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.

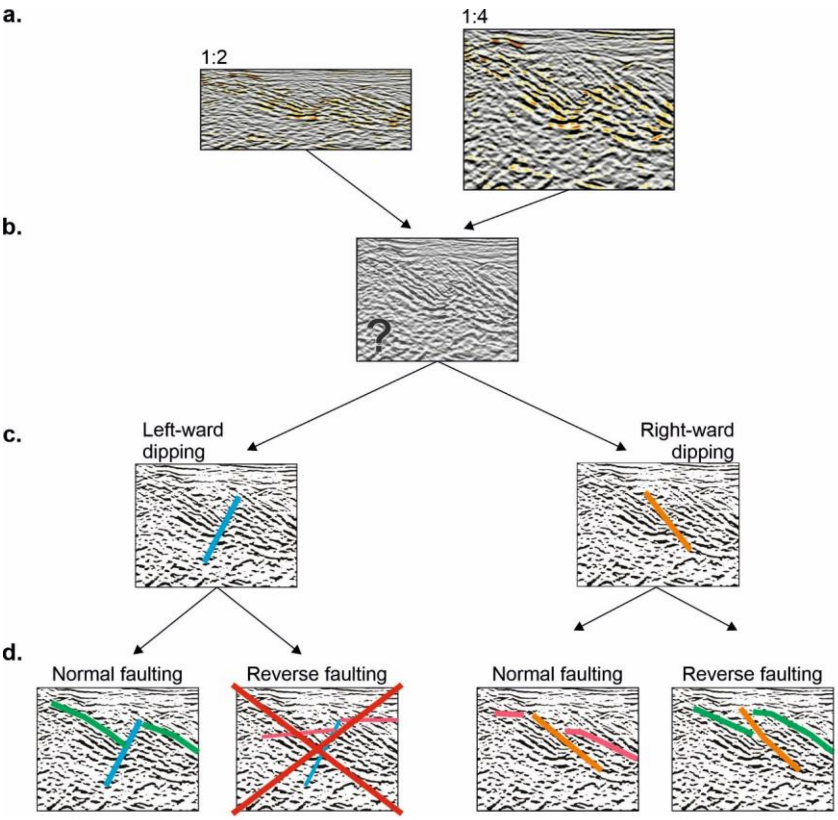
This comment was also raised by Reviewer #2 and a detailed response can be found in the responses to Reviewer #2. In summary, we did a broad assessment of the effect of experience in the interpretation results and did not find any conclusive correlation. There is a disparity in the number of undergraduate 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.



Responses to Reviewer #2

General comments

1. Effect of background and expertise. In the introduction, it is mentioned that previous studies have shown that the background of the interpreters is essential. However, the authors do not seem to take into account the workflows and methodologies used during the interpretation, which may be related to the knowledge and experience of the interpreter. I suggest that the result of the study should be filtered based on knowledge and experience. You can divide the result, e.g. as undergraduate vs postgraduate, and use additional information such as knowledge and attendance to courses in structural geology or seismic interpretation to rank the knowledge of the interpreter. That information is already in the data collected.

As suggested by Reviewer #2, we have conducted a broad assessment of the effect of experience of the participants in the interpretation results. A summary of this assessment is summarised in the table below:

Table 1: Statistics for the interpreted fault directions (left 'L' or right 'R'), and motions (normal 'N' or reverse 'R'), of the total interpretations (left, in blue) and separated by education of the participants (undergraduate, centre, in orange, or postgraduate students, right, in green).

Total interpretations			
Direction		Type	
L	R	N	R
52	67	67	32
43.7%	56.3%	67.7%	32.3%

Undergraduate			
Direction		Type	
L	R	N	R
39	50	52	23
43.8%	56.2%	69.3%	30.7%

Postgraduate			
Direction		Type	
L	R	N	R
13	14	13	9
48.1%	51.9%	59.1%	40.9%

An issue encountered in considering the data in 'experience' cohorts is that there are a different number of participants in each cohort, with 122 undergraduate students and only 34 postgraduate students. Nonetheless, separating the results into experience cohorts shows little difference in the overall ratios of the types of fault interpreted. The general trend observed in the total interpretations (i.e. greater percentage of right-dipping faults and normal fault types) are conserved.

In this experiment, the original hypothesis was that vertical exaggeration could have a strong effect in interpretation. However, our results turned out to show that conceptual models might play a stronger role than perceptual biases such as changes in vertical exaggeration. Experience, as highlighted by Reviewer #2 and by studies referenced in the manuscript introduction (e.g. Bond et al., 2012, *Geology* or Alcalde et al., 2017b, *Journal of Structural Geology*), can impact interpretation results, but we do not see this here.

Nevertheless, we have added a paragraph to the results section to describe this issue and to encourage the study of this effect in future works. If the editor would like us to add the above tables into a data repository, to support our findings, we are willing to do so:

To check if other factors, specifically: educational background and experience, were influencing interpretation outcome we also analysed the data for disparities between different University cohorts and between undergraduate and postgraduate students. There are no major differences in the analysed results across student cohorts from different Universities, *or between undergraduate and postgraduate students. For the latter cohort the difference in numbers (undergraduate (126) vs postgraduate (35) students) is large and does not allow easy comparison; despite this the ratios of leftward and rightward* *dipping faults and the sense of off-set is consistent across the cohorts. 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) and we suggest that this is still an area of interest for future work.*

Our analysis of experience within the cohort has been taken further in this revision following a comment made by Reviewer #2 in the comments in the manuscript Section 4.1: Can these result from the lack of experience of the interpreters? How does this impact your results?

We analysed the full conceptual models by experience level (undergraduate vs postgraduate), and did not find any remarkable difference between the models interpreted and the level of studies, given the number of participants per cohort:

	Total interpretations (with sense of fault motion)*		Right-ward dipping, normal faults		Right-ward dipping, thrust faults		Left-ward dipping, normal faults	
Undergraduate	68	76%	19	28%	22	32%	27	40%
Postgraduate	21	24%	4	19%	9	43%	8	38%

*Note that the number of interpretations do not add up to 161 when summing the subcohort results: this is because not all the interpreters featured a clear interpretation model (i.e. the motion of the faults was not clear).

2. Anchoring and bias effect. The authors interpreted that defining a reflection or set of reflections as horizons or faults may represent a form of anchoring. However, this seems different from the concept and examples described in the introduction where new data is given after an initial interpretation, and the interpreter does not see the necessity to adjust their interpretation. In the discussion of anchoring, the authors mentioned “horizons cutting reflections” as an element that suggest anchoring. However, it is not possible to know how and where in the seismic section the interpreters defined the horizons and the faults, or where did they start the interpretation. Moreover, the authors also discussed that they could not know if the students changed their minds during the interpretation, or what elements within the seismic section were considered in the process. Therefore, it is not clear how the anchoring bias was defined. The fact that experience was not taken into account is also a problem. “horizons cutting reflections” may reflect a lack of understanding of seismic interpretation.

See comments on anchoring made to reviewer 1, copied here for clarity:

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:

On the interpretation process – we asked participants to interpret the faults first and then a horizon to show fault off-set. Although we cannot be sure they followed this process, the final interpretation outcomes (multiple faults interpreted and a single horizon) leads us to believe that the participants followed the workflow as instructed. We make this clear in the manuscript, section 2. Experimental set up, lines 23-25:

' The participants were asked to "interpret the main faults crossing the section as deep as possible", as well as to add a "sedimentary horizon to mark the displacement",..'

In the discussion on anchoring we discuss the potential for participants not to have followed the workflow requested. We felt that it was important to raise this as a potential weakness in our methodology, but as stated do not feel it impacts our findings. We have updated this sentence to reflect that (page 8, lines 8-9):

'We should state that we cannot be sure that all participant followed this workflow, but we have no evidence to suggest that they did not.'

On how we define conceptual models and anchoring – We have added in a new section to the introduction that discusses previous work and how this relates to our study (pasted below – new text in *italics*):

Geoscientists employ mental models (or "conceptual models") that integrate their observations and that conform their understanding of the world (Shipley and Tikoff, 2016). When confronted with geological data, interpreters need to apply different conceptual models, acquired during their training and past experience (through learning), 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 the key elements (e.g. growth geometries, regional level) and employ different validation techniques (e.g. balancing or restoration) that allow differentiating between (a priori similar) conceptual models (Bond, 2015). The conceptual models therefore incorporate all the elements that shape the knowledge of the geologist of a certain aspect of the geology; for example, the conceptual model of a turbidite system will include characteristics about their origin and evolution, common stratigraphic sequences, lithological composition, stratigraphic structures associated, etc. 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). Conceptual models are therefore the basis of the interpretation, as they provide the necessary criteria to make sense of the data (Frodeman, 1995).

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*.

In the discussion we now introduce a new figure (9) that shows the interpretation process. We agree with the reviewer that “horizons cutting reflections” may reflect a lack of understanding of seismic interpretation, but we do not believe this to be the case. All the students had experience in seismic interpretation and there were no models of left-ward dipping faults with a reverse sense of motion have been interpreted, in which horizons would very distinctively have cut seismic reflectors. We have added in a new sentence in the discussion on this and refer to a new figure (9d).

On experience more generally see response to comment 1. We have also added the following paragraph into the discussion:

‘Experience and knowledge are 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.’

3. Vertical exaggeration analysis. Although the exercise is related to the perception of the interpreter to the scale, the authors based their analysis on the interpretation of seismic sections in time. It seems that the authors consider 1:2 and 1:4 represent vertically exaggerated displays of a 1:1 section. These scales are more likely to be display factors for seismic sections in time. A 1:1 display factor in time will be significantly different from a 1:1 section displayed in depth. The same problem will apply for other display factors in time, which will not be representative of the real scale of the structures in depth. Moreover, the depth section will depend on the velocity model used for depthconversion. For example, assuming an unrealistic and simple

model of a constant velocity of 5000 m/s throughout the section, the 1:2 display ratio you used will be equivalent to a 1:~0.7 depth ratio, whereas at a velocity of 4000m/s, this will be equal to a 1:~0.9 depth ratio. This means the section with 1:2 display is more likely horizontally stretched and not vertically exaggerated. The analysis of the impact of vertical exaggeration should, therefore, be performed in a depth section and possible in PSDM section. I suggest you depth-convert a section, so you have a reference for what a 1:1 section in depth looks like. Then compare these to the time sections given to the students and analyse the result taking into account the depth section. You can also consider additional display factors (e.g. 1:6, 1:8) in time to complement your study.

