Journal of the Geological Society

Dynamics of uncertainty in geological interpretation

Debbie Polson and Andrew Curtis

Journal of the Geological Society 2010; v. 167; p. 5-10
doi:10.1144/0016-76492009-055

Journal of the Geological Society, *London*, Vol. **167**, 2010, pp. 5–10. doi: 10.1144/0016-76492009-055.

Dynamics of uncertainty in geological interpretation

DEBBIE POLSON[1,2]* & ANDREW CURTIS[1,2]
[1] *University of Edinburgh, School of GeoSciences, Edinburgh EH9 2JT, UK*
[2] *ECOSSE (Edinburgh Collaborative of Subsurface Science and Engineering)*
**Corresponding author (e-mail: dpolson@staffmail.ed.ac.uk)*

Abstract: Interpretation of geological data is based on both personal judgement and previous experience of related scenarios. In combining such information geologists employ heuristics (rules of thumb), and are therefore subject to biases that are well known in cognitive psychology and are common to all expert judgements. Here we analyse dynamic uncertainty in an evolving geological interpretation. Through a well-designed elicitation process we show how the inclusion of multiple experts influences interpretational bias. In particular, group convergence of opinion is observed, and we show how this can be differentiated from 'herding' behaviour similar to that observed in economic bubbles by forcing a consensus to be reached. Thus we can identify when and why the judgement of a single geological expert should be treated with caution. This process can be applied to any geological interpretational scenario.

Geologists are often required to make judgements and interpretations in situations of uncertainty where data are inadequate to fully constrain any particular interpretation. Any interpretation by an expert is then dependent on the prior knowledge and experience of that expert, and hence the result is both subjective and qualitative in nature. The geological prior information employed is difficult to assess or quantify, as experience and knowledge vary from expert to expert, as do the methods an expert may use to generalize and categorize information (Rankey & Mitchell 2003; Curtis & Wood 2004; Wood & Curtis 2004; Bond *et al*. 2007).

This prior information may be thought of as a prior probability that an expert places on each hypothesis. As new information (e.g. data) becomes available this initial probability is updated by combining it with the data to produce a new, 'posterior' probability of each hypothesis. However, this approach assumes that the way in which an expert forms and updates their beliefs follows some rational, ordered approach. Research has shown that all experts are subject to biases when making probabilistic assessments, which result in inaccurate and uncertain judgements (e.g. Baddeley *et al*. 2004; O'Hagan *et al*. 2006). For example, numerous studies have shown that all individuals find it difficult to assign numerical probabilities to judgements (e.g. Kahneman *et al*. 1982; Anderson 1998). This is because heuristics (rules of thumb) are used to assess probability, and these introduce bias. Group interaction has the advantage of allowing knowledge and experience to be shared amongst the experts. It provides a method of aggregating individual opinions (e.g. Phillips 1999) and evidence suggests that group interaction can reduce some effects of individual bias (Sniezek 1992). However, this can also introduce other group biases (Sniezek 1992).

This paper investigates uncertainty in geological interpretation by individual and multiple experts. It describes an elicitation process designed to demonstrate individual bias, and the effect of group dynamics both on the final interpretations and on the perceptions of individual experts. Expert elicitation theory and practice have been investigated (e.g. see Bonano & Apostolakis 1991) and used in numerous studies, including in the Earth and environmental sciences (Morgan *et al*. 2001; Curtis & Wood 2004; Arnell *et al*. 2005; Bond *et al*. 2007; Lowe & Lorenzoni 2007; Ye *et al*. 2008). However, a well-structured and well-managed elicitation process is essential to avoid group biases such as overconfidence, to prevent the group becoming dominated by opinion over knowledge and to ensure that the expertise of all individuals is recognized with no single individual dominating the group by force of personality (O'Hagan *et al*. 2006).

The results show that individual judgements can be contradictory, and that group interaction radically alters individual perceptions. They demonstrate that a group consensus may not reflect the opinions of all constituent and consenting experts, and overall the results show that any probabilistic assessment, whether from individuals or a group, should be elicited with a carefully structured process such as the one we propose here if the results are to be properly valued and understood.

