Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Structural model creation: the impact of data type and creative space on geological reasoning and interpretation

C. E. BOND[1]*, G. JOHNSON[2] & J. F. ELLIS[3]

[1]*Geology and Petroleum Geology, University of Aberdeen, School of Geosciences, King College, Aberdeen AB24 3UE, UK*

[2]*School of Geosciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK*

[3]*Midland Valley Exploration, 144 West George Street, Glasgow G2 2HG, UK*

**Corresponding author (e-mail: clare.bond@abdn.ac.uk)*

Abstract: Interpretation of sparse or incomplete datasets is a fundamental part of geology, particularly when building models of the subsurface. Available geological data are often remotely sensed (seismic data) or very limited in spatial extent (borehole data). Understanding how different datasets are interpreted and what makes an interpreter effective is critical if accurate geological models are to be created. A comparison of the interpretation outcome and techniques used by two cohorts interpreting different geological datasets of the same model, an inversion structure, was made. The first cohort consists of interpreters of the synthetic seismic image data in Bond *et al.* ('What do you think this is?: "Conceptual uncertainty" in geoscience interpretation', *GSA Today*, 2007, **17**, 4–10, http://dx.doi.org/10.1130/GSAT01711A.1); the second cohort is new and interpreted borehole data. The outcomes of the borehole interpretation dataset support earlier findings that technique use, specifically evidence of geological evolution thought processes, results in more effective interpretation. The results also show that the borehole interpreters were more effective at arriving at the correct interpretation. Analysis of their final interpretations in the context of psychological and medical image analysis research suggests that the clarity of the original dataset, the amount of noise and white space may play a role in interpretation outcome, through enforced geological reasoning during data interpretation.

Gold Open Access: This article is published under the terms of the CC-BY 3.0 license.

Subsurface geological models play an important part in decision-making in industrial geology. Geological models delineate the key features of interest in the subsurface. The depths and geometries of features define calculations of volumes of potential hydrocarbons and mineral systems, which determine the economic viability of further exploration and ultimately production or extraction. In the storage industries (gas, radioactive waste or CO_2), similar determinations are made based on the geological model with the addition of assessing a storage site for leakage of the stored material. As subsurface models are based on incomplete datasets with limited resolution (i.e. boreholes, seismic images), constructed models are necessarily an interpretation. The subsurface model created could be one of several conceptual solutions (Bond *et al.* 2007), as well as there being precisional uncertainty in the placement of key elements, for example, horizons and faults (Torvela & Bond 2011; Lark *et al.* 2013). Understanding the techniques and interpretation strategies employed to create subsurface geological models is therefore an important element in risking models.

Previous studies have investigated: differences in interpretational outcomes in creating geological models (Rankey & Mitchell 2003; Bond *et al.* 2007); the dynamics of group decision-making in geological model interpretation (Polson & Curtis 2010); the experience and techniques used to make an effective interpretation (Bond *et al.* 2012); workflows for optimizing model choice to minimize risk and deal with model uncertainties (Refsgaard *et al.* 2006; Bond *et al.* 2008); and the calculation of differences between uncertainties in possible models (Tacher *et al.* 2006; Wellmann *et al.* 2010; Wellmann & Regenauer-Lieb 2012; Lindsay *et al.* 2013). Most previous studies have focused on the interpretation of seismic data or integrated seismic and borehole datasets. Work on choosing between multiple possible models, using expert elicitation, has been undertaken in hydrology (Ye *et al.* 2008) and the recent work of Lindsay *et al.* (2012, 2013) has focused on terranes of interest to the mining sector.

In this study we present multiple interpretations by geoscientists based purely on synthetic borehole data created from a known geological model.

From: Richards, F. L., Richardson, N. J., Rippington, S. J., Wilson, R. W. & Bond, C. E. (eds) 2015. *Industrial Structural Geology: Principles, Techniques and Integration*. Geological Society, London, Special Publications, **421**, 83–97. First published online February 11, 2015, http://dx.doi.org/10.1144/SP421.4
© 2015 The Author(s). Published by The Geological Society of London.
Publishing disclaimer: www.geolsoc.org.uk/pub_ethics

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

In addition, we compare these borehole-based interpretations with interpretations by geoscientists from a synthetic seismic image covering the same geological model. Our aim is to better understand interpretational strategies and techniques given different datasets, through comparison of the differences in interpretational outcome and techniques employed by the two cohorts.

Methodology

The two cohorts of geoscientists (our interpreters) were faced with different geological datasets of the same subsurface geological model. As a basis for the experiment we used the dataset of Bond *et al.* (2007). Our first cohort are the interpreters of Bond *et al.* (2007) who interpreted a synthetic seismic image; the second cohort are new, and were given borehole data based on the geological model used to create the synthetic seismic image.

Details of cohort 1, the 2D geological model and the synthetic seismic image they interpreted are given in Bond *et al.* (2007, 2012). We show in Figure 1 a series of images depicting the forward modelling used to construct the final geological model (Fig. 1d) from which the synthetic seismic image was created. In summary, the geological model is based on an initial layer-cake stratigraphy, which has undergone extensional faulting, followed by compression, inverting the structure. The overall tectonic concept for the model is inversion, which is picked out by syntectonic sedimentation (Fig. 1d). The interpreters in cohort 1 were self-selected geoscientists who undertook the interpretation exercise at a range of petroleum geology conferences and workshops, between 2005 and 2006. Cohort 1 consisted of 411 individual interpreters. Cohort 2 all attended the workshop 'Modelling Structural Evolution to Improve 3D Models for Exploration and Mine Development' organized by the Society of Economic Geologists and Midland Valley Exploration in Denver, 25–26 October 2012. They had self-selected to attend the two-day workshop. As with the synthetic seismic cohort (1), the borehole cohort (2) filled in a questionnaire about their background experience, training and education. Bond *et al.* (2007) describe the details of the questionnaire.

In total 19 geoscientists completed the borehole interpretation and questionnaire to form cohort 2. Analysis of the questionnaire returns showed that the borehole cohort were mainly mineral deposit geologists, unlike cohort 1, who were mainly oil and gas professionals or academics. Cohort 2 ranged from having basic to specialist knowledge of structural geology and interpretation and all but four interpreted at least monthly (80%), seven (37%) daily, three (16%) weekly, five (26%) monthly, one (5%) 6-monthly, two (10%) yearly and one (5%) almost never. Their experience ranged from 1.5 to 40 years in exploration and mine development. A comparison of the backgrounds and prior knowledge of the interpreters in the two cohorts is given in Table 1. Apart from the difference in professional background, the main difference between the two cohorts is in dominant career tectonic

