Dear Dr. Clarke:

I have received two expert reviews of Manuscript MC-ORIG-22-025 entitled "Six of one, half dozen of the other: Suboptimal prioritizing for equal and unequal alternatives" that you submitted to Memory & Cognition. The reviewers are generally positive about the paper and I agree. However, there are a few concerns that need to be addressed before the paper can be accepted.

As you can see, both reviewers provides some comments that I think are relatively easy to address. Reviewer 2 did make a comment about conducting another experiment using a larger sample size. I am not suggesting that you should conduct another one, but I do expect you to seriously look into this comment and argue whether or not you want to follow up with it.

**The previous version did not sufficiently justify the effect size we designed our experiments to detect, and we have edited the motivation to make it clearer why we set the effect size to the value we set it to in our a priori power calculations. Please see the response to this reviewers comment for a more detailed response.**

I also had two minor issues. In the introduction, you nicely introduced several theories related to decision making (prospect theory, regret theory). They were clearly explained. What I however missed was linking these theories with the predictions that were made at the end of the Introduction. In the sections of the separate experiments, you sometimes do make these links, but I can imagine that for readers, it makes sense to immediately link them from the beginning.

**Prospect theory and regret theory, mentioned on the first two pages, situate the decision failure we are investigating here in the larger context of other factors that make utility an imperfect predictor of decision. These two theories are often raised by other researchers in the field of decision-making when we have presented this work, so we felt it important to make it clear we have considered their relevance to the current topic. We have added a note to the end of this opening paragraph that we do not believe these theories address the special case of choosing between equal alternatives that we are investigating in this paper, just to make that completely clear (see page 4). Our experiments are more directly motivated by the other theories we raise (the independent vs competitive classes of the dynamic accumulator models), and we have modified the introduction to stake out this motivation more directly. Thank you for pointing this out; I think our exposition here was weakened by the fact that the results that would have had the biggest theoretical impact in relation to the theories were not the results we ultimately obtained (this makes them no less valid, of course!). But we had a solid theoretical grounding for running the experiments that we neglected to make clear, and have done so in the current version (see the added text on page 10 and page 11).**

Second, I might have missed it (and if so, sorry), but I did not see whether the data (and materials (e.g., scripts) are made publicly available on for example the Open Science Framework. If not, I would suggest doing so, as it might help replication studies and meta-analytic work.

**Yes, all data and data processing scripts are publicly available, although we neglected to clearly associate the repository with the submitted version of the manuscript. We have now uploaded the current version of the paper to psyarxiv [link] and linked the manuscript to the open data on the osf: https://osf.io/y5f9d/**

ACTION: I am rejecting this version of the manuscript, but strongly encouraging a revision if you are able to address the concerns above. Should you revise the manuscript, I will read the revision to make a decision and only send it out to reviewers if I feel unsure whether these concerns have been adequately addressed.

I do not want to have a lengthy process with multiple re-revisions, so I will try to decide on the next version whether the paper should be accepted or not. Thus, please make sure a revision is your best attempt at addressing these issues. If you submit a revision, please include a detailed letter explaining the changes you have made to address the comments here and in the reviews. Please submit the revised manuscript to the website within 180 days. If you cannot meet the deadline, please email journals@psychonomic.org to request an extension.

Once again, thank you for submitting your manuscript to Memory & Cognition, and I look forward to receiving your revision.

Sincerely,

Henry Otgaar

Action Editor, Memory & Cognition

Editor:

Ayanna K. Thomas

Associate Editors:

Sarah Barber, Dana Basnight-Brown, Monica Bucciarelli, Jen Coane, Eliana Colunga, Felipe de Brigard, Bridged Finn, Wendy Francis, Steve Majerus, Henry Otgaar, Laura Thomas

WEBSITE REVISION OF THE MANUSCRIPT:

When you have new files of the revision and a cover letter explaining the changes, please log on to [https://mc.manuscriptcentral.com/mc](https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fmc.manuscriptcentral.com%2Fmc&data=05%7C01%7Ca.hunt%40abdn.ac.uk%7C296fe915dbac4ce4adc708da42de1558%7C8c2b19ad5f9c49d490773ec3cfc52b3f%7C0%7C0%7C637895818853763353%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=LU0fQmudqu%2FEfUSnmF07%2B9H68csFT0PdSsbh1Xybxqc%3D&reserved=0). Enter your Author Center, where you should click on the manuscript listed under "Manuscripts with Decisions." Under "Actions," click "Create a Revision." Your manuscript number will be appended to denote a revision. Follow the steps indicated. Be sure to upload your new files and delete any old versions of the same files.

PLEASE DO NOT SUBMIT YOUR REVISION AS A NEW SUBMISSION. If you have missed your deadline, email journals@psychonomic.org for assistance before going ahead.

\*\*\*IMPORTANT: Please be sure to enter your response to the editor's and reviewer's concerns in the author's response field, not as a separate cover letter or attached file.