Context

A set of 2D seismic lines and contextual geological data for the Firth of Forth has been interpreted by experts on at least three occasions to produce 3D subsurface geological models. The Firth of Forth is in the Midland Valley of Scotland. The Midland Valley is an 80 km wide and >150 km long NE–SW-trending graben structure consisting of a series of sedimentary basins formed between 390 and 280 Ma ago in Devonian, Carboniferous and Permian times. These form a vertical succession of 5 km of mainly fluvio-lacustrine and marginal marine sedimentary rocks (Ritchie *et al*. 2003). The rock types are a variety of intercalated mudstones, siltstones, sandstones, coals, limestones, and extrusive and intrusive igneous rocks. There are a series of NNE–SSW-trending anticlinal and synclinal folds with numerous large faults.

Geological data

The seismic dataset consisted of 33 2D seismic reflection profiles in the Firth of Forth in Scotland (see Fig. 1 for an example). The lines are arranged in an approximate 1 km × 1 km grid and have a combined length of approximately 600 km. A number of commercial wells and boreholes were used to calibrate the data.

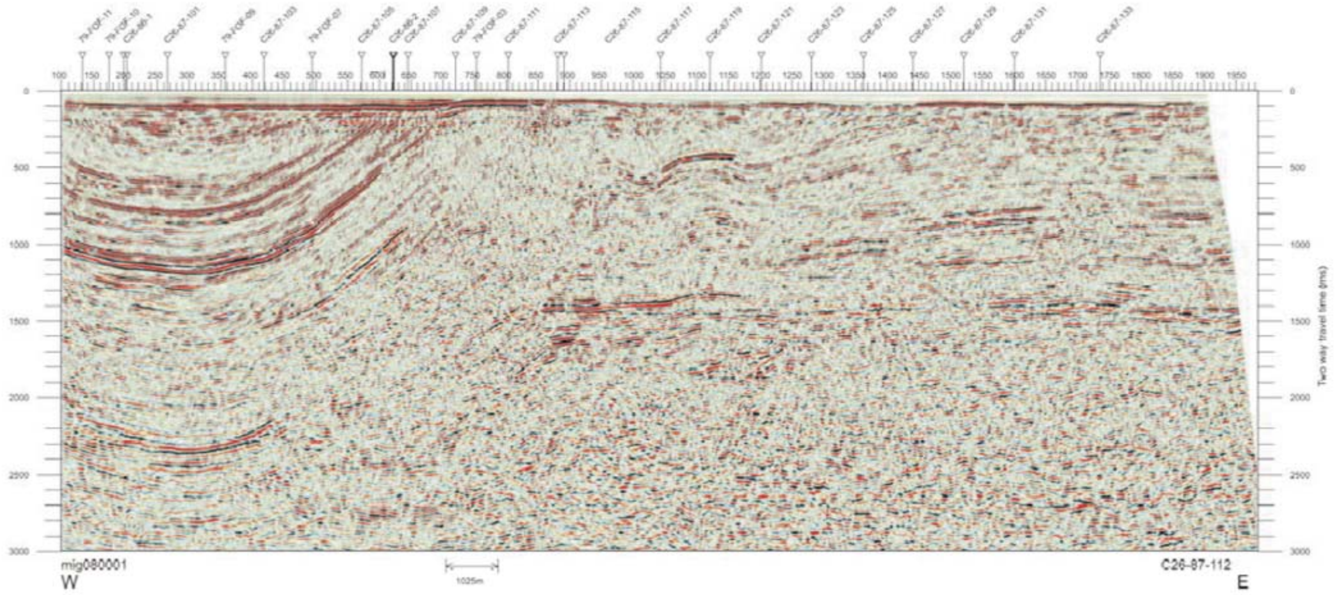

Fig. 1. Example of Firth of Forth seismic data. Line C26_87_112 oriented SW–NE through the 25/26-1 tied well. BGS©NERC (IPR/114-29DR), seismic data shown with permission of Phoenix Data Solutions.

