Responses to Reviewer #1

This manuscript addresses a very interesting and useful problem, the bias we, as geologists, seismic interpreters or geophysicists, have when interpreting seismic data. This study has even more impact as no context at all was given to the interpreters and they were limited in time for their interpretation. The results highlight quite a variability in the finale interpretation with a majority of the students falling into the most common (easiest?) interpretation (extensional settings, with faults dipping towards the right). Does it reflect the background and familiarity of the students towards extensional over compressional tectonic settings? The title and the abstract are promising, however the content of the paper lacks of clarity, clear objectives and outcomes, and conclusions that support the initial statements. The two terms that require a clear definition are conceptual model and uncertainty (I think you are mixing uncertainty vs variability in the paper).

I recommend a better organisation of the manuscript and its content, and a title (the second part of the title 'evidence of conceptual model uncertainty and anchoring' is misleading in view of the results/conclusions) that reflects better the content before publication in Solid Earth. I have ticked major revisions but this could be considered as minor as it is mainly regarding the writing and presentation of the results rather than the scientific approach (I do believe that this type of study is very important and exciting!).

Please find attached the manuscript with modifications added to reflect reviewers' comments and concerns. Unless specifically mentioned below, we have introduced all changes suggested in the supplementary by the reviewer.

1. I think the paper requires a clear and exhaustive definition of a conceptual model and what goes into it. Then you can refine the possible and plausible conceptual models that are appropriate to this case study. By reading the manuscript I often got the impression that the conceptual model was the result of the final interpreted section but at the same time that the conceptual model influenced the interpretation. It is not clear what is the relationship between the conceptual model and the interpretation: - Does the conceptual model influence the interpretation (c. model => interpretation)? - Does the interpretation become the conceptual model (interpretation => c. model)? - Are the conceptual model and interpretation developed at the same time/simultaneously (c. model <=> interpretation)? This becomes quite important when you talk about uncertainty and anchoring of the conceptual model.

We have addressed this comment by adding a clear definition of conceptual model in the introduction as suggested by Reviewer #1, and located it upfront in the second paragraph of the introduction:

Geoscientists employ mental models (or "conceptual models") that incorporate their observations and that conform to their understanding of the world (Shipley and Tikoff, 2016). These conceptual models are dynamically modified or renewed with the arrival of new observations (input information), and are used to produce predictions (inferences) that can help to answer questions about the world (Shipley and Tikoff, 2017). When confronted with geological data, interpreters need to apply different conceptual models, acquired during their training and past experience, together with robust interpretation methodologies, in order to produce interpretations that honour the data, particularly in areas of great uncertainty (Bond et al., 2007; Bond et al., 2015). Interpreters need to be able to identify key elements (e.g. growth geometries, regional level) and employ different validation techniques (e.g. restoration, attribute analyses) that allow differentiation between (a priori similar) conceptual models (Bond, 2015). Conceptual models are therefore the basis of the interpretation, as they provide the necessary criteria to make sense of the data (Frodeman, 1995).

We have then better linked this to the following section on anchoring bias. New text added in italics:

To deal with uncertainty, interpreters employ heuristics (or 'rules of thumb') in the process of generating the conceptual models, and that makes them subject to a broad range of cognitive biases (Kahneman et al., 1982). One of these biases is related to the capability of interpreters to adjust their interpretations from their initial ideas or conceptual models. This type of bias, called anchoring, has been identified in many decision-making processes since it was first described by Tversky and Kahneman (1974), and takes place in the seismic interpretation process. Rankey and Mitchell (2003) investigated the effect of anchoring in an interpretation experiment by asking interpreters to reassess their seismic interpretations after being provided with additional well data. Their work shows that most interpreters did not feel that their interpretations needed to change substantially, in spite of data showing changes in porosity and net-to-gross predictions that did not fit with their initial interpretations. Their results suggest that interpreters were anchored to their initial conceptual models, and that they were reluctant to change their mind in light of new data. In a different experiment, Bond et al. (2007) observed that participants asked for the geographical location of the section and suggested that interpreters could use this information to build their conceptual models, by using geographically specific knowledge of e.g. the relevant tectonic setting to anchor their interpretation. For example, an interpreter knowing a seismic section was from the North Sea may assume a conceptual model based on an extensional tectonic regime and will consciously and unconsciously look for normal faults in the seismic data. However, if the conceptual model is wrong, e.g. there is significant inversion in the seismic section, the interpretation could be compromised. *So although conceptual models can be dynamically modified or renewed with the arrival of new observations, as described by Shipley and Tikoff (2017) and others, anchoring bias often results in limited adjustment from initial models.* Thus, although conceptual models are needed to develop geologically sound interpretations, they can also create anchors to potentially erroneous outcomes.