The seismic section used was extracted from the Virtual Seismic Atlas (www.vsa.org) where it is stored with an approximately 1:1 display, as explained in the description of the section. Even if it was not the case, and the display was 1:1.3 (or 1:0.7) our results would be consistent, because we vertically exaggerated the original image 2 times and 4 times to use in the experiment, and then converted the interpretation back to the 1:1 (or 1:1.3-1:0.7) for analysis. We have added a sentence to make clear that the original image had no vertical exaggeration, according to the source of the data:

The section used in this experiment was originally downloaded with no vertical exaggeration (i.e. with an approximate horizontal to vertical ratio of 1:1), according to the Virtual Seismic Atlas information. In a series of interpretation experiments, this seismic image was presented to participants with horizontal to vertical exaggeration of 1:4 (Figure 2a) or 1:2 (Figure 2b), hereafter called *1:4* and *1:2* sections.

We agree that the velocity model has a strong impact in the interpretation outcomes. This issue was partly studied in an experiment by Alcalde et al. (2017), whose research shown that the depth conversion had a major impact on the interpretation, even if this was mostly related to the changes in the image quality derived from the depth conversion process. We have no data to inform velocity models for our seismic section to run a robust depth conversion, and therefore this option was dismissed during the experiment design phase.

Alcalde, J., Bond, C.E., Johnson, G., Ellis, J.F. and Butler, R.W.: Impact of seismic image quality on fault interpretation uncertainty. GSA Today, 2017.

Finally, on adding more display factors would be helpful to constrain the conceptual model vs vertical exaggeration issue. However, we did not include more displays due to the subject availability (160 students would have meant only ~40 participants per display, reducing the statistical meaning of the results), and to the display options (i.e. a 1:6 display would not fit an A4 sheet, or would require reducing the 1:2 section too much). Nevertheless, we have added the suggestion of Reviewer #2 to the result section, to acknowledge that using more displays would help to constrain the results:

Similarly for the right-ward dipping fault interpretations normal fault dip angles are low 24º-27º, but not as low as those interpreted to the right, suggesting that the angle of dip of the fault is driven more by the seismic image data than by any effects of vertical exaggeration. *Testing with more display options (e.g. 1:6 or 1:8 vertical exaggeration) could be helpful to confirm this finding, and would be interesting lines for further enquiry.*

4. Fault dip data. The author should clearly state in the text that the measurements were made for comparison and are not representative of real fault dips.

We have amended this issue in Section 3 to make this clear (interpretation results):

The interpretation results were digitised manually and then converted to a 1:1 vertical exaggeration (VE=1:1) for comparison; therefore, the fault dip angles presented in this work are VE=1:1 in time. *As the sections were interpreted in TWT, the analysed dips of the faults are not real dips (i.e. these observed in sections in depth), but their relative differences are still comparable.* Individual examples of the interpretation results after digitisation from both the *1:2* and *1:4* sections are shown in Figure 3.

We also added a table (new table 1 in the manuscript) with the median dip angles depth converted using a velocity of 3000 m/s (according to Stewart, 2011). This way we are able to compare the dip results with the Andersonian models in the discussion, and provide more realistic fault values than the ones calculated in TWT.

Table 1: median values in two-way traveltime and their depth-converted equivalent of the 1:2 and 1:4 sections, divided by dip direction and fault motion.

Section	*1:2*			*1:4*		
Dip direction	*Right*	*Right*	*Left*	*Right*	*Right*	*Left*
Fault motion	*Normal*	*Reverse*	*Normal*	*Normal*	*Reverse*	*Normal*
Median (TWT)	*13º*	*23º*	*16º*	*10º*	*21º*	*22º*
Median (depth-converted)	*19º*	*33º*	*23º*	*14º*	*31º*	*30º*

The analyses of fault dips should be divided based on the conceptual models, as different assumptions were made for these. Mixing data from different conceptual models based on dip direction as in figures 4 & 6 seems inadequate. The authors mentioned that right-dipping reverse faults required higher angles than right-dipping normal faults so that both populations will differ due to the assumption made during the interpretation. Hence, these should be treated separately as in figure 5 & 7. The data shows significant variability suggesting that the average should not be used. In figure 8, the analysis of fault dips should take into account the position of the faults within the section. As they are, the results show too much dispersion and mix different conceptual models (right-dipping faults). It is not possible to correlate faults between different display scales. Hence, it is not possible to know if the perception changed due to the scale. As currently displayed (rose diagrams (Figure 7) and curves (Figure 8)), the results are difficult to interpret and should not be used. I suggest that the authors subdivide the dataset based on the conceptual models and use plots of horizontal distance vs fault dip. For example, you can compare the horizontal distribution of the interpreted faults to see which faults are recurrent in the interpretations, and are located in similar places. Then for each location where faults are repeated, you can analyse the “dip” distribution and calculate a representative value for that population. This can then be compared between scales and plotted in figure 8.

We generally agree with Reviewer #2 that the average values are not fully representative of the fault dips, given the skewness featured by some of the distributions. To solve this, we have substituted the average values with the median values, which, together with the standard deviation (already shown in figures 6 and 7) provides a clearer picture of the characteristics of the dip distributions. We have also modified Figure 8 to consider the medians instead of the average dip values, and have included the medians of the dip direction (left or right) and fault motion (normal or reverse) sub-cohorts for comparison. We have also amended the discussion in the text with the new median data. The median results do not produce significant changes in the overall results, but the suggestion of adding sub-cohorts to the analyses (particularly in Figure 8) has helped to disentangle the effects and interrelationship of vertical exaggeration and conceptual modelling in interpretation outcomes. With these additions to the text and the figures, we hope that the results are clearer to the reader.

We disagree with the suggestion to remove figures 4 and 6. Figures 4 and 6 show the preliminary split of the results by cohorts, to which the sub-cohorts presented in figures 5 and 7 belong (i.e. figures 5 and 7 are sub-cohorts of these in figure 4 and 6). We believe that the current structure of the methodology is the most appropriate to describe the rationale followed, both in terms of the steps followed and in the order they were applied: we first analysed the interpretations as a whole and identified the three potential cohorts (i.e. 1:2 vs 1:4, left dipping vs right dipping and reverse vs normal). After analysing these cohorts we split even more these cohorts into the cohorts presented in Figs 5 and 7, based on the hypothesis that fault dipping direction had more influence in the outcomes than vertical exaggeration, but this hypothesis can only be formulated once the overall results have been discussed. We see no reason to remove figures 4 and 6 from the manuscript.

Regarding the subdivision by faults, we chose not to further subdivide the results fault by fault, as this would only add even more complexity to the results and the population (i.e. the number of faults interpreted per faulted section) will be reduced dramatically, potentially below statistical significance. Reviewer #2 has already shown their concerns about the complexity of the results, but we do not believe that further splitting of the data into even more sub-cohorts (as many as the faults present in the seismic image, 5-10 times more than the current) will help in this matter. Furthermore, these results would be relevant if the two vertically exaggerated sections were presented to the same subject, and this was not the objective of the experiment. The seismic section was selected because it presented faulting in domino blocks, with little variability in fault dip. The presence of the variability in fault dip is captured within the SD and the rest of statistical analyses, and these are averaged out across the entire section. The differences in variability are already discussed in the text, in section 4.2 Fault dip variability, which has already modified as suggested by Reviewer #2.

- Final remarks Vertical exaggeration and anchoring are important aspects that should be taken into account during seismic interpretation. Research on the impact of these in the outcome of the interpretations can contribute to the seismic interpretation workflow. However, the way the experiments are designed and the way results are present are crucial. Although the authors try to discuss the importance of vertical exaggeration and anchoring, the paper in its current state does not give support for their conclusions. The effect of anchoring is based on assumptions, and the existence of bias cannot be evaluated from the data and the analyses presented. The experiment, as described, is unlikely to support a discussion on anchoring and bias. The use of time sections makes the results uncertain, and their analyses do not include the background and experience of the interpreters. I suggest the methodology used to analyse the data needs to be modified. Current diagrams are difficult to analyse and sometimes combine the result of different conceptual models. Some of these do not support their discussion points. I suggest the author should revisit the methodology and parameters used to analyse the data, as well as the way the results are displayed. This inevitably requires major changes in the way the paper is presented, including significant modifications to the text and figures. The discussion and conclusions should be reconsidered after these modifications.

We are very thankful for the thorough review and comprehensive comments. We acknowledge that the methodology presented is not the perfect experiment, but in spite of the complexity of the study and the multiple inter-relationships, we have attempted to quantify the impacts. From this we believe that the data and analysis support discussion of the issues of interpreting vertically exaggerated seismic images and in conceptual model anchoring. We believe we have been open in qualifying the limitations to our study, and hope that it presents a starting point for future study in these areas. We have made significant changes to the manuscript according to the comments made by Reviewers #1 and #2 resulting in a clearer manuscript.

Specific comments in the supplementary file

Unless specifically mentioned below, we have introduced all changes suggested in the supplementary by the reviewer.

Section 1 – Introduction

Seismic images are indirect representations of ~~complex changes in~~ the physical properties of rocks in the subsurface.

Seismic images are indeed representations of the changes in physical properties of the rocks; if there were no changes, the seismic image would be blank. So we decided not to remove this from the text.

Consider putting together all the examples of the impact of vertical exaggeration on interpretation.

We do not understand this comment. The current structure presents positive examples of vertical exaggeration (paragraph 1) and after that examples where the vertical exaggeration disturb the perception of the interpreters and can hence lead to erroneous interpretations (paragraph 2). In any case, both groups of examples are one after another.

Are faults planar, listric, both? was this taken into account?

The participants' results included 70 interpretations (44% of the total interpretations) showing curved faults and 62 interpretations (39%) showing planar faults. However, the dip analysis was calculated in a single point (1.1 ms TWT) crossing at approximately the mid-point of all the faults, that we assume is a representative value for the whole of the fault, and this way the calculation is independent of the fault curvature.

It seems that this interpretation did not fulfil the requirement of marking a horizon. I suggest this is not included and discussed the reasons with the other interpretations that were not taken into account.

We use this single left-ward dipping reverse fault interpretation as a proof that this kind of interpretation is largely impossible. Interpreting the horizons was secondary with respect to interpreting the faults, which was really the main requirement.