These include the 25/26-1 well drilled by Conoco in 1990, which reached a depth of 2009.5 m, and shallow stratigraphic boreholes drilled by the Institute of Geological Studies (now the British Geological Survey) and the National Coal Board (Thomson 1978). Structural contour information from mine-working operations, published refraction work and evidence from field-based studies also contributed to the dataset. A full description has been given by Ritchie *et al*. (2003) and Underhill *et al*. (2008).

Interpretation of the geological data

From the geological models of the region based on these data, three features were identified for the elicitation exercise: a large fault, a sandstone formation and an overlying mudstone formation.

One interpretation of the data includes a fault running north–south, referred to as the Mid-Forth Fault (Ritchie *et al*. 2003), which is absent from later interpretations (e.g. Underhill *et al*. 2008).

A particular formation, the Knox Pulpit sandstone formation, is known to exist to the north of the Firth of Forth from borehole and outcrop data. Palaeogeographical reconstructions suggest that this formation is likely to be present at depth within the Firth of Forth (Browne *et al*. 1987), but this remains unproven.

Above the Knox Pulpit Formation is the mainly siltstone and mudstone Ballagan Formation, which is know to exist both north and south of the Firth of Forth (Mitchell & Mykura 1962) but is also unproven in the Firth of Forth.

Method

An elicitation exercise was carried out with the four experts to investigate the current level of uncertainty in individual expert interpretations of this dataset, and the effects of group interaction.

The experts were asked to assess the probability that three structures in the Earth model existed in a subregion beneath the Firth of Forth: the Knox Pulpit sandstone formation (a potential reservoir), the Ballagan Formation above the reservoir, and the Mid-Forth Fault, hereafter referred to as reservoir, seal and fault respectively. They also assigned lower- and upper-bound probabilities to their best guess, allowing each expert to make an assessment of their own uncertainty.

A six-step elicitation exercise took place: (1) information was elicited from each expert independently before the group session; (2) at the start of the group session they were alerted to the various common biases in expert judgements; (3) the experts repeated the assessment individually; (4) these individual assessments were shared amongst the group with each expert being asked to explain their reasoning; (5) this led to group discussion through which the experts were prompted to reach a consensus assessment; (6) finally the experts again repeated assessments individually.

Results

Figures 2–4 show the probability distribution (p.d.; from linear interpolation of lower bound, best guess and upper bound) for each structure. These are normalized distributions that give the approximate probability of percentage, p, lying within a particular range by integrating the probability density function, $f(p)$, over the given range. Expert E4 was not available prior to the session and so has no pre-session results. During the introduction stage of the group session it became clear that expert E1 had assessed the wrong fault. This resulted in a radically different distribution at session start compared with that pre-session (compare Fig. 4a and b). During discussions regarding the existence of a seal, it was revealed that this expert had also incorrectly assumed that instead of assessing the probability that the seal exists they should account for the quality of the seal. Figure 2b and c shows the change in perception of expert E1 from the start to the end of the session, with the assessment at the end showing a significant change.