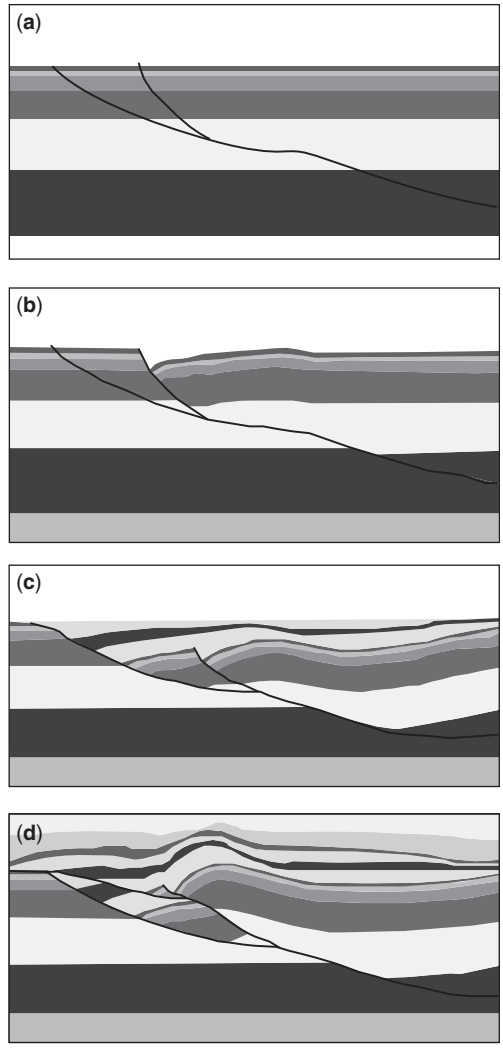

Fig. 1. The forward geological model used to create the two datasets: (**a**) original layer-cake stratigraphy and position of fault system; (**b**) initial normal faulting; (**c**) continued normal faulting with syntectonic sedimentation and fault step-back; (**d**) the final geometry used for the geological model showing inversion of the normal fault, with fault linkage, during continued sedimentation.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Table 1. *The background experience and prior knowledge of the two cohorts*

	Professional experience	Number of years of experience	Education	Dominant tectonic setting (career experience)	Specialist in structural geology
Cohort 1	Academia 32% Oil and gas 39% Other 25% No response 4%	None 5% Student 27% 0–5 years 14% 5–10 years 10% 10–15 years 7% 15 + years 33% No response 5%	Bachelors 29% Postgraduate degree 71%	Extension 18% Inversion 2% Thrust 13% Salt/diapirism 4% Strike-slip 12% Other 36% More than one choice 7% No response 8%	Specialist in structural geology 27% Other categories 61% No response 12%
Cohort 2	Exploration/mine development 63% No response 37%	None 0% Student 0% 0–5 years 32% 5–10 years 11% 10–15 years 5% 15 + years 47%	Bachelors 32% Postgraduate degree 68%	Extension 26% Thrust 26% Strike-slip 11% Polyphase 26% All 5% No response 6%	Specialist in structural geology 16% Other categories 84% No response 0%

Significant differences: professional experience – cohort 1 are mainly oil and gas professionals or academics, whereas cohort 2 is dominated by geoscientists working in the mining and mineral exploration sector; and dominant career tectonic setting – 26% of cohort 2 choose polyphase deformation, compared with only 2% of cohort 1 choosing inversion. Despite the differences in sample size, both the years of experience and education levels are similar for both cohorts, as is the percentage of specialist structural geologists.

setting, where 26% of cohort 2 chose polyphase deformation, compared with only 2% of cohort 1 choosing inversion.

The interpreters in the borehole cohort (2) were given 20 min to complete the exercise. The length of time given was determined by the average time taken for interpreters in cohort 1 to complete their interpretations. Cohort 2 completed their interpretations at the start of the workshop, before they had covered any of the course material. Structuring the experiment in this way minimized the potential for bias resulting from participation in the workshop. Common biases that have the potential to affect geological interpretation include availability bias, the most readily available model or interpretation that is brought to mind (Baddeley *et al.* 2004; Bond *et al.* 2008). The timing meant that the interpreters had only the prior knowledge that they brought to the workshop in advance of undertaking the exercise.

Cohort 1 were given the seismic image data to interpret, on a sheet of A4 paper with no other information to aid their interpretation (Fig. 2a; Bond *et al.* 2007). In contrast, cohort 2 were given 11 boreholes coloured by stratigraphic unit and with faults identified, surface outcrop dips and stratigraphic information, and a stratigraphic column with regional thicknesses. In contrast to cohort 1 the data were given in depth rather than time (Fig. 2b) and the information was printed over three landscape-oriented A4 sheets. Figure 2c shows the seismic image interpreted by cohort 1 overlain with the original geological model (transparent), with the topography and boreholes interpreted by cohort 2 annotated for reference. The collected interpretations and questionnaires completed by cohort 2 were analysed and categorized. The physical expressions of the interpretations were assessed and the questionnaire data collated into a spreadsheet for direct comparison with the existing cohort 1 data.

Each interpretation of the borehole data (cohort 2) was also digitized for direct comparison, and comparison with the original model. Interpretations were scanned and digitized in Adobe Illustrator, using a constant pixel size (line thickness) for all interpretations, with the centre of the original interpretation line forming the digitized trace. Image analysis software (Fiji) was used to construct a stack of digitized interpretations for comparison; the topography and boreholes were used to ensure perfect scaling of all the images. The stacking allows analysis of interpretation intensity based on their spatial overlap.

Results

Impact of experience and techniques

Similarly to the statistical analysis of Bond *et al.* (2012), comparisons were made for the two cohorts based on: (a) the tectonic concept they applied to the dataset; (b) the number of techniques they used to interpret the data; and (c) their self-assessed