I suggest you clearly divide the definition of the seismic unit and then how the conceptual model is created based on how the units were defined. For example, you can use subheadings like 1.1 seismic units, 1.2 conceptual model 1, 1.3 conceptual model 2 & 1.4 conceptual model 3). You should consider including your preferred model.

We do not agree with this comment. The units and conceptual models are already clearly separated and numbered. The discussion is also continuous, and we do not envisage how adding subsections will help the flow of the discussion. We also prefer not to include any "preferred model(s)", as we want to study the interpretation results objectively (i.e. we would rather avoid talking about "right" and "wrong" interpretations). Having a preferred model (and clearly stating this preference) would bias the reader towards this model, and could deviate the discussion from its original purpose.

Is this anchoring comparable to the one that you described in the introduction when additional data is given?

See response to question 2 above.

This paragraph should be in the discussion. However, this is not clear from your results.

We disagree with this comment. This paragraph summarises a recommendation (that awareness of biases is important) based on the fact that we found that interpretations were affected by anchoring bias. We do not think that this should be part of the discussion, as this is a further recommendation extracted from the results discussed previously.

Evidence of anchoring to initial conceptual models during ~~Fault~~ interpretation ~~in~~ of a vertically exaggerated seismic section~~: evidence of conceptual model uncertainty and anchoring~~

Juan Alcalde[1,2], Clare E. Bond[2], Gareth Johnson[3], Armelle Kloppenburg[4], Oriol Ferrer[5], Rebecca Bell[6], Puy Ayarza[7]

[1] Department of Structure and Dynamics of the Earth, Institute of Earth Sciences Jaume Almera, ICTJA-CSIC, Lluis Sole i Sabaris s/n, 08028 Barcelona, Spain
[2] Department of Geology and Petroleum Geology, School of Geosciences, University of Aberdeen, Aberdeen AB24 3UE, UK
[3] Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, G1 1XZ, UK.
[4] 4DGeo, Daal en Bergselaan 80, 2565 AH The Hague, The Netherlands
[5] Institut de Recerca Geomodels, Departament de Dinàmica de la Terra i de l'Oceà, Facultat de Ciències de la Terra, Universitat de Barcelona, Barcelona, c/ Martí i Franquès s/n., 08028, Spain
[6] Basins Research Group (BRG), Department of Earth Science & Engineering, Imperial College, Prince Consort Road, London, SW7 2BP, UK
[7] Department of Geology, University of Salamanca, 37008 Salamanca, Spain

Correspondence to: Juan Alcalde (juan.alcalde@abdn.ac.uk)

Abstract.

The use of conceptual models is essential in the interpretation of reflection seismic data. It allows interpreters to make geological sense of seismic data which carries inherent uncertainty. However, conceptual models can create powerful anchors that prevent interpreters from reassessing and adapting their interpretations as part of the interpretation process, which can subsequently lead to flawed or erroneous outcomes. It is therefore critical to understand how conceptual models are generated and applied to reduce unwanted effects in interpretation results. Here we have tested how interpretation of vertically exaggerated seismic data influenced the creation and adoption of the conceptual models of ~~160~~ 161 participants in a paper-based interpretation experiment. Participants were asked to interpret a series of faults and a horizon, off-set by those faults, in a seismic section. The seismic section was randomly presented to the participants with different horizontal-vertical exaggeration (1:4 or 1:2). Statistical analysis of the results indicates that early anchoring to specific conceptual models had the most impact on interpretation outcome; with the degree of vertical exaggeration having a subdued influence. Three different conceptual models were adopted by participants, constrained by initial observations of the seismic data. Interpreted fault dip angles show no evidence of other constraint (e.g. from the application of accepted fault dip models). Our results provide evidence of biases in interpretation of uncertain geological and geophysical data, including the use of heuristics to form initial conceptual models and anchoring to these models, confirming the need for increased understanding and mitigation of these biases to improve interpretation outcomes.

1 Introduction

Reflection seismic data is used to image and understand the subsurface structure of the earth, across scales and tectonic settings (e.g. Park et al., 2002; Simancas et al., 2003; Martí et al., 2008). As with other geophysical methods, seismic images are indirect representations of complex changes in the physical properties of rocks in the subsurface. Seismic images therefore carry inherent uncertainty related to the geological architecture, but also to the acquisition, processing and visualisation methods applied to the seismic data. ~~, making~~These uncertainties combined make seismic images~~ them~~ subject to multiple geological interpretations or, in other words, non-unique solutions (Frodeman, 1995; Rankey and Mitchell, 2003; Bond et al., 2007; Saltus and Blakely, 2011). ~~Interpreters need to apply different conceptual models, acquired during their training and past experience, in order to produce interpretations that honour the data, particularly in areas of great uncertainty (Bond et al., 2007; Bond et al., 2015). These conceptual models are therefore the basis of the interpretation, as they provide the necessary criteria to make sense of the data (Frodeman, 1995).~~

Geoscientists employ mental models (or “conceptual models”) that integrate their observations and that conform ~~with~~to their understanding of the world (Shipley and Tikoff, 2016). When confronted with geological data, interpreters need to apply different conceptual models, acquired during their training and past experience (through learning), 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 the key elements (e.g. growth geometries, regional level) and employ different validation techniques (e.g. balancing or restoration) that allow differentiating between (a priori similar) conceptual models (Bond, 2015). The conceptual models therefore incorporate all the elements that shape the knowledge of the geologist of a certain aspect of the geology; for example, the conceptual model of a turbidite system will include characteristics about their origin and evolution, common stratigraphic sequences, lithological composition, stratigraphic structures associated, etc. 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). Conceptual models are therefore the basis of the interpretation, as they provide the necessary criteria to make sense of the data (Frodeman, 1995).

The use of tectono-sedimentary conceptual models in seismic interpretation has been extensively documented in literature (e.g. Strecker et al., 1999; Nielsen et al., 2008; Alcalde et al., 2014). Understanding what elements influence conceptual model development and application in seismic interpretations is useful to better grasp how ~~erroneous~~ interpretations are made. Applying the appropriate conceptual models requires assessment, by the interpreter, of objective uncertainty, such as considering errors in data processing or acquisition, and of subjective elements, such as the potential biases they bring to the interpretation from their background and experience (Bond, 2015). Alcalde et al., (2017a) argue that image presentation also has a subdued effect in the way seismic image data is perceived and interpreted. Here, we develop this theme investigating how presentation of vertically exaggerated seismic image data influences conceptual model choice and application, and the subsequent interpretation outcome.

Modern computer-based methods provide important advantages to the interpretation of seismic data, such as the generation of 3D models, attribute analysis or the easy access to multiple display options (e.g. change in scales, colour palettes). However, the use of computers generally results in the onscreen interpretation of a vertically exaggerated seismic image, due to the conflicting ratios of a 1:1 seismic image with screen dimensions (Bond, 2015). Furthermore, m~~M~~ost 2D seismic cross-sections published in literature are displayed vertically exaggerated (Stewart, 2011), although it is likely that multiple displays were employed during the interpretation stage.~~,~~ ~~and modern computer based interpretation methods generally result in the onscreen interpretation of a vertically exaggerated seismic image, due to the conflicting ratios of a 1:1 seismic image with screen dimensions (Bond, 2015).~~ Vertical exaggeration of seismic image data creates images with apparent reflection continuity and exaggerates dips of structures and horizons. Conscious application of seismic image stretching is used in the seismic interpretation process because it helps to enhance certain aspects of the display that ease the interpretation (Stewart, 2011). It helps for instance to amplify low relief structures, that appear otherwise compressed and difficult to differentiate (Feagin, 1981; Bertram and Milton, 1996). For example, ~~and~~ Brothers et al. (2009) report that vertical exaggeration helped them to delineate small changes in stratal geometry, otherwise imperceptible, in their seismic interpretation study of the Salton Sea. Vertical exaggeration can also be used to mitigate the difference between vertical and horizontal sampling, which can be considerable depending on the acquisition parameters, the impact of which is to make images appear stretched (Stewart, 2011). These examples highlight the usefulness of scale variation during interpretation.

However, changes in appearance of seismic image data through, sub-conscious or conscious, vertical exaggeration change an interpreter's perception of an image. The change in image character is often unintentional, and can result in unwanted perceptual bias during interpretation, and subsequently lead to misinterpretations, particularly if the interpreted geological structures are complex (Stone, 1991). Vertical exaggeration can also make features, like gas escape chimneys, appear narrower than they are (Horozal et al. 2009). Black et al. (1994) noticed that vertically exaggerated seismic sections can result in gently dipping reflections being perceived as more steeply dipping; which may lead to the erroneous conclusion that migration of the seismic data is required. Similarly, Stewart (2012) investigated the impact of vertical exaggeration on fault dip and observed that structural restoration of interpretations conducted in exaggerated sections lead to unrealistic subsurface models. Thus, vertical exaggeration in seismic interpretation can have positive and negative influences on interpreter perception of the image and interpretation outcome.

Here we test the theory that the presentation of seismic image data in a vertically exaggerated format impacts the perceptions of interpreters, influencing the conceptual models they apply in their interpretation and their ~~final interpretation~~ outcome. We focus on analysis of fault and horizon interpretations in a clipped seismic image. Interpreters were randomly presented with different vertical exaggerations (1:2 and 1:4) of the same seismic image. Statistical analysis of fault and horizon placement, fault dip angle, fault dip direction and fault type, allow us to draw conclusions on the effect of vertical exaggeration on interpretation.

2 Experiment set up

The interpretation experiment consisted of a c. 15 km long clipped portion from a 2D seismic image from the Browse Basin, NW Australia (Figure 1) available on the *Virtual Seismic Atlas* (www.seismicatlas.org). ~~The~~ This seismic image has been interpreted as a series of normal faults dipping to the NW (left hand-side of the section) overlain by post-tectonic sediments, These faults could potentially have been formed in the Late Carboniferous to Early Permian rifting event (Struckmeyer et al., 1998; Keep and Moss, 2000). The area has undergone different stages of reactivation since the Early Triassic, so inversion structures can also be found (Keep and Moss, 2000).