Your original files are carried into the revision. Please delete any redundant files (such as the earlier cover letter, manuscript, or figures no longer the same) before completing the submission of your revision.

Reviewer Comments to Author (if any):

Reviewer: 1

Comments to the Author

Thank you for this manuscript, I enjoyed reading it. In this paper, the authors are motivated by humans' seeming inability to prioritize when given two tasks, which leads to performing worse on both compared to if they had prioritized one. The authors conduct three experiments, in which one option was easier than the other, and found that participants still did not prioritize the better option.

I found the paper relatively easy to read, the motivation to be clear, and the experiments to be largely concise and effective for the present purposes. I would characterize my comments as minor.

**It’s great to hear this - thank you for the kind assessment of our work and presentation.**

I'm a little confused by how exactly Hoop Delta is defined (as well as what C and F correspond to numerically).

**Hoop Delta was defined with reference to the point/slab at which participants were approximately 50% accurate (Δ = 0). Other points were then calculated by simply subtracting or adding one or two units of distance (in this case, slabs) from this value. For example, if a participant’s accuracy was 50% for slab 14, then their Δ = 0 would be 14. The points C and F were the points at which participants were 90% and 10% accurate respectively. The figure uses this notation since participants varied in their ability to accurately throw the bean bag into the hoop, but this way the distance for each participant was standardised based on their ability. Clarification has been added to the figure caption.**

To validate the prediction model in Experiment 1, did participants' accuracy from their actual standing positions correspond to what the model predicted?

**To validate the predictions of the model from Experiment 1, an ROC value was calculated for accuracy in session two using the original model (fit to session 1) and one fit to the data from session 2. The area under the curve for model one was 0.81, and the value for model two was 0.83. Although there was a slight difference between these models, both values were high suggesting that the model was able to accurately predict each participant’s accuracy given their distance from the hoop and the size of the target.**

I'm not sure I fully understood the rationale for Experiment 2. From how it was set up, the authors seemed to motivate it by trying to reduce the change participants were inferring the next targets. But, it seems that in the Two-Throw condition, that problem is still there because they don't know which color to throw. Any clarification on this would be helpful.

**Thanks for pointing out the lack of clarity here, we have corrected this. The misunderstanding comes from the fact that there are two details that are randomly selected on each trial:**

**1) which pair of colored hoops to aim for. This is determined by the participant randomly drawing a beanbag from the bag, the color of which tells them which of the two hoops are potential targets on that trial. Based on this knowledge, they choose a place to stand. We explained this in the methods for Experiment 1, and to make it clearer we have added an example to that text (page 16).**

**2) which of those two hoops they should throw the beanbag into on that trial. This is a sequence of “north” and “south” directions which we randomly shuffled for each participant before the experiment. Even though they are told the sequence of north and south designations is random, and the experimenter is clearly reading it off a list, the participant may question whether it is truly random (especially if they choose to stand near what turns out to be the “wrong” hoop for several trials in a row). By making the participants throw to both hoops, instead of designating just one of them as the target, we removed this second aspect of random selection from the experiment.**

**To make it clearer that we were referring to the second (north-south) random sequence and not the first, we have added additional explanation with more concrete wording to the preamble of Experiment 2 (page 22-23).**

For Experiment 3, why were participants given the option of which reward condition to proceed? Wouldn't it have been easier to measure the hypothesized effect if 50% of the trials were equal and the other 50% unequal?

**We mulled various options over when we designed the experiment, this reviewer’s suggestion here being an obvious good option. But we were interested in finding out how participants would choose to divide their reward, given the opportunity. This was a risk because, had our participants been highly consistent in choosing to always split the reward equally or unequally, we would have had to run a fourth experiment, splitting it evenly on half the trials the way this reviewer suggests, in order to find out whether they adjust their choices with the reward structure. Luckily (and interestingly), our participants were not very consistent, so we were able to contrast their choices under both conditions, as well as reveal their underlying preference between those conditions.**

Reviewer: 2

Comments to the Author

The authors investigated how participants use resources in prioritizing dilemmas. They conducted three experiments that presented participants with focus-or-divide problems. The results replicated previous findings in the literature and additional manipulations provided further insight into this question. None of the manipulations showed that participants applied optimal decision strategies for these problems, even though they took into account several relevant aspects of the problem (as shown in Experiment 2 and 3). My overall evaluation of this work is very positive. The paper is well written. Even though the manipulations do not really change the participants’ behavior, the question behind them is clear and justified based on the literature reviewed in the introduction. I also liked that they used a real-world task. Finally, the data analysis approach is state-of-the-art, with additional information provided in the supplementary materials.

Main comment:

The sample size calculation is a bit unconventional and results in a small required number of participants. Moreover, as the authors themselves discuss, there are outspoken interindividual differences in strategy use (in addition to the issue that some participants potentially were not naïve to the objective of the study). These differences in strategy use in combination with the small sample size can make the interpretation somewhat shaky.