Figure 2a shows pre-session distributions for the probability of

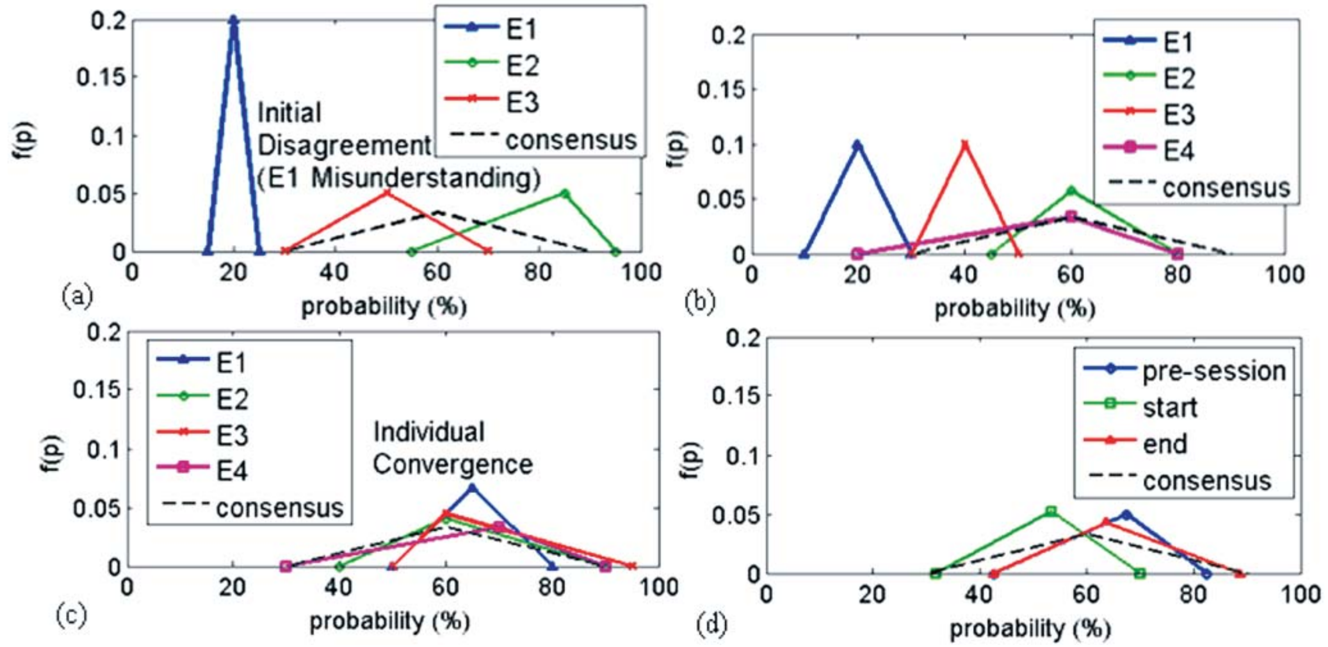

Fig. 2. Probability density function for the existence of the seal from individual (continuous line) and consensus (dashed line) assessment of the lower, upper and best-guess probabilities labelled with the interpretation of the results. The probability of the percentage, *p*, falling within a given range is given by the integral of *f*(*p*) over the range. (**a**) Pre-session assessments showing initial disagreement amongst experts and a misunderstanding for expert E1; (**b**) start of session assessments; (**c**) end of session assessments showing the individual experts converging towards to group consensus; (**d**) average. E1–E4 represent the four experts.