The section used in this experiment was originally downloaded with no vertical exaggeration (i.e. with an approximate horizontal to vertical ratio of 1:1), according to the *Virtual Seismic Atlas* information. In a series of interpretation experiments, the seismic image was presented to participants with horizontal to vertical exaggeration of 1:4 (Figure 2a) or 1:2 (Figure 2b), hereafter called *1:4* and *1:2* sections. The sections were presented in two-way traveltime (TWT) and no information about the actual depth of the sections was provided. The participants were asked to "interpret the main faults crossing the section as deep as possible", as well as to add a "sedimentary horizon to mark the displacement", and were given 15-30 minutes to complete their interpretations. The experiment as presented to the participants can be found in the Supplementary Information.

The participants also completed an anonymous questionnaire designed to collect information about their background, training, knowledge and experience in structural geology and seismic interpretation. The interpretation experiment was completed by ~~160~~ 161 students of which ~~61~~ 126 participants (~~38~~78% of the total) were undergraduate students and ~~99~~ 35 participants (~~62~~22% of the total) were postgraduate students, from different universities in the UK, France and Spain. The participants have mostly geology (72.5%) and geophysics (12.5%) backgrounds and considered themselves as having basic to good proficiency in structural geology and seismic interpretation (>93% of the participants). We focused this experiment on students only to observe the potential variability in interpretation of the same section in a group of people with similar experience and background.

3 Interpretation results

The two vertically exaggerated seismic images were assigned randomly to the participants: the *1:2* section was interpreted 88 times (55%) and the *1:4* section 72 times (45%). The interpretation results were digitised manually and then converted to a 1:1 vertical exaggeration (VE=1:1) for comparison; therefore, the fault dip angles presented in this work are VE=1:1 in time. As the sections were interpreted in TWT, the analysed dips of the faults are not true dips (i.e. these observed in sections in depth), but their relative differences are still comparable. Individual examples of the interpretation results after digitisation from both the *1:2* and *1:4* sections are shown in Figure 3.

Initially, interpretations were grouped based on fault dip direction,~~.~~ ~~The majority of the interpretations dipped in a single direction, either~~ to the left or to the right. Those interpretations with faults dipping in both directions (15 interpretations, 9.4% of the total interpretations), e.g. systems of faults and their conjugates, blank or equivocal interpretations were not included in further analyses. Of the remaining 119 interpretations, m~~M~~ost participants interpreted faults dipping to the right (67 interpretations, 56% of the total interpretations), rather than to the left (52 interpretations, 44% of the total) (Figure 4). The relative proportion is greater in the *1:4* sections (39 interpretations~~59%~~ to the right, 59%) compared to the *1:2* sections (28 interpretations to the right, 53%~~ to the right~~). These two groupings were identified as it was apparent that participants interpreting faults dipping to the right and those interpreting faults dipping to the left had employed two different conceptual models to the data. This resulted in four datasets with two pairs of properties (i.e. 1:2-left, notified as ‘1:2L’, 1:2-right or ‘1:2R’, 1:4-left or ‘1:4L’, and 1:4-right ‘1:4R’) that were further analysed in detail. This subdivision allows us to study if the potential differences can be attributed to the section interpreted (i.e. *1:2* or *1:4*), or to the conceptual model used in the interpretation.

We analysed the fault type (i.e. normal or reverse) and measured the fault dip angle interpreted by the participants. The fault type results do not show significant differences between the *1:2* and *1:4* section interpretations, with 32-33% of the participants interpreting reverse faults and 67-68% interpreting normal faults (Figure 4). However, difference in fault type can be correlated to the dip-direction of the fault (Figure 5). Only one participant (3%) amongst the left-ward dipping datasets (i.e. 1:2L and 1:4L) interpreted the fault motion as reverse, while the vast majority (35 participants, 97% of the total) interpreted leftward-dipping normal faults. In contrast, most right-ward dipping faults were interpreted as reverse (31 interpretations, 56%) instead of normal (24 interpretations, 44%). This result is more pronounced in the 1:4R, with 61% of faults interpreted as reverse (14 interpretations), compared to the 53% in the 1:2R (17 interpretations).

The dip angles of the faults were calculated by drawing a horizontal line at the approximate mid-depth point (1.1 ms TWT) of the seismic section, with the aim of crossing the majority of the faults around their midpoint. Similar numbers of fault interpretations were made on the *1:4* section (a total 300 faults interpreted by 72 participants, over 4 faults interpreted per participant), and the *1:2* section (272 faults by 88 participants, over 3 faults interpreted per participant) (Figure 6). The fault dip angle analyses were compared across the four datasets (Figure 7). ~~Here we observe t~~The ~~biggest~~ largest difference between the *1:4* and *1:2* sections is highlighted here, with the median ~~average~~ dip angle of faults of ~~24º~~ 22º in the right-ward dipping, reverse *1:4* section vs ~~19º~~ 16º in the *1:2* section (Figures 7c and 7d). The differences in normal interpretations, either left-ward (Figure 7a and 7b) or right-ward dipping faults (Figures 7e and 7f), show only differences of 2-3º, and therefore are less conclusive. The fault dip of the only participant interpreting left-ward dipping, reverse faults was 23º on average, ~~halfway between~~slightly higher than the other two groups.

To check if other factors, specifically: educational background and experience, were influencing interpretation outcome we also analysed the data for disparities between different University cohorts and between undergraduate and postgraduate students. There are no major differences in the analysed results across student cohorts from different Universities, or between undergraduate and postgraduate students. For the latter cohort the difference in numbers (undergraduate (126) vs postgraduate (35) students) is large and does not allow easy comparison; despite this the ratios of leftward and rightward dipping faults and the sense of off-set is consistent across the cohorts. 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) and we suggest that this is still an area of interest for future work.

4 Discussion

1. Conceptual model anchoring

Analysis of participants’ interpretations shows that fault interpretations in the seismic image fall into three main categories (Figure 3): (1) left-ward dipping normal faults with right dipping horizons (Figure 3b)~~, corresponding to normal faulting~~; (2) right-ward dipping thrust faults with right-dipping horizons (Figure 3c)~~, corresponding to thrusting~~; and (3) right-ward dipping normal faults with left-dipping horizons (Figure 3d)~~, corresponding to normal faulting~~. Only one interpretation showed left-ward dipping faults with left-dipping horizons and marked the motion of the faults as reverse (Figure 5). In addition, this interpretation did not show any evidence of correlating horizons across the fault and simply used arrows to mark the motion instead. The low number of interpretations of this type (one) and the difficulty in correlation suggests that interpreting left-dipping faults with reverse fault motions is largely impossible, given the reflection seismic characteristics of the data.

Faults and horizons (red and blue lines in Figure 3, respectively) are interpreted in three ways ~~(Figure 3)~~: (1) along left-dipping discontinuous and chaotic reflections, these align with breaks in right-ward dipping reflections that together give the appearance of a left-ward dipping ‘chaotic seismic fabric’ (Figure 3b); (2) along ‘packages’ of right-dipping reflections with greater continuity (Figure 3c); and (3) at an angle to these right-dipping reflections where reflection continuity is less strong (Figure 3d). Irrespective of the vertical exaggeration of the seismic image interpreted, most participants interpreted faults dipping right-ward instead of left-ward (Figure 4). At the same time, the majority of right-ward dipping faults (56%) were interpreted as reverse, in contrast to left-ward dipping faults, which are mostly interpreted as normal (97%) (Figure 7). We suggest that this is as a consequence of the seismic reflection characteristics of the different features that are being interpreted as faults and horizons. ~~Faults and horizons are interpreted in three ways (Figure 3): (1) along left-dipping discontinuous and chaotic reflections, these align with breaks in right-ward dipping reflections that together give the appearance of a left-ward dipping ‘fabric’; (2) along ‘packages’ of right-dipping reflections with greater continuity; and (3) at an angle to these right-dipping reflections where reflection continuity is less strong.~~ The continuity of the right-ward dipping reflections makes them a more ‘certain’ interpretation than the left-ward dipping fabric. When the right-ward dipping reflections are interpreted as horizons, leaving the left-dipping fabric to be interpreted as faults, this invariably leads to interpretation of faults with normal offsets due to the angular relationship between the fault and horizon interpretations and potentially due to the participants interpretation, consciously or sub-consciously, of the nature and geometries of the basin sediments above (Figure 3b). When the right-ward dipping reflections are interpreted as faults, the sedimentary packages are harder to interpret and horizon interpretations are often forced to cut reflections (Figure 3d). When participants have interpreted faults at an angle to the right-ward dipping reflections, where reflection continuity is less strong, this results in steeper fault dip angles, and interpreters often interpret the right-ward dipping reflections as sedimentary packages in horsts between reverse faults (Figures 3c and 7).

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; ~~so a likely~~an -interpretation sequence as shown in ~~(F~~figure 9~~)~~. ~~is that participants interpreted faults first, although w~~We should state that we cannot be sure that ~~this was the case~~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, t~~d~~his has previously been reported by Rankey and Mitchell (2003) and Torvela and Bond (2011). A~~)~~lthough, 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). ~~This is a different anchoring bias as the one described in Rankey and Mitchell (2005), but the effect in the interpretation is similarly strong. There is no evidence in the interpretations that the participants started off on one interpretation track and then changed this to another. Of course, e~~

Experience and knowledge are 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.~~Experience and knowledge is expected to have played a key role in as well, asinforming the initial observations that lead to the selection of a the appropriate 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. (e.g. selecting the appropriate seismo-stratigraphic markers) might be wrong, due to a lack of understanding or of familiarity with conceptual models of normal or thrust faults. However, oThis analysisur data does not shows no strong evidence that experience had an effect oin the interpretation outcomes. We therefore interpret our data as showing that 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 model, irrespective of whether later interpretative elements conform to the data (e.g. horizons cutting reflections, as seen in Figure 3d). In such cases, although i~~although ~~initial interpretations are informed by the data, these first conceptual models are~~ become anchored to and are ~~applied irrespective of whether they later conform to~~ all ~~the data~~, albeit to a threshold~~.~~

~~This has been reported by Rankey and Mitchell (2003) and Torvela and Bond (2011). This suggests that initial conceptual models play a dominant role in interpretation outcome.~~

2. Fault dip variability

Although we purport that the impact of conceptual model application and anchoring to models has the greatest influence on the interpretation outcomes of this experiment, the experiment results show certain differences in fault dip direction and dip angle between the *1:2* and *1:4* vertically exaggerated section interpretations (Figures 4, 6 and 7). Figure 8 shows a projection of the interpreted fault dip angles and their ~~averages~~ median values for both the *1:2* and *1:4* sections on a graph of exaggerated vs unexaggerated dip angles. The interpreted dip angles are projected onto the corresponding curves of vertical exaggeration to show the equivalent unexaggerated dip angle. The same faults interpreted in sections with differing vertical exaggeration should have the same un-exaggerated dip angle (x-axis), but a differing exaggerated dip angle (y-axis). This is the case for the ~~average~~ median of the right-ward and left-ward dipping normal fault interpretations (magenta and dark blue circles in Figure 8, respectively). ~~By inference, this suggests that the same features were interpreted as right-ward dipping faults in both the *1:2* and *1:4* vertically exaggerated seismic sections.~~ In contrast, the ~~average~~ median fault dip angle of the ~~left~~right-ward dipping reverse interpretations in the *1:2* and *1:4* sections (dark pink~~blue~~ circles in Figure 8) are not aligned vertically, indicating that the two cohorts, i.e. participants interpreting the *1:2* and *1:4* sections, did not interpret the same ~~left-ward~~ dipping features as reverse faults. Interpretations of ~~left~~right-ward dipping faults (at least these interpreted as reverse motion) show an apparent impact of vertical exaggeration on interpretation outcome, whereas the ~~right~~left-ward dipping normal fault interpretations do not. In the *1:2* section, interpretations of left-ward dipping faults have higher dip angles on average than those interpreted in the *1:4* section ~~(Figure 8)~~, and a greater spread in fault dip angle (Figure 6e and 6f).