2. The title states "evidence of conceptual model uncertainty", however uncertainties are not clearly defined, addressed or dealt with within the manuscript. If there is no clear definition of uncertainty (qualitative and/or quantitative) and paragraph/section in the Discussion, this should be removed from the title. When at the end of the conclusions, you say "uncertain geological and geophysical data", it seems that uncertain here only means that there is no unique interpretation. This case your study only illustrates that that there is a range of different yet plausible interpretations for one given seismic line (which was provided with no colour scale, which is supposed to be as standard information than a scale bar). It doesn't really highlight any clear uncertainty (why these different interpretations, are they related to the background, to the colour scale, or else? As you state that vertical exaggeration does not have a real impact).

We have changed the title of the manuscript to better reflect the findings presented throughout the text, and revised the presentation of the objectives of the study in the introduction and their interrelation in the discussion and conclusions sections. The new title is:

Evidence of anchoring to initial conceptual models during interpretation of a vertically exaggerated seismic section

Regarding uncertainty vs variability, we use the variability or range in interpretations as indicative of the range of interpretational uncertainty (see also Schaaf et al. 2019 (this volume)). In doing so we have considered what influences the variability in the interpretations proposing that anchoring to initial conceptual models appears to influence the range in interpretations. We also considered whether vertical exaggeration can introduce a greater range in interpretations and hence interpretational uncertainty. However, we found that vertical exaggeration had a subdued influence in the interpretation compared with anchoring to conceptual models. We have made this clearer in the updated introduction and the discussion/conclusions and recommendations sections.

3. I am a bit dubious on the fact that this study proves/shows that the conceptual model is anchored. Given the short time students were left with for the interpretation (15-30 min), it seems that they would have not been able to supply a different conceptual model to the one they started with. Indeed, if the students were provided with additional data or context after a first interpretation, would they update their model or will they keep it unchanged? Is it possible to define 'anchored' conceptual models without taking that into account? If the authors are satisfied with this simplified definition of anchoring, they should discuss it or at least make it clearer in the manuscript.

The reviewer is correct in that we surmise the anchoring from the interpretation process we asked the participants to undertake and the outcome of that interpretation process rather than through provision of additional data. In the original Tversky and Khaneman experiment anchoring was demonstrated by providing an initial value from which the participants were then asked to give an estimate (they were not provided with additional information). In contrast interpreters in the experiment by Rankey and Mitchell (2003) were given additional information and showed that interpreters were reluctant to adapt their interpretations to new information. We suggest that the final interpretation outcome in our experiment results from participants initial fault feature selection (i.e. right or left dipping elements in the seismic image data). In this way their initial conceptual model and its application provides the anchor, in much the same way as the initial values given by Tversky and Khaneman in their experiment provide the anchor to future value estimates. We described this in the third paragraph of the discussion:

In summary, from the analysis of the fault and horizon interpretations of participants, three conceptual models are identified (Figure 3) that have been applied in interpretations of the data. What we do not know is how the individual participants honed onto their 'chosen' conceptual model. The participants were prompted to interpret the faults as their main task in the experiment instructions, and as a secondary element to interpret a horizon to show fault motion; *an interpretation sequence as shown in figure 9*. We should state that we cannot be sure that all participants followed this workflow, but we have no evidence to suggest that they did not.