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

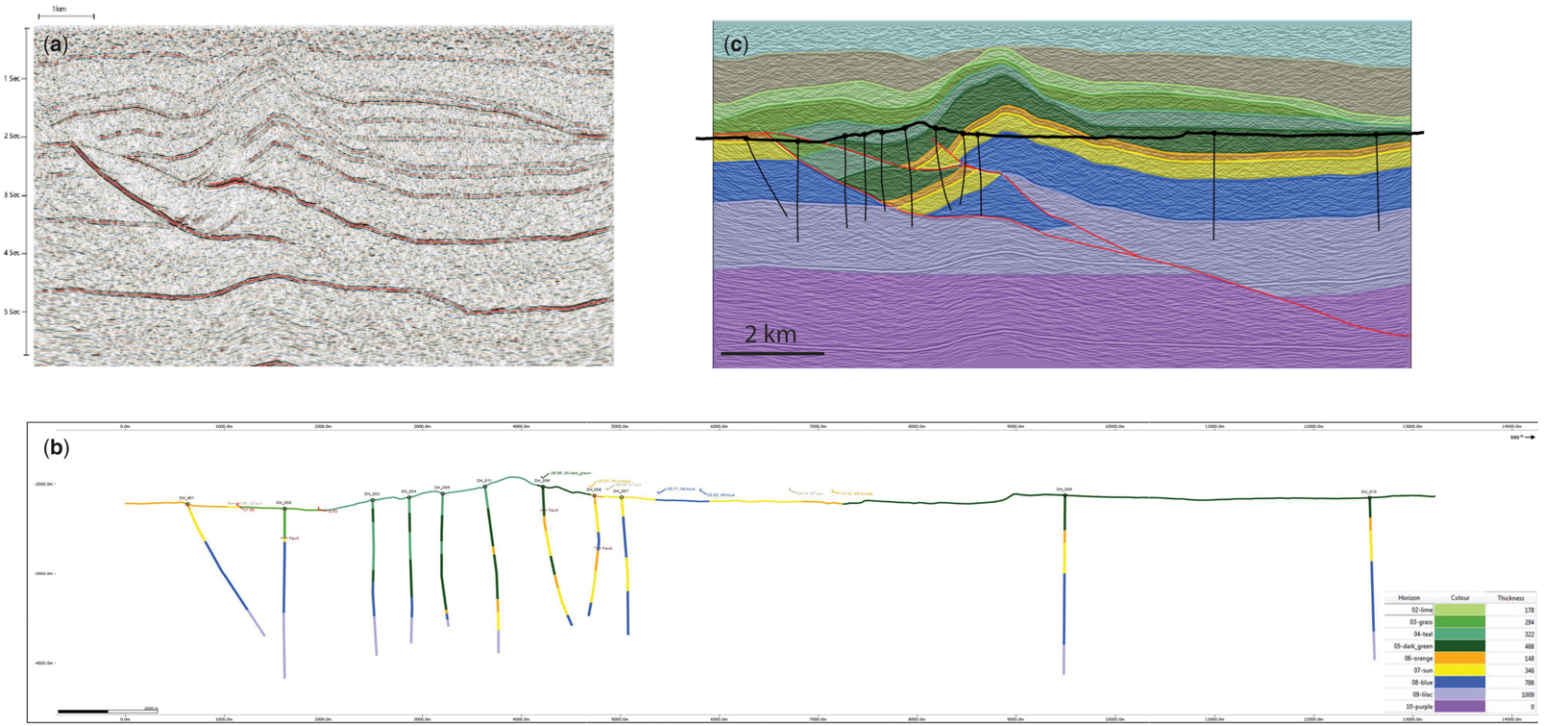

Fig. 2. Geological model and datasets used in the interpretation exercise. (**a**) Synthetic seismic dataset shown as presented for interpretation by cohort 1 (from Bond *et al.* 2007). (**b**) The borehole and topographic data as presented for interpretation by cohort 2. (**c**) Original geological model, shown transparently coloured by stratigraphic unit, faults in red; behind this is the synthetic seismic created from the model interpreted by cohort 1. The black lines show a topographic surface and 11 boreholes used to create the dataset for cohort 2.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

experience as a structural geologist and interpreter. Finally we looked at (d) the types of technique used and if these had a bearing on interpretational success. The analysis allowed a qualitative assessment of the cohort 2 dataset to assess if the same factors as those identified for cohort 1 affected interpretation effectiveness when interpreting borehole data.

Tectonic concept. The first comparison of the two cohorts is based on the tectonic concept applied to the data by the interpreters. The following tectonic concepts were used to categorize all interpretations: extension, inversion, thrust, strike-slip, salt/shale, other and unclear. The criteria for tectonic concept categorization are outlined in full in Bond *et al.* (2007), based on the premise that specific illustration of fault movement is required to define each concept (e.g. arrows on faults, offset correlated horizons, text stating the concept). Issues with the categorization criteria used are discussed in Bond *et al.* (2007), but in summary result in few strike-slip interpretations and a significant 'unclear' category. Figure 3 shows the range of tectonic concepts applied as a percentage by each cohort. The synthetic seismic interpreters (cohort 1, Fig. 3a) showed a greater range of concepts applied to the dataset than those interpreting the boreholes in cohort 2 (Fig. 3b). The most commonly applied tectonic concept by cohort 1 was thrusting (26%). The second most commonly applied tectonic concept was the 'correct' or modelled concept of inversion (23%). In contrast 69% of cohort 2 interpreted the borehole data as inversion, with the same percentage interpreting a thrust concept (26%) as cohort 1.

Number of techniques used. Technique use may also be thought of as interpretational style or method. The categorizations for techniques are based on analysis of the final interpretations of cohort 1. The technique categorizations are: horizon interpretation; fault interpretation; annotations (arrows and text, or key words: e.g. fault, syncline, extension, growth); geological evolution (any evidence that the interpreter has thought about the geological evolution, e.g. evolutionary sketches, numbering sequences of events); sticks (straight lines to represent faults, see Fig. 6c for a good example of use of this technique); and descriptive writing (multiple sentences of text). These categorizations were found to adequately represent the techniques used by cohort 2, despite the difference in data they had to interpret. Overall the borehole cohort used fewer techniques and specifically there was no evidence for the use of 'sticks' as in cohort 1.

Based on the findings of Bond *et al.* (2012), we initially consider the number of techniques used, as this was found to be a statistically significant factor in interpretational success. Figure 4 graphs the number of techniques used by each cohort as a percentage (number of interpreters in brackets) of those who created an interpretation based on the 'correct' tectonic concept of inversion. Dashed blue (cohort 1) and red (cohort 2) lines show the overall percentage in each cohort achieving an inversion interpretation (irrespective of the number of techniques used). The mean number of techniques used by cohort 1 was 1.7, with two techniques used by 202 (49%) of the 412 interpreters; the mean for cohort 2 was 2.9, with three techniques used by 12 (63%) of the 19 interpreters. The results for cohort 1 show that using three or more techniques improved an interpreters' chance of creating a 'correct' interpretation (Bond *et al.* 2007). For cohort 2 using three techniques improved the interpreters' chance of a 'correct' interpretation by 14%, as compared with the cohort as a whole. The data for the use of four techniques by cohort 2 is less convincing, showing a fall in the percentage of interpreters obtaining a 'correct' interpretation. The low number of interpreters in cohort 2 may be a factor (see the number of interpreters shown in parentheses on the graph).

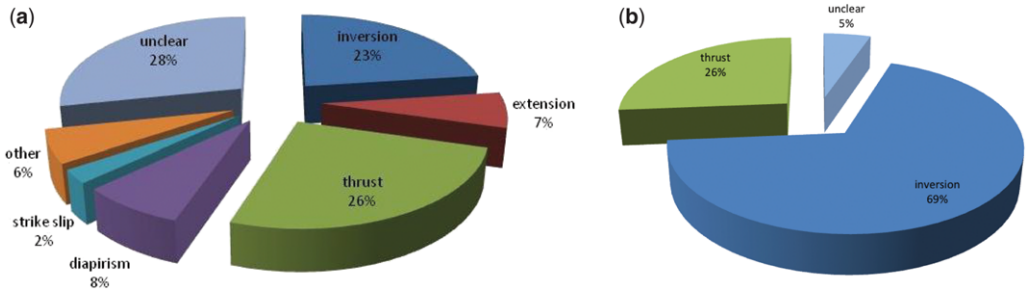

Fig. 3. The range of tectonic concepts applied to the dataset by each cohort. (**a**) Cohort 1 shows 23% with the correct inversion interpretation; the most common concept applied is thrusting (26%) (pie chart after Bond *et al.* 2007). (**b**) Cohort 2 shows a greater percentage (69%) with an inversion interpretation, followed by 26% with a thrust interpretation.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