The observations of fault dip angle and motion consistency suggest that those interpreting normal faults (either right-ward or left-ward dipping) ~~faults~~ were unaffected by vertical exaggeration. Note that the interpreted ~~average~~ median right-ward dipping fault dip angles are low, ~~20-21~~15-17º; when these separated into normal and reverse faults, the right-ward dipping normal faults are very low angle ~~14-15~~10-13º (Figure 7e-f), with the reverse faults having higher average dip angles of ~~19-24~~16-22º (Figure 7c-d).

We did not provide the velocity model for the section used, but just for comparison, we converted the faults from TWT to depth assuming a seismic velocity of 3000 ms-1 for the area (following the assumptions and caveats outlined in Stewart, 2011) (table 1). For the reverse motion faults, the resulting dip angles in depth (31-33º) are~~,~~ closer to an Andersonian-predicted reverse fault dip of (30º) and falling within the range of common reverse fault dips of 10º-30º (Anderson, 1905; 1951). The ~~right-ward dipping~~ normal fault angles (14-30º), however, do not conform to predicted Andersonian fault dips of 45-60º ~~(Anderson, 1905; 1951)~~, that are predominant in teaching materials (Alcalde et al., 2017c). The participants did not have access to the regional seismic line, that would have provided context for such low angle normal faults, nor to the actual depth of the sections, so participants may have been expected to attempt to interpret faults with higher dip angles to conform to accepted dip models of normal faults. We see no evidence of this and interpret this observation as data and conceptual model co-confirmation acting dominantly over other reasoning (if any took place).

Table 2: median values in two-way traveltime and their depth-converted equivalent of the 1:2 and 1:4 sections, divided by dip direction and fault motion. The dips were depth-converted using a uniform velocity model of 3000 m/s (as per Stewart, 2011).

Section		1:2			1:4	
Dip direction	Right	Right	Left	Right	Right	Left
Fault motion	Normal	Reverse	Normal	Normal	Reverse	Normal
Median (TWT)	13º	23º	16º	10º	21º	22º
Median (depth-converted)	19º	33º	23º	14º	31º	30º

For the interpretations of left-ward dipping faults, the extent of the vertical exaggeration of the interpreted seismic image appears to have an impact on interpretation outcome. Analysis of fault dip angle from the left-ward dipping fault interpretations of the 1:2 seismic section show a greater range in fault dip angle (standard deviation SD=16º) and a higher ~~average~~ median fault dip angle of ~~34º~~29º, compared to the 1:4 section interpretations with an ~~average~~ median dip angle of ~~24º~~21º, SD=13º (Figure 6e-f), that is, an ~~10º~~ 8º higher ~~average~~ median fault dip in the 1:2 section. If we now consider only the participants’ interpretations that had also interpreted a horizon showing fault motion (Figure 7a & b), the difference in fault dip angle between the 1:2 and 1:4 sections decreases to only ~~3º~~2º, with similar standard deviations of 14º and 13º. We suggest that the differences observed between the 1:2 and 1:4 sections are dominated more by erroneous seismic interpretations than by vertical exaggeration, with those making ‘dubious’ left-ward dipping fault interpretations not completing horizon interpretations. ~~Similarly~~Similarly, for the right-ward dipping fault interpretations normal fault dip angles are low 24º-27º, but not as low as those interpreted to the right, suggesting that the ~~angle of dip of the fault is~~faults are defined more by their seismic character ~~driven more by the seismic image data~~ than by any effects of vertical exaggeration. ~~Tes~~ Testing with more display options (e.g. 1:6 or 1:8 vertical exaggeration) could be helpful to confirm this finding, and would be interesting lines for further enquiry.~~ting with more display options (e.g. 1:6 or 1:8 vertical exaggeration) could be helpful to confirm this output, and is suggested for further work.~~

If we consider the observations described in the light of our knowledge of the perceptual impact of vertically exaggerated seismic images (e.g. Stone, 1991; Black et al. 1994; Horozal et al. 2009; Stewart, 2012), the *1:4* section should perceptually have better reflection continuity due to data compression (Stewart, 2011). The higher apparent reflection continuity in the *1:4* section could make the right-ward dipping reflections appear more dominant and the discontinuities between the sediment packages less dominant and narrower. The smaller range in dip angles for the *1:4* section compared to the *1:2* section (SD=14º vs 16º, respectively, Figure 6a, b) may be the result of this perceptual change. But the lack of consistency in this observation when the data is split between right-ward and left-ward dipping faults (Figure 6) and also into normal and reverse faults (Figure 7), leads us to conclude that vertical exaggeration has ~~little~~ a subdued impact. Our interpretation of these observations is that the seismic data and conceptual model have a more dominant influence on interpretation than any perceptual bias resulting from vertical exaggeration.

Our work does not provide evidence, in this case, to support the conclusions of Stone (1991), Black et al. (1994), Stewart (2011 and 2012) that vertically exaggerated seismic sections causes perceptual bias, compared with the dominant effect of anchoring to conceptual models. We still suggest however, that multiple visualisations of the data should be made, including at a scale of 1:1 and that care should be taken when interpretations of seismic image data have been made in a vertically exaggerated form. Other experimental work (Alcalde et al. 2017b) showed that interpreters and interpretation outcomes were influenced by seismic reflection contrast and continuity, factors that can be enhanced in vertically exaggerated seismic images. We suggest that future work should further investigate the effect of vertical exaggeration on seismic image properties and interpretation outcomes.

We have shown in an interpretation exercise by ~~160~~ 161 participants that:

1. Conceptual models have greater dominance on the interpretation outcome than perceptual bias from interpreting vertically exaggerated seismic sections.
2. Initial conceptual models are anchored to and there is no evidence for reassessment by participants when data does not conform to their initial model.
3. When conceptual models are confirmed, at least initially, by the data, there is no evidence that accepted models, for example in fault dip, have an impact on interpretation outcome, and that variability in interpretation (e.g. fault dips in our experiment) is minimal even if it does not conform to accepted models (e.g. Andersonian dips). Instead, the data drives the interpreted fault dip, and the conceptual model and data co-confirm each other.

Our results support the conclusions of other workers (Rankey and Mitchell, 2003; Bond et al. 2007; 2008) that seismic interpreters need to be aware of potential biases when interpreting seismic image data particularly in the application of conceptual models; and of the high likelihood of anchoring to initial conceptual models even when data does not confirm or conform to the model. Research has shown that awareness of biases (e.g. George et al., 2000) can help mitigate the potential impacts of bias. Thus, seismic interpreters and their employers should employ bias awareness in their interpretation workflows, and obtain multiple opinions to test a broader range of conceptual models (see Bond et al., 2008 for workflow ideas; for reasoning tests to avoid anchoring see Bond, 2015; and Macrae et al., 2016; and for the potential impact of single conceptual models on decision making see Richards et al., 2015). Research into the effectiveness of different bias awareness techniques and their impact in geological interpretation is an obvious focus for future research.

~~Our work does not provide evidence, in this case, to support the conclusions of Stone (1991), Black et al. (1994), Stewart (2011 and 2012) that vertically exaggerated seismic sections causes perceptual bias, compared with the dominant effect of anchoring to conceptual models. We still suggest however, that multiple visualisations of the data should be made, including at a scale of 1:1 and that care should be taken when interpretations of seismic image data have been made in a vertically exaggerated form. Other experimental work (Alcalde et al. 2017b) showed that interpreters and interpretation outcomes were influenced by seismic reflection contrast and continuity, factors that can be enhanced in vertically exaggerated seismic images.~~

~~We suggest that future work should further investigate the effect of vertical exaggeration on seismic image properties and interpretation outcomes.~~ and geophysical data. The resultant interpretation outcomes are not only based on uncertain data, but these uncertainties are compounded by interpretation biases including using heuristics to form initial conceptual models and anchoring to these. Understanding how to better mitigate bias in interpretation and the competing impacts on outcomes of different biases remains a significant challenge in the geosciences.

Competing interests

The authors declare that they have no conflict of interests.

Both JA and CEB conceptualised and designed the interpretation experiment. CEB, AK, OF, RB and PA conducted the experiments at the University of Aberdeen, the University of Grenoble, University of Barcelona, Imperial College of London and University of Salamanca. JA was responsible for the project administration and data analyses. CEB and JA were responsible for the writing, reviewing and editing of the manuscript with help from GJ, AK, OF, RB and PA.

Acknowledgements

We would like to thank both Charlotte Botter and Javier Tamara for their detailed review and suggestions to improve this paper. Ignacio Marzán is also thanked for his help with the figure designs.

Alcalde completed the work presented whilst supported through a NERC industry partnership grant (NE/M007251/1) and is currently funded by EIT Raw Materials – SIT4ME project (17024). Bond is currently funded through a Royal Society of Edinburgh research sabbatical on uncertainty in seismic image interpretation. Johnson is funded by the University of Strathclyde Faculty of Engineering. Ferrer has been supported by the SALCONBELT Project (CGL2017-85532-P), the Geomodels Research Institute and the Grup de Geodinàmica i Anàlisi de Conques (2017SGR-596).

