what one needs to make it go away.

We will look at that paper as a touchstone for some things we might do differently. I am familiar with it, though have only skimmed it. Uri Simonsohn has a nice analysis of the studies reported there, suggesting that those showing the celebrated effect are p-hacked, whereas those showing the more expected effect are not. I might conclude from this secondary analysis that there is no effect to decompose; just file drawers and publication bias.

I am less cynical about the attraction effect, since we *do* find the effect using non-bizarre stimuli, such as gambles. But the effect is certainly much narrower than the way it is typically characterized. Certainly, clearer specification of boundary conditions is something toward which to strive. I think this paper makes some progress, but leaves room for more.

“More realistic” is clearly too blunt (e.g., Ratneshwar, Shocker, and Stewart ’87).

For example, one of the long-standing assumptions is that for the attraction effect to hold there must be ambiguity or uncertainty about the meaning of (or preference for) at least one of the attributes; it doesn't work for money, nor would it work for vivid, familiar, stimuli like preferences for views from apartments.

I disagree that that is a widely held assumption, and disagree that the typical characterizations of the effects suggest the boundary condition we posit. Nor do I think it perfectly maps onto the data. Upon receipt of the data, the AE and R1 offer strident post-dictions that the results were predictable, but many of the stimuli were constructed in response to asking champions of the effect to design stimuli that would show it, with the constraint that no more than one dimension could be numerically specified. Many candidate stimuli were offered up, though we've not yet found one that yields an effect. In some cases, even *these* people later find reasons why these stimuli weren't right for testing the effect urged us to run a 28th study with their candidate stimulus du jour. Moreover, even if we cede n examples as being, for some reason or reason(s), inappropriate for testing the effect, the remaining $27-n$ studies still require an account.

Likewise, as the AE suggests (also R1), it is important to wear the hat of someone on the other side, someone who believes the many studies that seem to show the effect. Are you sure that, in fact, that people actually perceive the dominance in all these studies?

Yes. I think there is no other sensible account for the overwhelming preference of the target over the decoy.

So here is where I think we are, which is basically where we were last time. There a couple of reasons that I would ideally like to see some version of this paper eventually make it in print. There is a widespread sense that we should be encouraging more replication work in our field, and this paper does that, and does it reasonably well. Publishing it would send a nice signal. Then there are the studies themselves, which represents an impressive pile of empirical evidence on one of our field's most well-known effects. It ought to be in print. All of that said, I agree with both reviewers and the AE that in its current form the paper does not appropriately acknowledge the long history of scholarship that has tried to do the same thing. One might argue that the endeavor is novel because there has been only one major published paper that has tried to look at the effect of using meaningful stimuli, but think about it: once the RSS paper was published, it would have been hard for anyone else to publish the same thing again (it's hard enough getting a replication study published, much less a replication of a replication). You are not the first to do this, not the first to wonder why we see the effect in the lab with numerical quality scores, but choice models do not seem to suffer from it when predicting real-world choices of toasters.

Well, I think the autism analogy I drew earlier is relevant here. I don't think our paper quite settles the matter either, though it makes

a substantial advance.

Regarding RSS, I think it needs to be considered in the context of the other papers that have examined boundary conditions (around six such papers including ours). I'll reprise the questions I raised in my email to you: (1) Is that paper enough? (For someone interested in this phenomenon and where it works and doesn't, does this paper provide all the answers?) (2) Are *those* results replicable? (3) Are the designs there appropriate? (4) Is the authors' account of their results accurate? (5) Are the manipulations conducted there "essentially" the same as those we use. One can, for instance, distinguish between *supplementing* numbers with verbal descriptions (e.g. Ratneshwar, Shocker and Stewart, 1987), *replacing* numbers with verbal descriptions (e.g., Sen, 1998), and replacing numbers with perceptual representations (Frederick, Lee, and Baskin, 20XX).

We've actually conducted some direct and conceptual attempts to replicate RSS. I will have to excavate the details, but the upshot is that they don't look especially similar. I'm not convinced that *supplementing* numbers with verbal descriptions typically has the same effects as the sorts of manipulations we use. For reasons I can elaborate upon later, I don't believe their effects are, in fact, caused by enhanced meaningfulness; I don't think that verbal descriptions appended to the numeric summaries actually enhances meaningfulness very much in that context. Respondents get it; these are three orange juices that can be ranked in quality and price; the higher priced ones with higher ratings presumably taste better and the supplementary verbal descriptions confirm this – though they have other confounding effects, including suggesting that a property of the cheapest orange juice is shared by the intermediate one.

We avoided extensive reference to this paper because I think it is the worst of the five that precede ours, because it uses an unnatural manipulation that essentially preserves the 2X2 numeric matrix from which we wanted to depart, because it uses a manipulation that we believe would be largely ineffectual save for an artifact caused by an inappropriate design, and because the results don't really replicate anyway. Furthermore, as noted earlier, it seemed odd to spend time critiquing this study specifically, when subsequent studies using *stronger* manipulations (i.e., more realistic stimuli) appear to show that the effect was *not* attenuated. Of course, we'd be happy to more explicitly acknowledge the temporal priority of this work, and to discuss and critique it at length. The later studies were more pertinent to our central interest, because they represented more substantive departures from the typical design, and, thus, a more ambitious exploration of the envelope of the effect.

To be a contribution to the literature in this area you need to build a more

explicit case for how the findings BUILD on what has been done before. It's not enough simply to say, "it doesn't work for window views". Why doesn't it? Or, alternatively, what is it about numerical quality scores that causes it to work? Go back to the physics example. As I mentioned why people think the AE is important is not that it discredits standard modeling practice, but rather that gives an insight into a nuanced feature of the psychology of choice that standard field data would be too coarse to pick up. It seems that with the vast amount of data you have here you ought to be able to tell us much more about what the original AE designs were saying about choice psychology than what we had before.

We speculate about the reasons, but fail to isolate the precise psychological mechanisms. I agree the paper falls short of our ideals in this respect – perhaps due to lack of imagination, though certainly not from lack of effort or lack of familiarity with similar work. But I think the paper goes a long way toward establishing the contours of the boundary, and that it will foster further work to help settle their exact location.

Hence, while there has not been great convergence in this round, I would nevertheless like you to try again to write a paper that addresses the concerns raised by myself and the review team. Ideally it's a paper we would like to publish, but you have to do your paper by writing a work that truly builds on prior efforts, not works in parallel to it.

I think it *does* build on prior results (though sometimes on the location where they've been razed). This seems like a good departure point for a conversation (though I'll be very sleepy).

Regards, Shane.