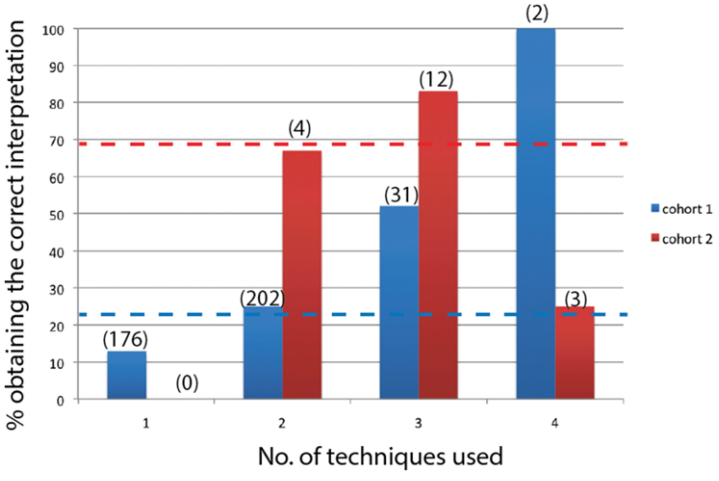

Fig. 4. Graph showing the number of techniques used by interpreters in each cohort as a percentage of those obtaining a 'correct' inversion interpretation. Numbers in parentheses give the number of interpreters. Dashed red line and dashed blue line show the percentages of cohorts 2 and 1, respectively, obtaining an inversion interpretation.

Experience. Each cohort was asked a series of questions, through a questionnaire, about their previous experience and knowledge. The findings of Bond *et al.* (2012) showed that prior experience in particular tectonic regimes or dominant tectonic regime expertise had no influence on interpretation outcome, although those who were self-proclaimed structural geology experts did better than those with a basic working knowledge (Bond *et al.* 2012). The results in Table 2A and B show that the findings of Bond *et al.* (2012), our cohort 1, are replicated for cohort 2. Those with specialist experience in both structural geology and interpretation out-performed the cohort mean (69%) in creating a 'correct' inversion interpretation, with 100% success rate in creating an inversion interpretation.

The picture for interpretational frequency is less clear (Table 2C) for cohort 2. It shows that those interpreting on a daily basis were less effective at interpretation in our experiment than those interpreting on a weekly, monthly or yearly basis. Given the results in Table 2A and B, those interpreting on a daily basis identified themselves as not having specialist structural geology or interpretation experience.

Analysis of cohort 1 (in Bond *et al.* 2007, 2012) shows similar trends with those indicating self-assessed expertise performing better than those without such experience. In Bond *et al.* (2012) the statistical significance of these findings was tested, particularly the possibility that experience was acting as a proxy for specific technique use.

Type of technique used. In our final comparison we look at the types of techniques the interpreters in cohort 2 used to interpret the borehole data and how types of technique use affect interpretational outcome. Table 3 summarizes the different techniques used and their impact on interpretational outcome. The table is divided into three sections based on cohort, these are: cohort 1; cohort 1 – specialists, a subset of cohort 1 who defined themselves as structural geologists or proficient in structural geology (this is the cohort analysed in Bond *et al.* 2012); and cohort 2.

Every interpreter in cohort 2 used both feature and horizon interpretation while completing their geological model. Those who additionally used annotations and showed evidence of thought about how the geological structure had evolved through time (geological evolution) were more likely to create a correct interpretation. Descriptive writing as an additional technique did not improve performance, although only one individual employed this technique.

As for cohort 1, the results for cohort 2 suggest that the interpretation techniques employed do have an impact on outcome. Testing of the statistical significance of technique use was undertaken in Bond *et al.* (2012) on a subset of cohort 1 – specialists. Those techniques that were statistically significant are highlighted in bold in Table 3. Notable of all the techniques was the use of 'thoughts about geological evolution' (row highlighted in Table 3). In the specialist cohort (cohort 1 – specialists), 94% of participants using this technique were

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Table 2. *Percentages and numbers of interpreters in cohort 2 who produced a 'correct' inversion interpretation, categorized on self-defined experience*

A			B			C		
Structural geology experience	No. interpreters	Percentage with 'correct' answer	Interpretation experience	No. interpreters	Percentage with 'correct' answer	Interpretational frequency	No. interpreters	Percentage with 'correct' answer
Specialist	3	100	Specialist	2	100	Daily	7	43
Good working knowledge	7	71	Good working knowledge	9	67	Weekly	3	100
Basic knowledge	9	56	Basic knowledge	8	63	Monthly	6	83
						Yearly	2	100
						Almost never	1	0

(A) Self-defined experience in structural geology from specialist to basic working knowledge; (B) self-defined interpretation experience from specialist to basic working knowledge; (C) self-selected interpretational frequency, from daily to almost never.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Table 3. *Types of interpretation technique used and their impact on successful interpretation*

Type of technique	Cohort 1 (412 participants)		Cohort 1 – specialists (184 participants)		Cohort 2 (19 participants)	
	(A) Percentage of participants (no.)	(B) Percentage of correct answer (no.)	(C) Percentage of participants (no.)	(D) Percentage of correct answer (no.)	(E) Percentage of participants (no.)	(F) Percentage of correct answer (no.)
Features	84 (344)	25 (86)	92 (170)	**37** (63)	100 (19)	69 (13)
Horizons	50 (206)	34 (70)	64 (118)	**45** (53)	100 (19)	69 (13)
Annotations	7 (30)	33 (43)	33 (61)	**51** (31)	79 (15)	80 (12)
Geological evolution	7 (28)	86 (24)	10 (18)	**94** (17)	11 (2)	100 (2)
Descriptive writing	19 (78)	28 (22)	14 (25)	44 (11)	5 (1)	0 (0)
Sticks	7 (30)	7 (2)	5 (10)	20 (1)	0 (0)	0 (0)

(A) Percentage of interpreters in cohort 1 using each technique, actual numbers in parentheses (applicable to all columns); (B) percentage using each technique (cohort 1) that resulted in a 'correct' inversion interpretation; (C) percentage of interpreters within a subset of cohort 1, those who self-defined themselves as structural geology experts (as analysed in Bond *et al.* 2012); (D) percentage of interpreters within the cohort 1 – specialist subset with the 'correct' answer – bold text denotes the techniques used that were statistically significant in achieving the 'correct' answer (see Bond *et al.* 2012); (E) percentage of interpreters in cohort 2 using each technique; and (F) percentage in cohort 2 using each technique that resulted in a 'correct' inversion interpretation. The technique 'geological evolution' is highlighted as it had the greatest impact on an interpreter's ability to produce the 'correct answer'.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

successful in creating the 'correct' modelled interpretation. The odds ratio was 40.51, meaning that those employing this technique were more than 40 times more likely to obtain the 'correct' interpretation (Bond *et al.* 2012). As highlighted in Bond *et al.* (2012), a very small percentage (10%) actually employed this technique even within this specialist subset. A similar percentage level (7%) used this technique across the whole of cohort 1, and had a similar interpretational success rate (86%). This level of success associated with technique use across the whole cohort and a combined statistical analysis of experience and techniques in Bond *et al.* (2012) suggest that experience appears to be statistically significant, but it is a proxy for good technique use.