Special issue statement

This article is part of the special issue “Understanding the unknowns: the impact of uncertainty in the geosciences”.

References

Alcalde, J., Marzán, I., Saura, E., Martí, D., Ayarza, P., Juhlin, C., Perez-Estaun, A. and Carbonell, R.: 3D geological characterization of the Hontomín CO_2 storage site, Spain: Multidisciplinary approach from seismic, well-log and regional data. Tectonophysics, 627, pp.6-25, 2014.

Alcalde, J., Bond, C.E. and Randle, C.H.: Framing bias: The effect of figure presentation on seismic interpretation. Interpretation, 5(4), pp.T591-T605, 2017a.

Alcalde, J., Bond, C.E., Johnson, G., Ellis, J.F. and Butler, R.W.: Impact of seismic image quality on fault interpretation uncertainty. GSA Today, 2017b.

Alcalde, J., Bond, C.E., Johnson, G., Butler, R.W., Cooper, M.A. and Ellis, J.F.: The importance of structural model availability on seismic interpretation. Journal of Structural Geology, 97, pp.161-171, 2017c.

Anderson, E.M. The dynamics of faulting: Transactions of the Edinburgh Geological Society, v. 8, p. 387–402, doi: 10.1144/transed.8.3.387, 1905.

Anderson, E.M.: The Dynamics of Faulting and Dyke Formation with Application to Britain, Second Edition: Edinburgh, Oliver and Boyd, 206 p, 1951.

Black, R.A., Steeples, D.W. and Miller, R.D.: Migration of shallow seismic reflection data. Geophysics, 59(3), pp.402-410, 1994.

Bertram, G.T., Milton, N.J.: Seismic stratigraphy. In: Emery, D., Myers, K., Bertram, G.T. (Eds.), Sequence Stratigraphy. Blackwell Science Ltd, Oxford, pp. 45-60, 1996.

Bond, C.E., Gibbs, A.D., Shipton, Z.K. and Jones, S.: What do you think this is? ``Conceptual uncertainty" in geoscience interpretation. GSA today, 17(11), p.4, 2007.

Bond, C.E., Shipton, Z.K., Gibbs, A.D. and Jones, S.: Structural models: optimizing risk analysis by understanding conceptual uncertainty. *First Break*, *26*(6), pp.65-71, 2008.

Bond, C.E.: Uncertainty in structural interpretation: Lessons to be learnt. Journal of Structural Geology, 74, pp.185-200, 2015.

Bond, C.E., Johnson, G. and Ellis, J.F.: Structural model creation: the impact of data type and creative space on geological reasoning and interpretation. Geological Society, London, Special Publications, 421(1), pp.83-97, 2015.

Brothers, D.S., Driscoll, N.W., Kent, G.M., Harding, A.J., Babcock, J.M. and Baskin, R.L.: Tectonic evolution of the Salton Sea inferred from seismic reflection data. Nature Geoscience, 2(8), p.581, 2009.

Feagin, F.J.: Seismic data display and reflection perceptibility. Geophysics, 46(2), pp.106-120, 1981.

Frodeman, R.: Geological reasoning: geology as an interpretive and historical science. Geol. Soc. Am. Bull. 107 (8), 960e968, 1995.

George, J.F., Duffy, K. and Ahuja, M.: Countering the anchoring and adjustment bias with decision support systems. Decision Support Systems, 29(2), pp.195-206, 2000.

Horozal, S., Lee, G.H., Bo, Y.Y., Yoo, D.G., Park, K.P., Lee, H.Y., Kim, W., Kim, H.J. and Lee, K. Seismic indicators of gas hydrate and associated gas in the Ulleung Basin, East Sea (Japan Sea) and implications of heat flows derived from depths of the bottom-simulating reflector. Marine Geology, 258(1-4), pp.126-138, 2009.

Kahneman, D., Slovic, P. & Tversky, A. (eds).: Judgement under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, 1982.

Keep, M. and Moss, S.J.: Basement reactivation and control of Neogene structures in the outer Browse Basin, North West Shelf. Exploration Geophysics, 31(1/2), pp.424-432, 2000.

Macrae, E.J., Bond, C.E., Shipton, Z.K. and Lunn, R.J.: Increasing the quality of seismic interpretation. *Interpretation*, *4*(3), pp.T395-T402, 2016.

Martí, D., Carbonell, R., Flecha, I., Palomeras, I., Font-Capó, J., Vázquez-Suñé, E. and Pérez-Estaún, A.: High-resolution seismic characterization in an urban area: Subway tunnel construction in Barcelona, Spain. Geophysics, 73(2), pp.B41-B50, 2008.

Nielsen, T., Knutz, P.C. and Kuijpers, A.: Seismic expression of contourite depositional systems. Developments in Sedimentology, 60, pp.301-321, 2008.

Park, J.O., Tsuru, T., Kodaira, S., Cummins, P.R. and Kaneda, Y. Splay fault branching along the Nankai subduction zone. Science, 297(5584), pp.1157-1160, 2002.

Rankey, E.C. and Mitchell, J.C.: That's why it's called interpretation: Impact of horizon uncertainty on seismic attribute analysis. The Leading Edge, 22(9), pp.820-828, 2003.

Richards, F.L., Richardson, N.J., Bond, C.E. and Cowgill, M.: Interpretational variability of structural traps: implications for exploration risk and volume uncertainty. *Geological Society, London, Special Publications*, *421*(1), pp.7-27, 2015.

Saltus, R.W. and Blakely, R.J.: Unique geologic insights from "non-unique" gravity and magnetic interpretation. GSA Today, 21(12), pp.4-11, 2011.

Shipley, T.F., Tikoff, B. Linking Cognitive Science and Disciplinary Geoscience Practice, in: 3-D Structural Interpretation. American Association of Petroleum Geologists, pp. 219–237, 2016. https://doi.org/10.1306/13561994M1113677

Shipley, T.F., Tikoff, B. The Role of Geoscience Education Research in the Consilience between Science of the Mind and Science of the Natural World. J. Geosci. Educ. 65, 393–398, 2017. https://doi.org/10.5408/16-252.1

Simancas, J.F., Carbonell, R., González Lodeiro, F., Pérez Estaún, A., Juhlin, C., Ayarza, P., Kashubin, A., Azor, A., Martínez Poyatos, D., Almodovar, G.R. and Pascual, E.: Crustal structure of the transpressional Variscan orogen of SW Iberia: SW Iberia deep seismic reflection profile (IBERSEIS). Tectonics, 22(6), 2003.

Stewart, S.A.: Vertical exaggeration of reflection seismic data in geoscience publications 2006–2010. Marine and Petroleum Geology, 28(5), pp.959-965, 2011.

Stewart, S.A.: Interpretation validation on vertically exaggerated reflection seismic sections. Journal of Structural Geology, 41, pp.38-46, 2012.

Stone, D.S.: Analysis of Scale Exaggeration on Seismic Profiles (1). AAPG Bulletin, 75(7), pp.1161-1177, 1991.

Strecker, U., Steidtmann, J.R. and Smithson, S.B.: A conceptual tectonostratigraphic model for seismic facies migrations in a fluvio-lacustrine extensional basin. AAPG bulletin, 83(1), pp.43-61, 1999.

Struckmeyer, H.I., Blevin, J.E., Sayers, J., Totterdell, J.M., Baxter, K., Cathro, D.L., Purcell, P.G. and Purcell, R.R.: Structural evolution of the Browse Basin. North West Shelf. new concepts from deep seismic data. 345-367, 1998.

Torvela, T. and Bond, C.E.: Do experts use idealised structural models? Insights from a deepwater fold–thrust belt. *Journal of Structural Geology*, *33*(1), pp.51-58, 2011.

Tversky, A. and Kahneman, D.: Judgment under uncertainty: Heuristics and biases. science, 185(4157), pp.1124-1131, 1974.

Vermeer, G.J.: Factors affecting spatial resolution. Geophysics, 64(3), pp.942-953, 1999.

Figure captions

Figure 1: Regional seismic image from the Browse Basin (NW Australia). The black box marks the area of the section used in the interpretation experiment. Note that the vertical exaggeration of the image is high (1:8). The full section, with approximately no vertical exaggeration (i.e. 1:1 display) can be downloaded from the VSA website (www.vsa.org).

Figure 2: Seismic sections used in the interpretation experiment with a) 1:4 vertical exaggeration and b) 1:2 vertical exaggeration.

Figure 3: The seismic section and sketch interpretations of the three ~~main conceptual~~interpretation models applied in interpretations of the seismic section (a) by participants. ~~a~~b) left-dipping normal faults (in red) with right-dipping horizons (in blue); ~~b~~c) right-dipping normal faults with left-dipping horizons~~right-dipping reverse faults with right-dipping horizons~~; and ~~c~~d) right-dipping reverse faults with right-dipping horizons~~right-dipping normal faults with left-dipping horizons~~.

Figure 4: Statistics for the interpreted fault directions (left 'L' or right 'R'), and motions (normal 'N' or reverse 'R') in the sections with 1:4 and 1:2 vertical exaggeration ('VE'). The number of participants is given in brackets. Note that ambiguous interpretations (e.g. left + right-dipping fault interpretations, or no faults interpreted), corresponding to 41 interpreters (25.6% of the total), were excluded from the count.

Figure 5: Statistics for the interpreted fault directions (left 'L' or right 'R'), and motions (normal 'N' or reverse 'R'), separated by vertical exaggeration ('VE') 1:2 or 1:4~~)~~.

Figure 6: Rose diagrams showing the dips of interpreted faults. Fault dips interpreted at a vertical exaggeration of: a) 1:4, b) 1:2, c) 1:4 dipping right-ward ("R") d) 1:2 dipping right-ward, e) 1:4 dipping left-ward ("L") and e) 1:2 dipping left-ward. The 'n' marks the number of faults analysed. 'SD' stands for standard deviation.

Figure 7: Rose diagrams showing the dips of interpreted faults and their motion. Fault dips interpreted at a vertical exaggeration of: a) 1:4, left-ward dipping and normal, b) 1:2, left-ward dipping and normal, c) 1:4 right-ward dipping and reverse, d) 1:2 right-ward dipping and reverse, e) 1:4 right-ward dipping and normal, f) 1:2 right-ward dipping and normal. Note that there are fewer faults presented here than in Figure 6 due to fewer participants interpreting the fault motion. The "n" marks the number of faults analysed. "SD" stands for standard deviation.

