**Review of “Revisiting diversification bias and partition dependence: Replication and extensions of Fox, Ratner and Lieb (2005), Studies 1, 2, and 5” by MeiYee Li and Gilad Feldman.**

Comments by Craig R. Fox:

I appreciate very much the interest and efforts of the authors in attempting a replication of our studies from the 2005 paper as a “registered report.” This is an important service to the field. Of the six studies reported in our original paper, one (Study 4) has had failed a literal replication attempt, but was subsequently conceptually replicated. A second study (Study 1) has been previously successfully replicated. The authors chose to attempt to replicate Studies 1, 2, and 5 (Studies 3 and 6 would be a little more difficult to replicate using an online sample of convenience). The authors also propose to collect data on potentially related constructs (desire for choice diversity as a possible predictor of the extent of diversification; patriotism as a possible covariate of partition dependence involving domestic vs. international charities).

Due to heavy teaching and service responsibilities this month I’ve only had a moment to quickly review the Stage 1 manuscript. I provide my major comments and specific comments below.

**MAJOR COMMENTS:** Overall I don’t see any fatal flaws with the authors’ proposal. This said, at the risk of coming across as (prematurely) defensive, I have a couple of general concerns/suggestions:

1. **Review of literature.** I think it may be important to contextualize the Fox, Ratner & Lieb (2005) paper within a larger literature on partition dependence. Although, as noted in the manuscript, there have been few attempts that I am aware of to literally replicate our findings, there are probably dozens of studies by now (many from my lab group and many from other groups) that *conceptually* replicate findings of partition dependence across a wide range of guises and contexts. The reason I bring this up is that a Bayesian reader will want to know how strong our prior beliefs should be that these phenomena are robust. I’d be happy to provide references, but for now I’ll just suggest that the literature review might be broadened, and it might acknowledge that these kinds of findings are not limited to the studies and paradigms outlined by Fox, Ratner & Lieb (2005).
2. **Methodological differences.** It is often impractical or impossible to conduct a literal replication of studies originally conducted in a specific location long ago. This said, while I expect that the first-order findings from our studies will prove robust to small methodological variations, I do think that some of the variations could possibly prove problematic. I presume that mTurkers completing three studies back-to-back-to-back in 2022 online will have less personal connection (and devote less concerted attention) to the content of these surveys than did Duke students in-person each responding to a quick single pencil-and-paper study in the early 2000’s. For instance, family incomes have increased (affecting Study 1), charities may be less familiar (affecting Study 2), wines are older (affecting Study 5). I outline more such variations that could be an issue in my specific comments below. Mind you, I could imagine everything replicating despite this (honestly, I don’t recall there being much that didn’t work in our file drawer); however, I also wouldn’t be shocked if some differences in methodology end up making a difference here. For example, there might not be enough variation in familiarity among 2022 mTurkers with wines from 20 years ago to replicate an expertise effect that we observed among Duke graduate students 20 years ago. At very least, I’d like to see the authors acknowledge a little more explicitly the methodological differences. I imagine that if everything *does* replicate as expected despite these limitations then that only makes the results more convincing, no? And if everything does *not* replicate perfectly these variations may provide clues to previously unknown moderators worth investigating in future studies.
3. **Correcting the record.** In a couple of cases I think there are mistakes or important omissions in citing prior literature. See specific comments below.

**SPECIFIC COMMENTS:**

1. p.8, background: FYI: I first used (“coined”) the term “partition dependence” in print in Fox & Rottenstreich (2003), in the context of judged probabilities.

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science*, *14*(3), 195-200.

1. pp.9-10, the film study cited at the bottom of p.9 is something that we successfully ran years ago (and alluded to in the discussion of Fox et al., 2005) but didn’t end up including in the Fox et al. (2005) paper writeup (we had more material than we needed). I surmise that the authors must have received the original materials for that study from Rebecca Ratner. But mentioning it here in the paper could be confusing to readers.
2. p.10, par. 3: It is not clear what the authors mean about individuals taking partition dependence into account in their own decision-making strategies. What exactly do you have in mind?
3. p.10, last paragraph: My reading of Xing et al. (2020) is that their analyses “did not provide evidence that financial aid status moderated the way income partitions influenced resource allocation.” What they *did* find is that students on financial aid allocated more financial aid to families with lower incomes in *both* partition conditions.
4. p.11, par. 1: actually, although Reichelson et al. (2018) failed to directly replicate close version our original Study 4 results among adults, those researchers did manage to create a successful conceptual replication among children.