If we consider the result for cohort 2 (Table 3, columns E & F), a similar picture appears. A low percentage of participants (11%) use the 'geological evolution' technique, but use of the technique leads to 100% success in creating the correct interpretation. Although the numbers of participants in cohort 2 are small, the results support the conclusions of Bond *et al.* (2012) in that technique use, specifically the use of geological evolution as a technique, has a clear impact on obtaining a 'correct' interpretation.

Physical expression of interpretations

The 19 interpretations of the borehole dataset are presented in (Fig. 5a–c). The stack of overlaid interpretations (Fig. 5a) highlight the main areas interpreted. Only one interpreter extended their interpretation to the furthest borehole (right-end of image), but did not interpret the borehole; rather, they extended a fault into this part of the section. Interpretations were focused around the area of structural complexity where the borehole density was greatest (to the left hand side of the section). The overlay (Fig. 5) also highlights the minimal interpretation above ground, with few interpreters extending interpretations into the air.

In general, although focusing on specific areas of the section, the interpretations appear to match well to the original model (Fig. 5b) with the high density of boreholes constraining the interpretation at the left-hand end. This observation is reflected in the image of average interpretation intensity (the number of interpretational lines spatially located in the same place on the model; Fig. 5c). As expected, the higher-intensity colours occur at the data points, where the interpretations intersect with the boreholes and topography. Higher-intensity colours are also prevalent between the boreholes where boreholes are closely spaced. To the right of the area of high borehole density the colours are grey between the boreholes (signifying zero interpretation overlap). The exception to this is the top dark blue horizon (horizon 08 – blue, see Fig. 5c), which shows intensities in the red zone, signifying between four and seven interpretations overlapping in places between the last borehole in the high borehole density area (to the left of the anticline) and the first borehole to the right of the anticline. The position of this horizon is constrained at two points on the topography between the boreholes. Few interpreters followed any of the other horizons across the whole section, as the interpreters generally did not interpret above the topographic line.

On further investigation of the average intensity image (Fig. 5c) it is apparent that not all interpreters physically marked horizons and faults at all the data points (borehole and topography intersections), as the intensity colours are at their maximum yellow and green (eight to 13 overlapping interpretations), rather than blue, the latter signifying co-location of 16 or more interpretations. In fact, individual interpretations are both sparse (Fig. 6a) and do not necessarily conform to the expected interpretation (Fig. 6b). A similar inspection of cohort 1 interpretations shows similar sparseness (Fig. 6c) and interpretational variation (Fig. 6d; see the results of interpretational outcome (tectonic concept), in the second paragraph of the Results section, Fig. 3).

Discussion

In this paper we have compared geological interpretations of laterally continuous seismic image data, published in Bond *et al.* (2007, 2012), with comparatively sparse borehole data, derived from the same geological model. The aim was to investigate the impact on both interpretational outcome and methodologies employed as a result of the different data types.

Techniques used

Technique use was found to be similar across both cohorts, with interpretation of features and horizons being prevalent, slightly more so (100% usage) for cohort 2, perhaps evidence that the cohort 1 interpreters see the seismic image as 'speaking for itself', with no need for physical interpretation. One technique that did not transport from the seismic image interpretation analysis of cohort 1 was the use of 'sticks' to interpret faults. This technique appears to be an artefact of seismic interpretation, notably the use of workstation interpretations in which 'fault stick' interpretations are built to create fault surfaces in a 3D model. Fault stick creation as a technique was shown not to be effective in creating a 'correct' interpretation for cohort 1 (Table 3).

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

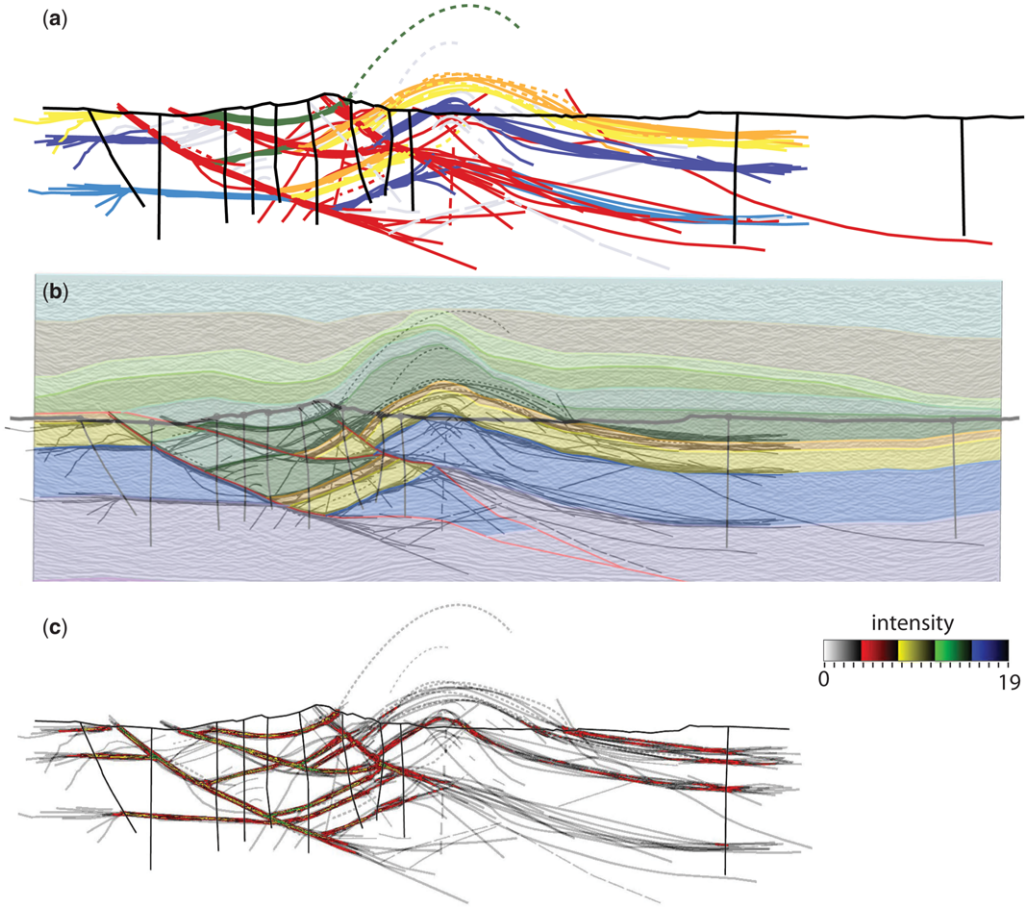

Fig. 5. (**a**) Digitization of all 19 interpretations of the borehole data overlain. Red lines are faults, horizon colour code is based on the original stratigraphy (colours assigned to lines if they crossed a borehole in the appropriate location). Pale grey lines are indistinguished interpretation line work, either faults or horizons. Dashed lines indicate that the interpreter drew dashed lines. (**b**) Transparent overlay of the original model, the combined 19 interpretations are shown with dark grey line work. (**c**) Image of the average intensity calculated from the 19 stacked interpretations. The intensity scale colour-bar is shown top right. Topography and boreholes are shown in black in all images.