Figure 8: Graph adapted from Stewart (2011) showing ~~of~~ exaggerated and un-exaggerated dip values for all fault interpretations, showing the average fault dips for left-ward and right-ward dipping faults interpreted at 1:2 and 1:4 vertical ~~exagertion~~exaggeration~~, graph from adapted from Stewart (2011)~~. The medians of the dip direction and fault motion sub cohorts are also presented in the 1:2 and 1:4 curves.

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.

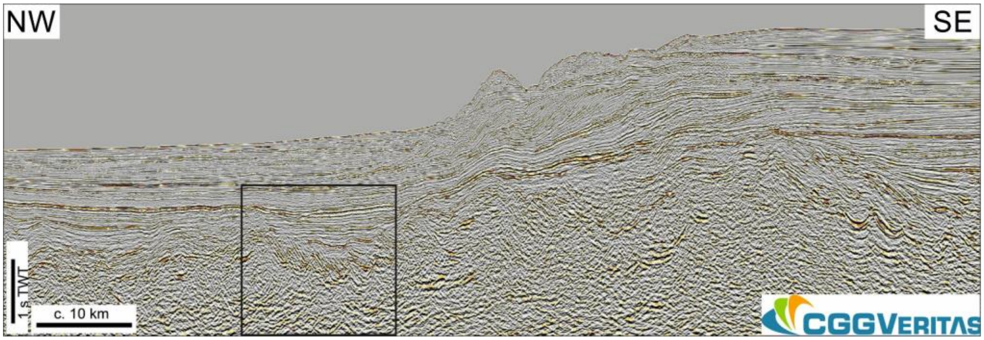


Figure 1

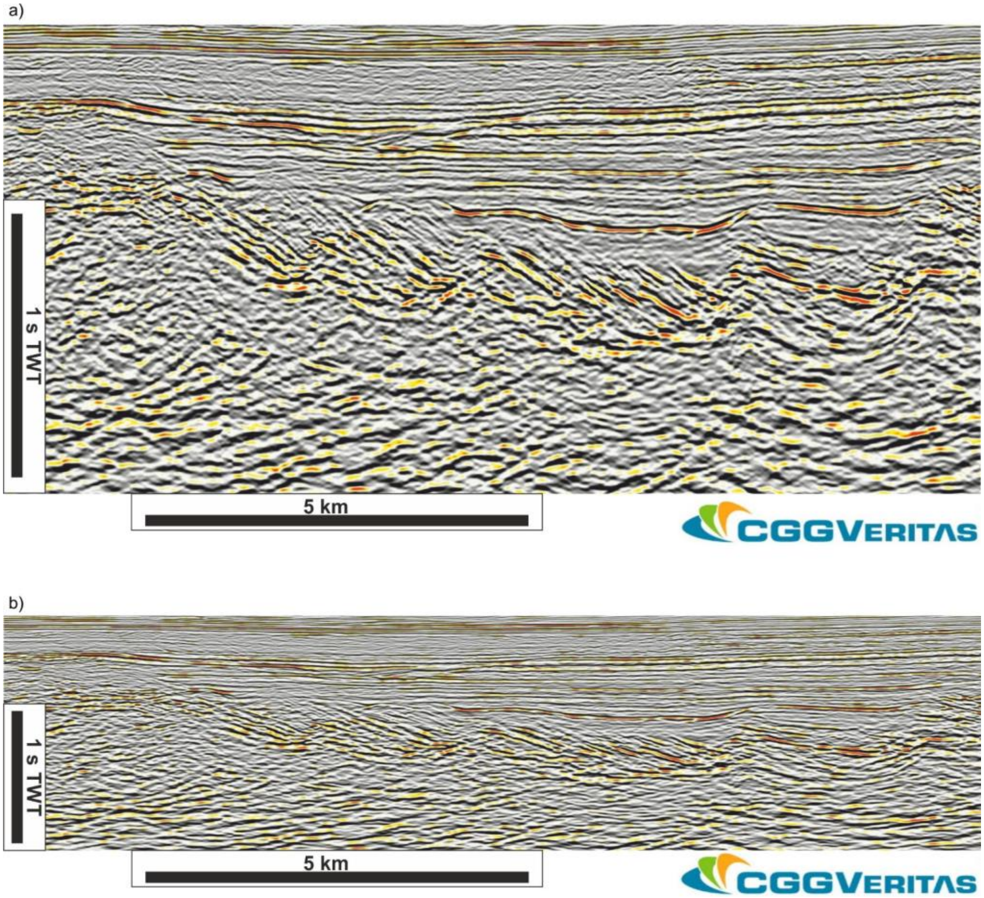


Figure 2

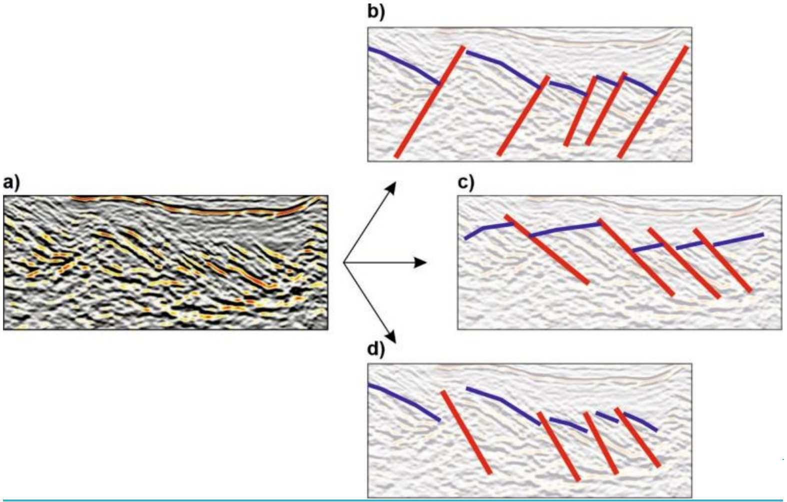


Figure 3

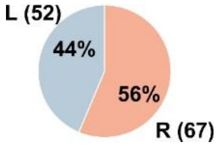
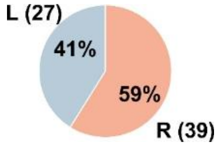
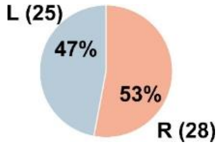
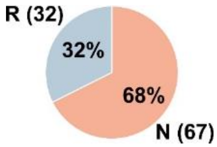
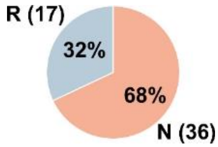
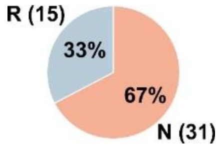
	Aggregated Total	**VE = 1:4** (88 interpretations – 55%)	**VE = 1:2** (72 interpretations – 45%)
Fault direction L – left R – right			
Fault type R – reverse N – normal			

Figure 4

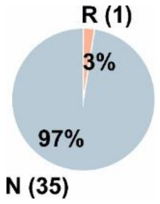
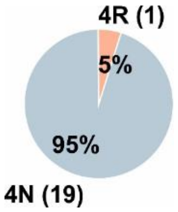
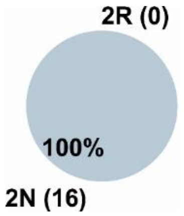
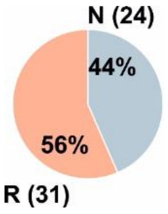
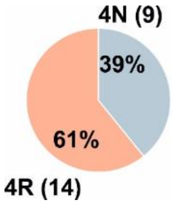
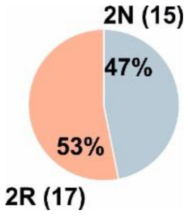
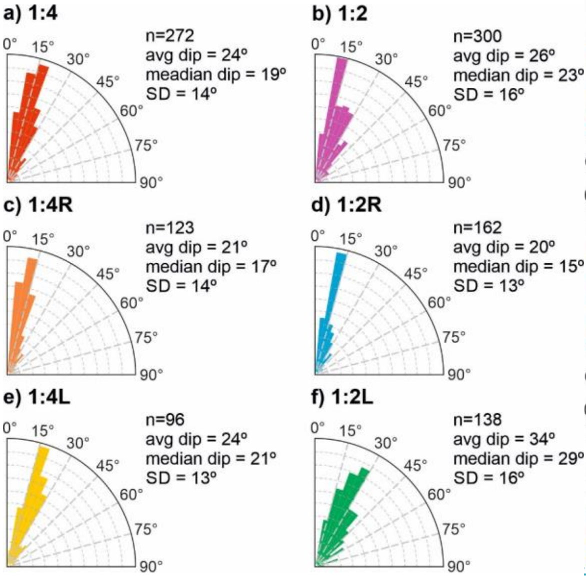
	Fault motion R – reverse N – normal	**VE = 1:4** + **Fault motion**	**VE = 1:2** + **Fault motion**
LEFT dipping			
RIGHT dipping			

Figure 5


a) 1:4
n=272
avg dip = 24°
meadian dip = 19°
SD = 14°
b) 1:2
n=300
avg dip = 26°
median dip = 23°
SD = 16°
c) 1:4R
n=123
avg dip = 21°
median dip = 17°
SD = 14°
d) 1:2R
n=162
avg dip = 20°
median dip = 15°
SD = 13°
e) 1:4L
n=96
avg dip = 24°
median dip = 21°
SD = 13°
f) 1:2L
n=138
avg dip = 34°
median dip = 29°
SD = 16°
0°
15°
30°
45°
60°
75°
90°