**The conventional NHST approach would be to estimate the required sample size to detect an effect of a size that is based on existing literature, but here we only have extremes to guide us; there is just one experiment (Clarke and Hunt, 2016, Experiment 3) in which naive participants make prioritizing decisions that are consistently modified by task difficulty, and in this experiment, they are *perfectly* guided by task difficulty, with all 12 participants choosing to focus when the targets are far away and to divide when they are close. In all other experiments (e.g. Clarke and Hunt, 2016, James et al., 2017; James et al., 2019, Morvan and Maloney, 2012, Hesse et al., 2020) participants’ choices are not modified *at all* in response to task difficulty. It seemed disingenuous to select this as our standard – if we did, the sample size recommended by conventional calculations would be about N=1! Instead we based our choice of sample size around simulations, creating a set of data that matched the distribution of data from Clarke and Hunt (2016) Experiment 2, but with a relatively small effect (a shift in normalised standing position of 0.05, i.e. 5% of the distance between the center and far hoop). We then ran repeated simulations of experiments drawing from these two distributions and calculating the mean difference as we increased the sample size to the point where our estimate of uncertainty around the mean difference plateaued. This plateau was around 15 participants. So we can say with confidence that if we increased our sample size beyond what we have used here, we would learn very little new information. Given that the purpose of the bayesian analysis we present is to estimate the size of the effects of our manipulations and our certainty around these estimates, a sensible standard for deciding when to stop is when certainty stops increasing (i.e. why waste additional participants that add no new information to the data?).**

**That said, I think our description of the power analysis made it sound like it was only designed to detect the expected effect of our manipulation of hoop size, when in fact it was a generic method to detect any shift of 0.05 towards one of the hoops. We have clarified this in the text (page 13).**

I think it would be informative to run a study with a larger sample size (whether this study needs to be part of the current paper, I leave up to the editor). Perhaps this can be done using a modification of the computerized version that was discussed in the paper, which will make it easier to recruit a larger and naïve population. In addition, measuring reaction times in this computerized version can help resolve another question that was brought up but remained unanswered: whether this process is best modeled with independent or competitive decision accumulators (also in line with a suggestion by the authors).

Finally, with this experiment, a modification can also be run where the experimenter sets the relative awards on each trial, which would be interesting.

**Thank you for these suggestions. We have indeed already developed a computerized version of this decision dilemma to try and answer questions about the individual differences we observe in how people deviate from optimal choices. This line of work has opened up a whole new set of questions, because computerizing the task introduces a set of additional interesting factors that are important in decision-making, namely task agency (because the computerized task involves making decisions on behalf of an avatar) and the source and range of outcome noise (because the computerized tasks can have none). Our computerized version of the task has therefore kicked off a long series of experiments exploring a very different set of questions, well outside the scope of the current findings (but not contradicting them in any way). We also agree that it would be interesting to run a study with a more nuanced set of relative rewards to explore the limits of participants’ sensitivity to these (although because of the complications alluded to above, we are not convinced a computer-based task would be appropriate for asking this question).**

Minor comments:

If I understand it correctly, the bags were not replaced until all 9 colors were played. Does this mean that it becomes easier to predict which loop will be selected as the game progresses (until the bags are all of them replaced again). This could be an important confound, given that participants might simply use this information to stand close the hoop in the remaining color (which is not a resource allocation strategy in the sense defined by the authors).

**We think there is a misunderstanding here, and we have gone back to our preamble to the experiment and description of the task and revised the text to clarify our method (page 12, page 16). Yes, the beanbags were not replaced until after they have pulled all nine out and thrown them, and yes, that means it becomes easier to predict which color will be drawn out of the bag as the participant approaches the 9th beanbag. But all the beanbag color tells the participant is which pair of hoops are relevant on this trial, and pulling them out of the bag at random was just a simple way of mixing up the trial order. If it’s the red beanbag, they will be asked to throw it into one of the two red hoops. They pull out the beanbag and based on its color, choose a place to stand between the two hoops of the same color. This color has no bearing on which of the two hoops they are then told to try and get the beanbag in on that trial - this is the “north” and “south” designation, which is determined by a list the experimenter is reading from, which was randomly determined for each participant and only revealed to them after the choose a place to stand.**

In Experiment 3 there is a shorter time interval between the session during which the participant’s skills are assessed and the test session compared to Experiment 1 and 2 (one day versus a week apart). Could this in part explain the differences found in this Experiment compared to 1 and 2?

**Yes, participants were overall slightly better in this experiment even in the condition with equal reward, which could be due to the reward itself, the population we sampled from, or the recency of the practice they got from session 1. We added this to the discussion (page 40-41). To be clear though, the main conclusions we have drawn from this experiment rest on the contrast in standing position between the equal and unequal split conditions and the participants’ preference for the equal splits, which are not undermined by overall better performance in the equal reward condition relative to previous studies.**