The number of techniques used by cohort 2 appears to support, although not conclusively, the conclusions of Bond *et al.* (2007) for cohort 1 in which the more techniques interpreters used, the more successful they were at achieving a 'correct' interpretation. In our analysis here, of cohort 2, this hypothesis falls down above three techniques, although the number of participants (three) in the sample is small (Fig. 3). As suggested by Bond *et al.* (2012), the number of techniques used may be acting as a proxy in the analysis for the specific types of technique employed.

Technique type usage by cohort 2 follows a similar pattern to cohort 1 (see Table 3). Notable amongst the specific techniques used is evidence of thought about the evolution of the geological model created by the interpretation. The trend in use of the geological evolution technique – high success rates in creating a 'correct' interpretation, but low numbers of participants – is mirrored across the cohorts. For cohort 2, the two interpreters using this technique did not categorize themselves as specialists in structural geology or interpretation and interpreted at a frequency of weekly and monthly respectively. This suggests that use of the technique 'thought about the geological evolution' is effective irrespective of specialism or experience. The low numbers of interpreters showing evidence of use of this technique in both cohorts does not necessarily mean that others did not utilize this technique mentally, just that they showed no physical marking to evidence their thoughts.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

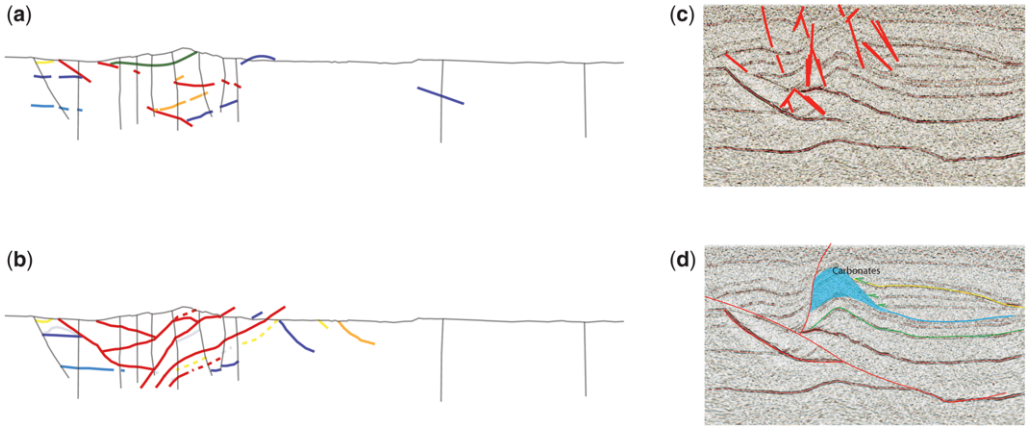

Fig. 6. Examples of individual interpretations from both cohorts. (**a**) Individual interpretation from cohort 2; note the sparse interpretation, the lack of linkage between boreholes and the lack of interpretation of many of the boreholes, even in the area of high borehole density. (**b**) Individual interpretation from cohort 2; note the choice in linkage of the faults between boreholes (antithetic to the original model). (**c**) Individual interpretation from cohort 1; the interpretation is both sparse and unfocused (interpretation from Bond *et al.* 2007). (**d**) Individual interpretation from cohort 1; the interpreter has marked some of the faults and interpreted a carbonate reef on the antiformal structure.

Initial data

The difference in 'white space' between the two initial datasets (Fig. 2a, b) is distinct. Cohort 1, interpreting the seismic image data, were presented with an A4 sheet 'covered' in printed data. The bounds of the interpretation were clear and there were data to interpret at all points in the 2D panel, even if it had limited resolution and clarity. In contrast cohort 2 started their interpretation with a white page, containing limited data along lines (the topography and boreholes). Interpretations with limited physical marking were produced by both cohorts (Fig. 6a, c).

Psychological studies of interpretation focus on the theories of occlusion and illusionary boundary completion (e.g. Shipley & Kellman 2003; and Kalar *et al.* 2010). Simply, occlusion occurs when only part of an object is seen (for whatever reason) and an illusionary boundary is created to complete the partially occluded object. The human brain is very effective at this type of activity, with evidence of perception of partially occluded objects by 4 month-old infants (Kellman & Spelke 1983). In the interpretation of the geological data in this experiment, both cohorts had to employ illusionary boundary creation, as the data were incomplete (or occluded), to create a final model. The ability of individual participants to complete the model in each cohort will be affected by the differences in the data presented for interpretation.

Psychological occlusion experiments often use 'clean images', similar to but with more information available than the borehole dataset given to cohort 2; this is in distinct contrast to the 'noisy' seismic image data faced by cohort 1. Studies of interpretation of more noisy or blurred data investigate the concept of scene gist (e.g. Oliva & Torralba 2006), in which participants are able to quickly determine an everyday scene, even when the images they are presented with are unclear. Evidence from medical science, specifically the interpretation of X-ray images, suggests that those with more experience are able to recognize objects more effectively in noisy data (e.g. Lesgold *et al.* 1988); simply, they know what they are looking for. Indeed, studies of imaging and image interpretation in medical science provide the best analogue for interpretation of noisy geological data such as seismic imagery.

For cohort 1, noise in the data (the complexity of the image) will limit the recognition of critical features and hence the interpreter's ability to see the geological structure (create an illusionary boundary). In radiology, random and structured noise in images has been shown to affect visual searches for nodules (Revesz *et al.* 1974; Kundel & Revesz 1976; Kundel *et al.* 1985). In summary there is a relationship between the conspicuity of the object of interest and the background complexity of the image. The research shows that an increase in structured noise decreases the probability of nodule detection.

For cohort 2, the challenge is in constructing the illusionary boundary from the limited information available, assuming that the interpreters have enough knowledge of what the geological structure (or object) should look like. The complexity of the object as well as the amount of data will affect the

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

outcome. A further element may arise from the psychological constraints of filling in the 'white space' that confronts the interpreter. The PhD thesis of Joyce (2009) explores constraint on creativity when faced with a blank page. Joyce concludes that a moderate amount of constraint helps generate creativity.

Unlike cohort 1, who could choose to start their interpretation anywhere on the seismic image, the starting points for interpretation by cohort 2 were more constrained, limited by the data. Cohort 2 were effectively forced to interpolate between the data lines, starting on one data line and working to another. The outcome of amalgamating all the interpretations to create an average intensity of interpretations for cohort 2 (Fig. 5c) suggests that there was enough data, particularly in the area of structural complexity, to constrain the interpretation. That is despite individuals showing an inability to commit to a fully outlined interpretation (Fig. 6a). The question that arises is whether the creative constraint of linking data points across a blank page or an ability to create an illusionary boundary (i.e. visualize the geological structure) controls the interpretation. In a similar manner interpreters of the seismic image (cohort 1) may have seen the image as 'speaking for itself', therefore requiring limited physical marking to complete the interpretation exercise. What is apparent from both interpretation datasets is that 'full' interpretations, those that resulted in physical marking across the majority of the interpretation space, were rare.

