Round 2: Response to Comments

Major Comment

- I find the equations for the models inconsistent and confusing. Making these more consistent would greatly improve readability and understanding. Specific examples below:
  - *Eq (3): Var hat and w bar are defined now as:* where $\bar{w}$ and $\widehat{Var(w)}$ are sample mean and sample variance of the weights w of survey sample S, respectively.
  - Model equations *We are sorry for the previous index salad. For all of the model equations, i, j, ℓ and k are now used in the same way (i main factor: aligned/unaligned, j secondary design factor, ℓ participant's decision, ordered from correct to wrong, and k as the kth participant's response, $1 \leq k \leq N$, N number of survey responses) We have explained the notation of $i(k)$ for the indices – essentially, we are trying to combine the advantage of the model matrix form that explicitly connects responses to outcomes (kth row of model matrix X is connects to kth response Y_k) with the interpretability of the parametric form, i.e. $j(k)$ is defined as the value of j for response k. This should address the comments below:*
    - Eq (5): the terms and the subscripts need to be defined; the parentheses in the subscript need to be explained (perhaps nested effects). Also, you say "let Y_{jk} describe the zooming behavior", but only use Y_k in Eq (5). I also don't understand what i, j, and l are referring to in this equation.
    - The indexing changes between Eq (4) and (6) (design is j in (4) and l in (6)).
    - k=1,…,N isn't defined until Eq (6), even though it is used earlier. It isn't clear what N is.

Minor Comments

- The title is much improved, great change. I'd recommend removing the question from the title, as it seems repetitive with the first sentence of the paper.
  - *We have removed the question from the title to address this.*

- Last paragraph of page 2 and first paragraph of page 3 (after numbered list) both start with "In this work"; one of these should be changed (could change the second to "To address these questions").

  – *We have changed the second use of "In this work" to "To address these questions", as suggested.*

- Pg 3: "…we either split a survey sample in two and show each subsample a distinct structural version of the chart or test variations of a chart across distinct rounds of the survey." Please add how much time passes between rounds.

  – *A clarification that one month passes between rounds has been added here.*

- Pg 4, 2nd paragraph: I think "aligned pieces" and "unaligned pieces" in the parentheses should be switched.

  – *This has been updated.*

- Figure 1 caption: "Please refer to the appendix for full size images of the stimuli." -> "Please refer to the appendix for a full size image of stimuli (a)."

  – *We have updated this caption accordingly.*

- Pg 5: I really like the last sentence of the Task 1 section: "Any differences in observed responses can therefore be attributed to this difference in presentation." A sentence like this after the other 2 tasks would be a helpful addition.

  – *Great idea! We have added some language to this effect for both Tasks 2 and 3.*

- Pg 7, 1st sentence: "While there is some risk…" A sentence like this should appear earlier, like on page 3 in the study design section.

  – *We have added a sentence in the paragraph starting with "We ask participants to determine which of two…" to clarify that the unaligned task comes first to reduce the likelihood of participants recognizing they are the same task and carrying over responses.*

- Pg 8, table 1 caption: Remove "as discussed above"

  – *This has been updated.*

- Pg 9, 1st paragraph: "the dark points" → "the black points and error bars"; "hashed lines" → "vertical lines"

  – *These wording updates have been made.*

- Pg 9, fig 3: increase the axis text size

  - *This has been increased as much as possible without the labels running into each other.*

- Pg 10, fig 4: I think the figure would be easier/more natural to read if you switch the ordering of the x axis (unaligned first). This better represents the order in which the questions are shown and would highlight the improvement of moving from unaligned to aligned. Also, this figure should be bigger. Finally, the sentence "There's a huge asymmetry in the number of responses where participants answered only one of the questions correctly." is very confusing to me, could you clarify this?

  - *We have rephrased the sentence about the asymmetry in responses for clarity and reordered the x axis to place the unaligned responses first.*

- Pg 11: Thanks for adding all the detail about the letter comparisons in the figures and tables, very helpful. I'd start a new paragraph for this description, starting with "The letters allow us to assess…". I'd also specify: "The letters in the figures and tables allow us to assess…". Finally, the one sentence paragraph on this page should be combined with another paragraph.

  - *We have made these edits.*

- Pg 11: "two estimates within the same column are significantly different (at the 5% level) if they do not have any letter in common": this works for the table, but is confusing for the figures (and the reader is mostly looking at fig 5 when reading this description), be more specific here

  - *We have expanded this description to explain the table and figure interpretation separately.*

- Pg 12, Table 3 caption: "at 5%" → "at a 5% level of significance"

  - *This has been updated.*

- Pg 14, fig 7: Could you rephrase this, I find it hard to understand: "Incorrect responses on the unaligned task show a significant boost in certainty in the vertical stacked bar." Also, in the next sentence "the difference…matters significantly" does not make sense.

  - *We have rephrased the first sentence to improve clarity, and removed the second.*

- Pg 15, 1st paragraph: "Our pattern of higher accuracy in selecting the correct response on the aligned task relative to the unaligned task holds across all demographic groups and structural variations." To make this comparison, shouldn't the aligned blue letters and the unaligned blue letters be different?

  - *In Figure 8, the letter assignment is completed separately for the two subfigures: aligned chart in Figure 8(a) and the unaligned chart in Figure 8(b). However, we can see lower accuracy visually if we do compare across Figures 8(a) and 8(b).*

- Pg 16, figure 8 caption: Some of the conclusions made seem to conflict with how I interpret the plot. Clarification is needed here. Specifically:
  - "For aligned tiles, gender, age, and education are not significant factors." It looks like gender is significant? It has a different blue letter.
    - *This has been updated to reflect that gender and income are associated with significantly different response patterns in the aligned task.*
  - "When income levels increase, the percentage of wrong answers decreases, while the percentage of correct answers increases slightly (significant at 5%)." Looking at the pattern of the orange letters in the "by income, aligned" plot (a,ab,b,ab), it seems like the only significant difference is between the <$30k group and the $60k-$100k group, so it seems incorrect to say "the percentage of wrong answers decreases … (significant at 5%)". Similarly, the blue letters in that plot are all a, so I don't think it is correct to say "percentage of correct answers increases slightly (significant at 5%)".
    - *We have removed this sentence.*
  - "higher levels of education and higher levels of income are associated with a significant increase of panelists choosing a response of 'they are the same', resulting in significant decreases of both correct answers and wrong answers." Again, I don't see evidence in the plot to say these differences are significant. Could you please clarify this? For example, in the "by income, unaligned" plot, the blue letters are all a, so there is no significant decrease of correct responses there.
    - *We have edited this caption to clarify that there is a decrease in wrong answers for increasing education and income, but only a decrease in correct answers for increasing levels of education.*

Overall

- **The paper suits better the Data Science in Action Section better than the Statistical Data Science Section, since this is a case study.**
- **Number the sections as the JDS style does.**