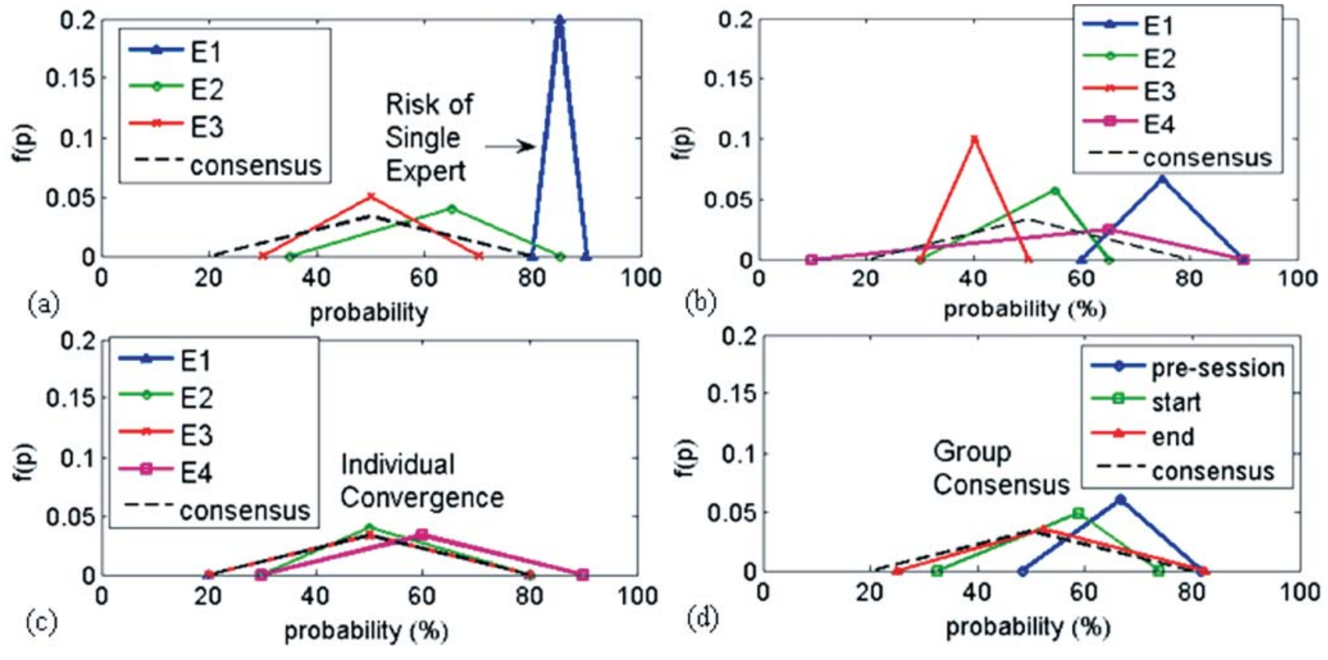

Fig. 3. Probability density function for the existence of the reservoir from individual (continuous line) and consensus (dashed line) assessment of the lower, upper and best-guess probabilities labelled with the interpretation of the results. The probability of the percentage, *p*, falling within a given range is given by the integral of *f*(*p*) over the range. (**a**) Pre-session assessments showing the potential risk of eliciting the judgement of only one expert, which may differ radically from the judgements made by other experts; (**b**) start of session assessments; (**c**) end of session assessments showing the individual experts converging towards the group consensus; (**d**) average. E1–E4 represent the four experts, showing the group convergence through the elicitation process.