The ability to see the geological structure and hence create an illusionary boundary would seem to be a key factor in interpretation ability. This ability to visualize relies on a level of expertise or at least exposure to the types of geological structures that may be seen. However, in contrast to the medical examples and the seismic image data provided to cohort 1, the data presented to cohort 2 had a limited visual expression. Perhaps this limited borehole data required the interpreters to question from the start the validity of correlating two data points in terms of their final geological model. That is, they do not visualize immediately the geological structure, but iterate to a solution, whereas seismic interpreters can follow an amplitude across the dataset without a requirement to question their interpretation, or to have a visual geological model in their heads that they are working with. In essence we suggest that interpreters of the seismic image may create non-linked illusionary boundaries, that is, they visually link elements of the object, but do not necessarily see the whole object, until a number of non-linked boundaries result in the final or near-final object. For cohort 2, visualization of non-linked boundaries, object segmentation, may also occur, but as there is little guide to linkage of these elements, each boundary link requires scrutiny and testing as the object (geological structure) is created.

What we recognize is that, although boundary illusion and scene gist are relevant to elements of interpretation, geological interpretation is more than just object recognition, it requires geological reasoning (see Frodeman 1995). The results, although for low numbers of participants, suggest that the borehole interpreters (cohort 2) were much more successful in creating interpretations, with 69% of the cohort achieving the 'correct' answer, than the seismic image cohort (21%). Bond *et al.* (2012) showed that interpreters who used thoughts about geological evolution increased their odds, by 40 times, of creating a correct interpretation. Thus, an interpretation situation that requires the interpreter to think about a complete model and its evolution during the interpretation (assesses the impact of joining data points) should have a positive effect on interpretation outcome. The object (geological structure) needs to be seen in its static final state, but importantly needs to be critically assessed to see if the final geometry can be created geologically through evolution of the geometries. Based on our observations, in situations where both seismic image and borehole data are available, giving interpreters only borehole data initially may create the white space required to force interpreters to question the validity of their interpretation as they undertake it. The seismic image data can then be introduced later to refine and test models.

Why do the borehole cohort perform better?

Overall the borehole cohort (2) outperformed the seismic interpretation cohort (1). This poses a question on the representativeness of the sample of individuals within cohort 2; were they more appropriately experienced or skilled? Cohort 2 self-selected to attend a structural geology workshop; however, their personal assessment of experience does not suggest a significant difference from cohort 1. Cohort 2's technique use was also similar to that of cohort 1 and the techniques identified in the earlier analysis of cohort 1 that were effective were seen to be effective for cohort 2, and with similar percentages employing these techniques. So why are they more successful?

One hypothesis is that the working backgrounds and experience of the interpreters in cohort 2 as mineral and mine-based geologists mean that they often work within more complex fold-and-thrust belt-style terrains with complex structural histories (Table 1; 26% chose their dominant career tectonic setting as polyphase), when compared with the oil- and gas-based geologists comprising cohort 1.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Essentially their prior experience and knowledge could have had an impact on interpretational outcome, through more exposure to the type of geological structure the exercise was based on and hence a greater ability to recall and visualize the structure. Those who defined themselves as specialists in structural geology and interpretation in cohort 2 were more likely to create a 'correct' interpretation than those that did not; this is also the case with cohort 1. However the work of Bond *et al.* (2012) suggests that experience acts as a proxy in the statistical analysis for effective technique use, namely interpretation query through consideration of the geological evolution. Interestingly these structural specialists in cohort 2 were not the same as those interpreting on a frequent basis. This suggests that junior staff interpret and that there is a need for careful quality control of interpretations in the mining and minerals sector given this structure.

A second hypothesis is that the use of the borehole data given to cohort 2 provides an advantage, through an initial stratigraphy and some clearly tied units. However a similar stratigraphy could have been made using the edges of the seismic image by cohort 1; in fact, those who clearly correlated horizons in cohort 1 outperformed those who did not (Table 3). Tying horizons across the seismic image or between the boreholes at either end of the section allows the interpreter to establish the concept of regional, a critical concept in understanding tectonics, and particularly useful for identifying inversion (Williams *et al.* 1989). Having highlighted the importance of establishing a regional by correlating horizons across the full section, it is perhaps surprising that none of the interpreters in cohort 2 interpreted the borehole on the far right of the section, which aids confirmation of the regional for the stratigraphy. The implication is that, if the concept of regional was used by cohort 2, it was not used explicitly, but as a subconscious control.

Finally, is there an impact of 'white space'? Does the borehole dataset provide white space for creativity to ensure critical thinking, model generation and evolutionary thoughts? The borehole cohort had almost no option but to interpret horizons between boreholes, suggesting that they were forced to create a model that fitted the data and potentially they tested this (mentally) as they progressed.

Conclusions

In our comparison of two cohorts' interpretations of distinctly different datasets (seismic image and borehole) created from the same initial geological model, we have shown that, for both cohorts:

(1) specific technique use increased the chance of creating a 'correct' interpretation, specifically if interpreters 'thought about the geological evolution';
(2) the percentage of interpreters who 'thought about the geological evolution' was low across both cohorts, despite its success if employed;
(3) those who self-defined themselves as specialists did better than those who did not.

For cohort 2:

(1) A greater percentage of the interpreters of the borehole data produced the 'correct' modelled interpretation.
(2) All interpreters in cohort 2 interpreted both horizons and features, whilst the seismic cohort did not. This is likely to have resulted in an increased overall performance of cohort 2, given the importance of the concept of regional and the statistical analysis of Bond *et al.* (2012) that suggests that horizon interpretation has a positive effect on interpretational outcome. However, this does not fully account for their comparative success in creating 'correct' interpretations.
(3) Specialism did not equate to frequency of interpretation.
(4) Interpreters did not use the technique of drawing fault sticks.

From our analysis of cohort 2, and comparison with cohort 1, it would seem that the type of data interpreted has an impact on interpretational outcome (more correct interpretations by cohort 2). The borehole data appear to force the interpreter into interpreting both horizons and features and we hypothesize in questioning their model as they interpret. Both of these factors have previously been shown to improve interpretational outcome (Bond *et al.* 2012). Prior knowledge and experience may also have played a role, with the professional experience of cohort 2 focused on terrains with a complex structural history.

Irrespective of the specific reasons behind the overall performances of the two cohorts, we interpret the data presented to support the conclusions of Bond *et al.* (2012), who identified specific technique use as having an impact on interpretational outcome, specifically the use of thoughts about the geological evolution of the interpreted model. We suggest that techniques involving the testing of the validity of interpretations through analysis of geological evolution during interpretation be employed to maximize effective interpretation, irrespective of the data type interpreted. Forcing the interpreter to engage in a geological reasoning process by creating a situation in which the amount of data v. creative white space is optimized may be a simple method to ensure consideration of the validity of the interpretation as it is created.