Irrespective of the exact interpretation sequence, we suggest that once participants started interpreting certain 'features' in the reflection seismic image data as faults or horizons, they became anchored to an initial conceptual model and fitted the rest of their interpretation to this model. *Consequently, we suggest that interpreters were likely anchored to their initial thoughts on the direction of dip of the faults and the rest of their interpretation is determined by this initial fault model, irrespective of whether later interpretative elements conform to the data e.g. horizons cutting reflections, as seen in Figure 3, this has previously been reported by Rankey and Mitchell (2003) and Torvela and Bond (2011). Although, there appears to be a threshold of tolerance for data dis-confirmation. Note that no left-ward dipping faults with a reverse sense of motion have been interpreted, in which horizons would very distinctively have cut seismic reflectors (see figure 9d).*

Experience and knowledge is expected to have played a key role in informing the initial observations that led to selection of a conceptual model at the beginning of the interpretation. We purposely chose a student only cohort to mitigate against the competing effects of experience and knowledge with other factors we wanted to test. To ensure this was the case we have analysed the data for differences in interpretation outcome between students from different Universities and between undergraduate and postgraduate students. This analysis shows no strong evidence that experience had an effect on interpretation outcome. We therefore interpret our data as showing that although initial interpretations are informed by the data, these first conceptual models become anchored to and are applied irrespective of whether they later conform to all the data, albeit to a threshold. This suggests that initial conceptual models play a dominant role in interpretation outcome.

4. I understand that this study is quite exciting but it would even be more if it were directly related to the background of the interpreter. In the appendix you provide the survey and questions handed to the students. A summary of the results of this survey should also be added to the paper/appendix to know if Normal vs Reverse fault is falling mostly for people that interpret often or with no knowledge about seismic interpretation. I think this survey is also part of the bias that forms the conceptual model.