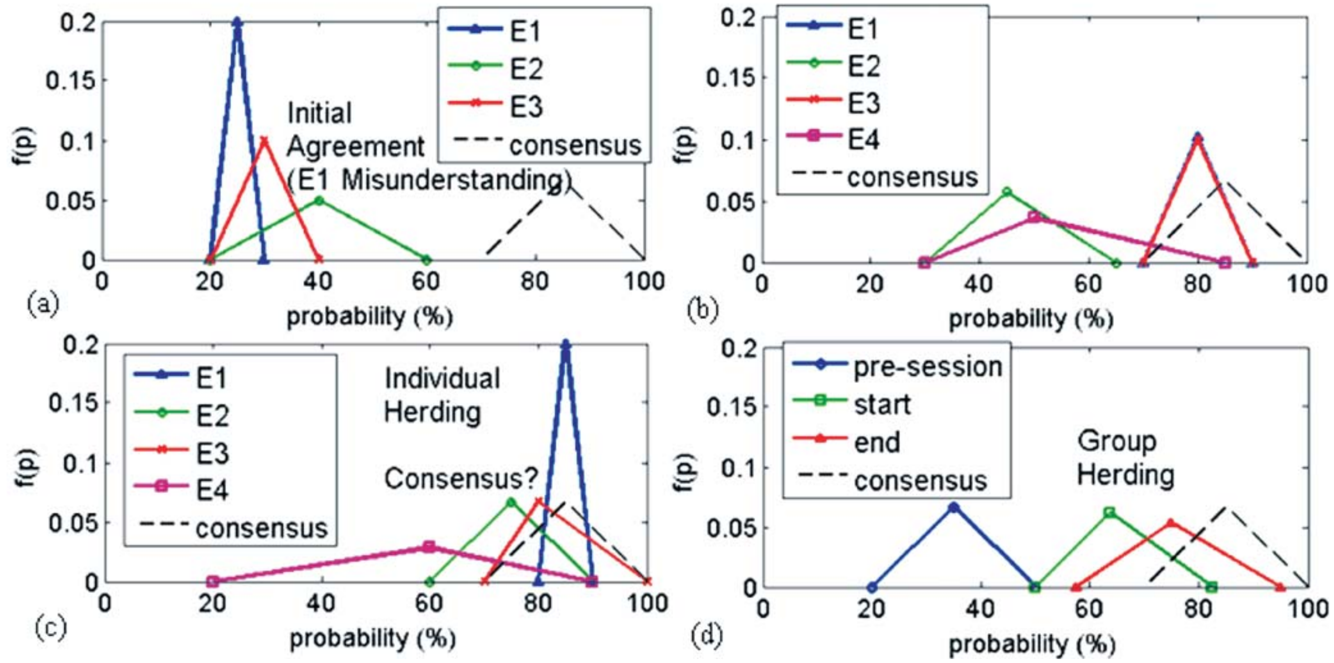

Fig. 4. Probability density function for the existence of the fault from the individual (continuous line) and consensus (dashed line) assessment of the lower, upper and best-guess probabilities labelled with the interpretation of the results. The probability of the percentage, p, falling within a given range is given by the integral of $f(p)$ over the range. (**a**) Pre-session assessments showing initial agreement between experts and one misunderstanding from expert E1; (**b**) start of session assessments; (**c**) end of session assessments showing the individual experts herding around expert E1; (**d**) average. E1–E4 represent the four experts, showing group herding through the elicitation process.

the seal. Expert E1 misunderstood what was being assessed, and experts E2 and E3 reached different conclusions from each other about the probability of the seal's existence, with their distributions overlapping only at their upper and lower bounds, respectively.

Figure 3a shows pre-session results for the existence of the reservoir. Whereas the distributions of experts E2 and E3 mostly overlap, expert E1 shows significant disagreement. Expert E1 is more certain of the reservoir's existence (higher probability) and is more confident of this assessment (narrow p.d.).

Figure 4a shows pre-session results for the existence of the fault. Expert E1 has misunderstood which fault they are assessing, and opinions of experts E2 and E3 overlap. However, at the start of the session shown in Figure 4b, expert E3 has radically changed their assessment, resulting in distributions from experts E2 and E3 that are mutually exclusive. This shows that individual experts can make interpretations based on the same set of data that are contradictory with others and with themselves, even taking into account the expert's estimation of their own uncertainty.

Individual assessments at the end of the elicitation session (Figs 2c, 3c and 4c) show a clear shift towards the consensus distribution compared with pre-session and start of session assessments. The experts have individually converged towards a group position during the elicitation process. Expert E1 in particular shows a dramatic change with respect to the reservoir, for which there was no misunderstanding of the task: the distribution in Figure 3a shows a high degree of confidence in the existence of the reservoir with low uncertainty before the session. By the end of the session, this expert has radically changed their opinion with a final assessment, shown in Figure 3c, that excludes their initial assessment entirely.

Results from parts a, b and c of Figures 2–4 individually were combined by averaging the lower bound, best guess and upper bound. Where there was a misunderstanding, results were excluded. The average individual assessments as shown in Figures 2d, 3d and 4d for the seal, reservoir and fault, respectively, show a clear convergence towards a consensus during the elicitation process. Although the average distribution from the end of the session differs from the consensus, it shows a significant shift towards the consensus for all three structures.

Combining the distributions for the seal, reservoir and fault gives the joint probability distribution for the existence of all three features at once (if the fault was impermeable); this might represent the p.d. of the existence of a viable bounded reservoir. Figure 5 shows the distributions for the seal, reservoir and fault combined into a single probability distribution for the individual pre-group elicitation, as well as the average of the pre-group elicitation, start of elicitation, end of elicitation and combined consensus distributions reached by the experts during the session. Figure 5a shows the simplified method of averaging the experts' distributions by averaging the lower bounds, best guesses and upper bounds elicited from the individual experts. Figure 5b shows the combined distributions produced by averaging the individual expert distributions for the seal, reservoir and fault. This method is more robust, as it shows the full width of all four experts' individual distributions. Figure 5c shows the combined individual expert pre-session distributions.