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Given our results, in situations where multiple data sources are available, the use of early borehole data interpretations prior to seismic image interpretation may encourage use of effective techniques. These initial models can subsequently be checked and tested against seismic image data. The constraints of software tools are not explored here, but will clearly impact on creativity and creative thinking and constrain interpretation practices. It is worth noting that interpreters of the borehole dataset (cohort 2) did not draw fault sticks, a common technique for annotating faults in seismic interpretation software packages, and that analysis of fault stick use in cohort 1 had a negative correlation with successful interpretation. Finally, as interpretations in the mining and minerals sector are not undertaken by those with the most experience (as evidenced by cohort 2), training in effective technique use and creating interpretation situations that maximize creativity and geological reasoning could be the key to effective geological interpretation.

The authors would like to thank participants in the interpretation exercises. Bond was funded by Midland Valley Exploration and a Scottish Executive SCORE award whilst collecting the cohort 1 interpretation dataset. Midland Valley Exploration's Move software was used to create the geological model and borehole dataset. Fiji (Image J) software was used to create Figure 5a and c. ION GXT are thanked for creation of the synthetic seismic image dataset.

References

Baddeley, M. C., Curtis, A. & Wood, R. 2004. An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cognitive bias and herding. *In*: Curtis, A. & Wood, R. (eds) *Geological Prior Information: Informing Science and Engineering*. Geological Society, London, Special Publications, **239**, 15–27, http://dx.doi.org/10.1144/GSL.SP.2004.239.01.02

Bond, C. E., Gibbs, A. D., Shipton, Z. K. & Jones, S. 2007. What do you think this is?: 'conceptual uncertainty' in geoscience interpretation. *GSA Today*, **17**, 4–10, http://dx.doi.org/10.1130/GSAT01711A.1

Bond, C. E., Shipton, Z. K., Gibbs, A. D. & Jones, S. 2008. Structural models: optimizing risk analysis by understanding conceptual uncertainty. *First Break*, **26**, 65–71.

Bond, C. E., Lunn, R. J., Shipton, Z. K. & Lunn, A. D. 2012. What makes an expert effective at interpreting seismic images? *Geology*, **40**, 75–78, http://dx.doi.org/10.1130/G32375.1

Frodeman, R. 1995. Geological reasoning: geology as an interpretive and historical science. *Geological Society of America Bulletin*, **107**, 960–968.

Joyce, C. K. 2009. *The Blank Page: Effects of Constraint on Creativity*. PhD thesis, University of California, Berkeley, CA.

Kalar, D. J., Garrigan, P., Wickens, T. D., Hilger, J. D. & Kellman, P. J. 2010. A unified model of illusory and occluded contour interpolation. *Vision Research*, **50**, 284–299, http://doi.org/10.1016/j.visres.2009.10.011

Kellman, P. J. & Spelke, E. S. 1983. Perception of partly occluded objects in infancy. *Cognitive Psychology*, **15**, 483–524.

Kundel, H. L. & Revesz, G. 1976. Lesion conspicuity, structure noise, and film reader error. *American Journal of Roentgenology*, **126**, 1233–1238.

Kundel, H. L., Nodine, C. F., Thickman, D., Carmody, D. & Toto, L. 1985. Nodule detection with and without a chest image. *Investigative Radiology*, **20**, 94–99.

Lark, R. M., Mathers, S. J., Thorpe, S., Arkley, S. L. B., Morgan, D. J. & Lawrence, D. J. D. 2013. A statistical assessment of the uncertainty in a 3-D geological framework model. *Proceedings of the Geologists' Association*, **124**, 946–958, http://dx.doi.org/10.1016/j.pgeola.2013.01.005

Lesgold, A. M., Rubinson, H., Feltocih, P. J., Glaser, R., Klopfer, D. & Wang, Y. 1988. Expertise in a complex skill: diagnosing X-ray pictures. *In*: Chi, M. T. H., Glaser, R. & Farr, M. (eds) *The Nature of Expertise*. Erlbaum, Hillsdale. NJ, 311–342.

Lindsay, M. D., Ailleres, L., Jessell, M. W., de Kemp, E. & Betts, P. G. 2012. Locating and quantifying geological uncertainty in three-dimensional models: analysis of the Gippsland Basin, southeastern Australia. *Tectonophysics*, **546–547**, 10–27.

Lindsay, M. D., Jessell, M. W., Ailleres, L., Perrouty, S., de Kemp, E. & Betts, P. G. 2013. Geodiversity: exploration of 3D geological model space. *Tectonophysics*, **594**, 27–37.

Oliva, A. & Torralba, A. 2006. Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research* **155**, 23–36.

Polson, D. & Curtis, A. 2010. Dynamics of uncertainty in geological interpretation. *Journal of the Geological Society, London*, **167**, 5–10, http://dx.doi.org/10.1144/0016-76492009-055

Rankey, E. C. & Mitchell, J. C. 2003. That's why it's called interpretation: impact of horizon uncertainty on seismic attribute analysis. *The Leading Edge*, **22**, 820–828, http://dx.doi.org/10.1190/1.1614152

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. & van der Keur, P. 2006. A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, **29**, 1586–1597.

Revesz, G., Kundel, H. L. & Graber, M. A. 1974. The influence of structured noise on the detection of radiologic abnormalities. *Investigative Radiology*, **9**, 479–486.

Shipley, T. F. & Kellman, P. J. 2003. Boundary completion in illusory contours: Interpolation or extrapolation? *Perception*, **32**, 985–999.

Tacher, L., Pomian-Srzednicki, I. & Parriaux, A. 2006. Geological uncertainties associated with 3-D subsurface Models. *Computers & Geosciences*, **32**, 212–221.

Torvela, T. M. & Bond, C. E. 2011. Do experts use idealised structural models?: insights from a deepwater fold-thrust belt. *Journal of Structural*

Downloaded from http://sp.lyellcollection.org/ by guest on January 24, 2019

Geology, **33**, 51–58, http://dx.doi.org/10.1016/j.jsg.2010. 10.002

Wellmann, J. F. & Regenauer-Lieb, K. 2012. Uncertainties have a meaning: information entropy as a quality measure for 3-D geological models. *Tectonophysics*, **526–529**, 207–216.

Wellmann, J. F., Horowitz, F. G., Schill, E. & Regenauer-Lieb, K. 2010. Towards incorporating uncertainty of structural data in 3D geological inversion. *Tectonophysics*, **490**, 141–151.

Williams, G. D., Powell, C. M. & Cooper, M. A. 1989. Geometry and kinematics of inversion tectonics. *In*: Cooper, M. A. & Williams, G. D. (eds) *Inversion Tectonics*. Geological Society, London, Special Publications, **44**, 3–16, http://dx.doi.org/10.1144/GSL.SP.1989.044.01.02

Ye, M., Pohlmann, K. F. & Chapman, J. B. 2008. Expert elicitation of recharge model probabilities for the Death Valley regional flow system. *Journal of Hydrology*, **354**, 102–115.