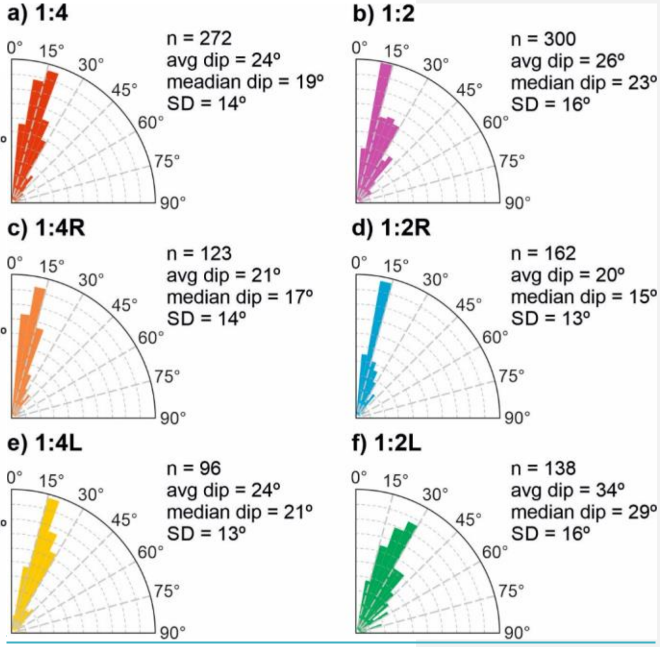
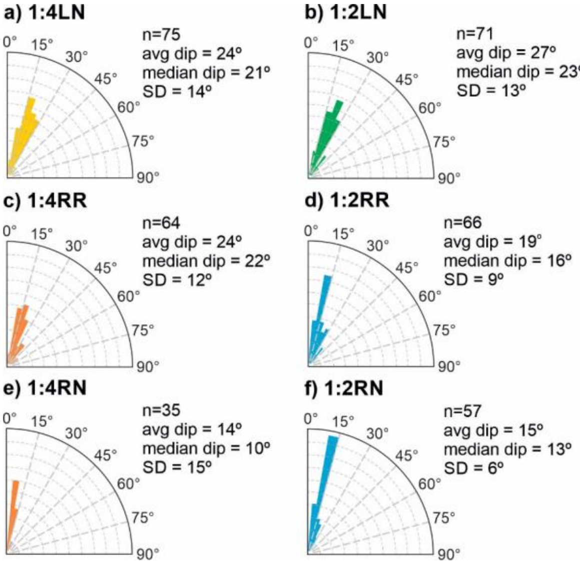

a) 1:4
n = 272
avg dip = 24°
meadian dip = 19°
SD = 14°
b) 1:2
n = 300
avg dip = 26°
median dip = 23°
SD = 16°
c) 1:4R
n = 123
avg dip = 21°
median dip = 17°
SD = 14°
d) 1:2R
n = 162
avg dip = 20°
median dip = 15°
SD = 13°
e) 1:4L
n = 96
avg dip = 24°
median dip = 21°
SD = 13°
f) 1:2L
n = 138
avg dip = 34°
median dip = 29°
SD = 16°
0°
15°
30°
45°
60°
75°
90°

Figure 6


a) 1:4LN
n=75
avg dip = 24°
median dip = 21°
SD = 14°
b) 1:2LN
n=71
avg dip = 27°
median dip = 23°
SD = 13°
c) 1:4RR
n=64
avg dip = 24°
median dip = 22°
SD = 12°
d) 1:2RR
n=66
avg dip = 19°
median dip = 16°
SD = 9°
e) 1:4RN
n=35
avg dip = 14°
median dip = 10°
SD = 15°
f) 1:2RN
n=57
avg dip = 15°
median dip = 13°
SD = 6°
0°
15°
30°
45°
60°
75°
90°

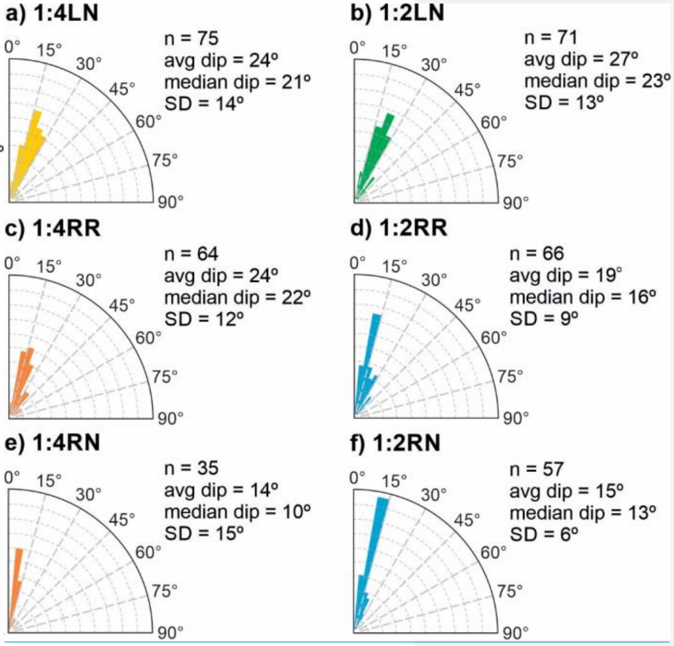

a) 1:4LN
n = 75
avg dip = 24°
median dip = 21°
SD = 14°
b) 1:2LN
n = 71
avg dip = 27°
median dip = 23°
SD = 13°
c) 1:4RR
n = 64
avg dip = 24°
median dip = 22°
SD = 12°
d) 1:2RR
n = 66
avg dip = 19°
median dip = 16°
SD = 9°
e) 1:4RN
n = 35
avg dip = 14°
median dip = 10°
SD = 15°
f) 1:2RN
n = 57
avg dip = 15°
median dip = 13°
SD = 6°
0°
15°
30°
45°
60°
75°
90°

Figure 7

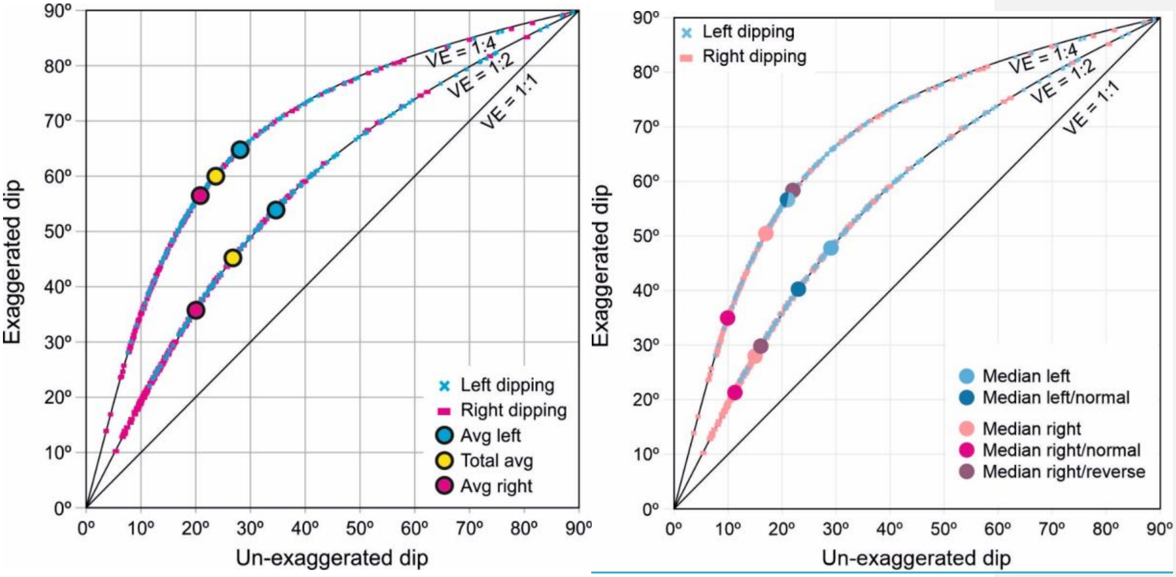

90º
80º
70º
60º
50º
40º
30º
20º
10º
0º
Exaggerated dip
0º
10º
20º
30º
40º
50º
60º
70º
80º
90º
Un-exaggerated dip
VE = 1:4
VE = 1:2
VE = 1:1
Left dipping
Right dipping
Avg left
Total avg
Avg right
90º
80º
70º
60º
50º
40º
30º
20º
10º
0º
Exaggerated dip
0º
10º
20º
30º
40º
50º
60º
70º
80º
90º
Un-exaggerated dip
Left dipping
Right dipping
VE = 1:4
VE = 1:2
VE = 1:1
Median left
Median left/normal
Median right
Median right/normal
Median right/reverse

Figure 8

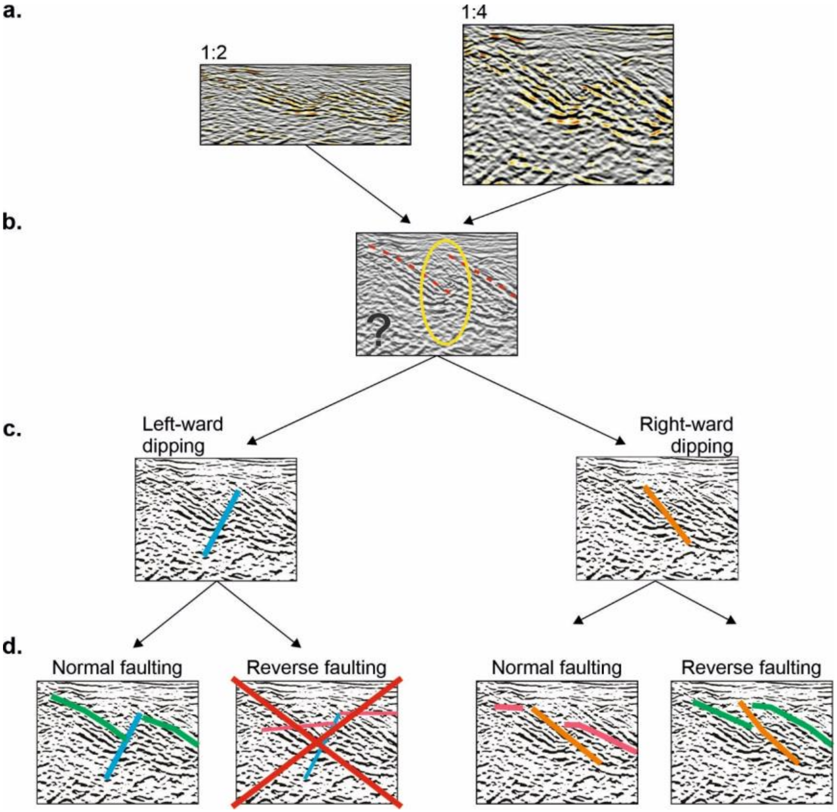

a.
1:2
1:4
b.
?
c.
Left-ward dipping
Right-ward dipping
d.
Normal faulting
Reverse faulting
Normal faulting
Reverse faulting

Figure 9