See: Reichelson, S., Zax, A., Patalano, A. L., & Barth, H. C. (2019). Partition dependence in development: Are children’s decisions shaped by the arbitrary grouping of options?. *Quarterly Journal of Experimental Psychology*, *72*(5), 1029-1036. They later concluded that the transparency of the task (to adults) may have been part of the issue.

Indeed, in a follow-up they found that their simple paradigm replicated among a separate sample of children but not adults:

Williams, K., Zax, A., Reichelson, S., Patalano, A. L., & Barth, H. (2020). Developmental change in partition dependent resource allocation behavior. *Memory & Cognition*, *48*(6), 1007-1014.

1. p. 11, Method: single data collection. The authors should make clear that the study is entirely within-subject rather than between-subjects in parallel.
2. p.14: Extensions: I’m surprised to hear the assertion that there are no off-the-shelf scales for preferences to diversify, but it is an interesting and still not settled question what drives partition dependence in this context. A mindless application of motive to diversify (which is not sensitive to the particular partition presented to participants) could be one mechanism, though there could of course be others. Of course preference to diversify is not really an explanation (several reasons for variety seeking have been proposed in the literature—such as seeking information, overgeneralized concerns about satiation, desire to look adventurous, etc.). As for patriotism, I’m not persuaded that this will moderate partition dependence. I think I understand the instinct that patriots might be more drawn to U.S. charities than international charities (on the other hand, many patriots like to sponsor international causes in the name of advancing US prestige abroad). This said, while I could imagine patriotism increasing motivation to donate to domestic causes, it is difficult to see why this measure would otherwise moderate partition dependence (except for the trivial case where some participants give zero to the International United Way regardless of partition).
3. pp.16-17 Participants: I’m a little concerned that the authors are using a single sample of participants for all 3 studies. Of course this could lead to correlated error across studies, and possible subject fatigue. Presenting participants with multiple partitioned option sets might also inflate attention to differences among partitions. (On the other hand, past research suggests that partition dependence is usually robust to transparency about the partitions used in a study—see e.g. Xing et al. 2020; Fox & Clemen, 2005; Sonnemann, Camerer, Fox & Langer, 2013).
4. p.21, Table 7: It is interesting that the authors chose to use the same income brackets (and same university) as we did in the study that we ran 20 years ago at Duke. Nominal incomes have risen considerably since that time and of course mTurkers generally have no ties to Duke University or, necessarily, a sense of university tuition and financial aid norms. This said, my guess is that these differences won’t stop the authors from replicating our effects.
5. p.21, Study 2 methods. Procedural differences from the original study that are mentioned on p.23 might be anticipated here. Of course the original study used Duke students (who live in Durham county), an incentive-compatible consequence (a random student’s preferences would direct a real donation of $2 per participant); also the funds involved Durham County vs. International United Way Charities—so the 2005 subjects had a closer connection to the decision they were making. This said, I am guessing if the partitioning elicitation is done right (it looks fine based on the supplementary materials) I expect that this shouldn’t stop the authors from replicating our original finding. I do think it would be helpful to readers to provide a visual depiction of how the one-stage versus two-stage elicitation is to be implemented.
6. p.21, Study 5 methods. Here again, it would be helpful to reproduce the elicitation to make clear how the partitioning was implemented and that the six wines included two wines from each of the grape and region categories. Perhaps it won’t make much difference but if I were trying to replicate this study today I probably wouldn’t use the same materials: the wines are now 20 years older (much more expensive if at all available, in some cases probably past their prime, etc.).
7. p.22 Measures: Of course it is an open question how the distribution of wine expertise will differ in this population compared to the 2005 sample of graduate students. For instance, if there is not sufficient variance in this population you would be less likely to find moderation of partition dependence by expertise. Note that in our sample the median number of bottles that distinguished “novices” from “experts” was 4, and of course the current sample is likely to differ. Also note that we only binarized our data for illustrative purposes. In any case, expertise and other indicators of strength of preferences for specific wines should in theory moderate the magnitude of partition dependence effects. The question is whether the measures here will tap into that.
8. p.24, Table 9: taking stock of the classification of the replication, I think that it is basically accurate, but elides reasons why differences could be important, as detailed above.
9. p.25: the proposed data analysis strategy seems fine to me and basically duplicates our approach from 2005.
10. p.26, outliers and exclusions. I’m glad that the researchers are planning to weed out non-human responses, but I do worry a bit about the attentiveness of an mTurk sample. Many researchers, for instance, use attention filters. I don’t feel strongly about this though. It might even be that *less* attentive participants are more likely to exhibit partition dependence.