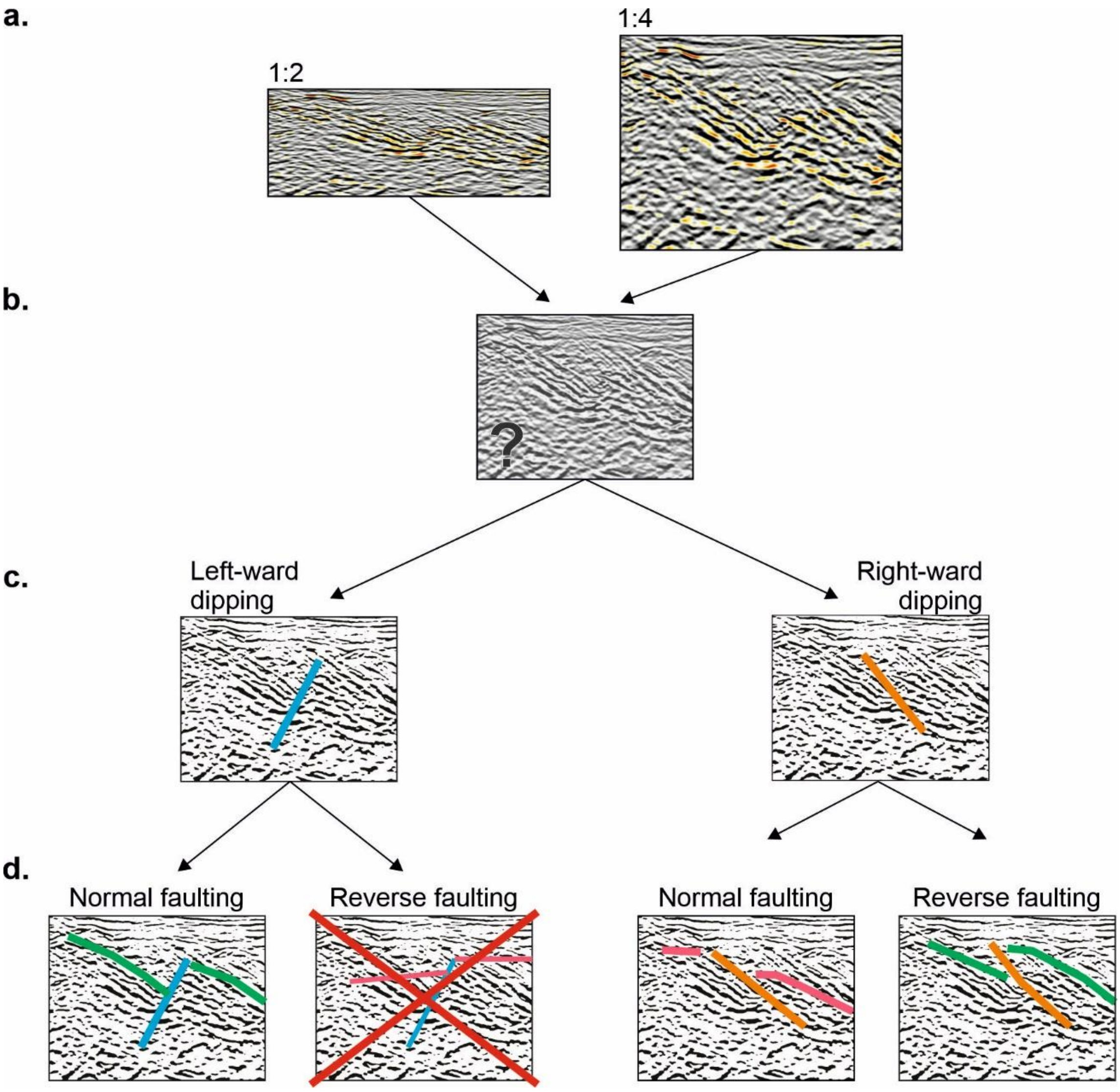
This comment was also raised by Reviewer #2 and a detailed response can be found in the responses to Reviewer #2. In summary, we did a broad assessment of the effect of experience in the interpretation results and did not find any conclusive correlation. There is a disparity in the number of undergraduate vs postgraduate participants (122 vs 34 participants, respectively), and we do not feel their experience levels are potentially so dissimilar to treat the dataset as two different cohorts. We nevertheless found a small difference in the fault types interpreted by the two cohorts, and therefore added a paragraph to the text to state that further research in this matter is needed:

There are minor differences between the fault type interpreted by undergraduate vs postgraduate students, but the disparity in the size of the two cohorts (122 vs 34 interpretations, respectively) does not allow us to pursue this line of research. The effect of level of education and experience in seismic interpretation has been raised in the past (e.g. Bond et al., 2012; Alcalde et al., 2017b), so further work in this area could provide fruitful in better understanding interpretation processes.

5. One additional figure summarising clearly the work, such as the different interpretations, conceptual models and implications (such as which interpretation is the most probable) is necessary to fully comprehend the implications of this work.

We have added a new figure (Figure 9) to summarise the proposed interpretation workflow. In this figure we propose that, independent of the vertical exaggeration of the seismic section interpreted (i.e. 1:2 or 1:4 vertical exaggeration), participants interpreted the faults first, as requested, and the rest of the interpretation was anchored by this initial fault selection.

Figure 9: Proposed interpretation sequence. (a) The seismic images were presented in both 1:2 and 1:4 vertical exaggerations. (b) Independently of the image interpreted, the participants of the experiment faced the problem of how to interpret the right-ward dipping structures and the chaotic seismic fabric. (c) Participants interpreted the central fabric either as a left-ward (blue) or right-ward (orange) dipping fault, which consequently triggered (d) the horizon interpretation determining the motion (normal, green horizons; and reverse, pink horizons) of the fault. The left-ward dipping, reverse faulting interpretation (crossed out in red) is too difficult to fit to the seismic data, and so only one participant chose this interpretation.


a.
1:2
1:4
b.
?
c.
Left-ward dipping
Right-ward dipping
d.
Normal faulting
Reverse faulting
Normal faulting
Reverse faulting