Figure 5a and b both show the joint probability increasing as the elicitation progressed, with the peak of the end of session distribution greater than that of the pre-session and start of session distributions. The distribution from the end of the session is also wider than the earlier distributions, as the experts become less confident through the elicitation process. In both cases the

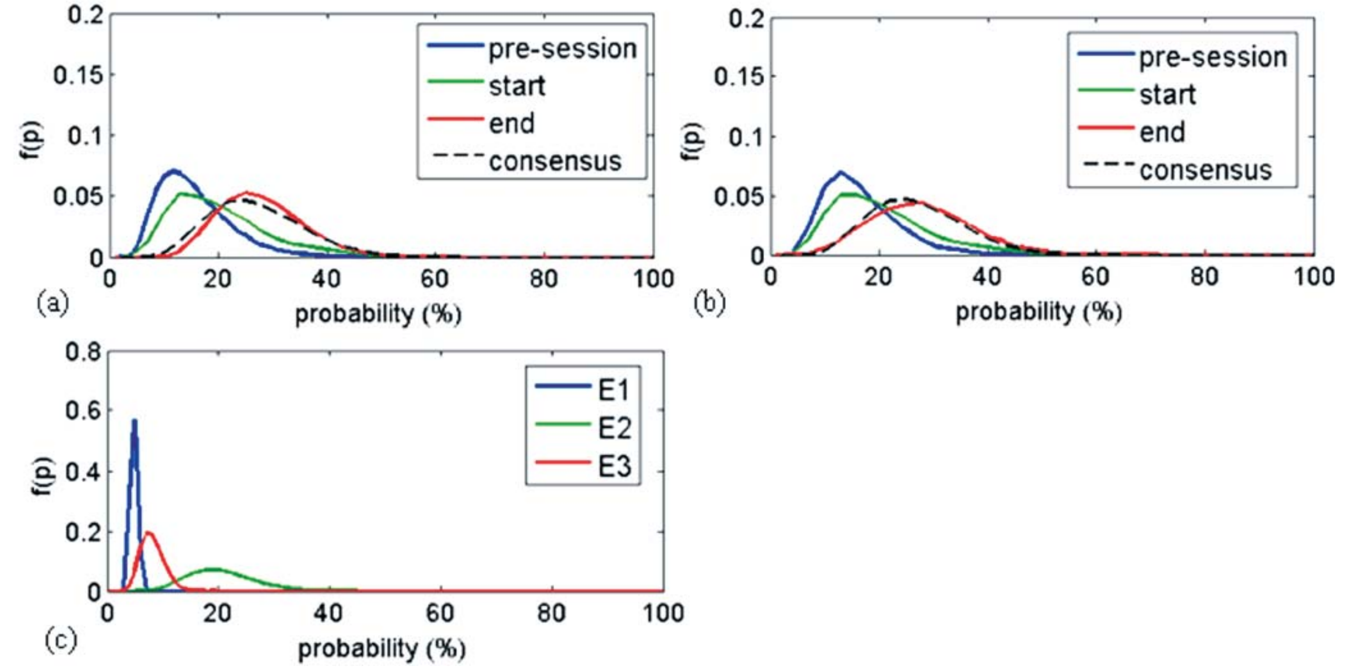

Fig. 5. Combined probability distributions for the reservoir, seal and fault. The probability of the percentage, *p*, falling within a given range is given by the integral of *f*(*p*) over the range. (**a**) Average lower, upper and best-guess probabilities; (**b**) average distributions; (**c**) individual expert distributions for experts E1, E2 and E3 for the pre-session assessment.

end of session distribution is similar to the consensus. From the peak of the consensus distribution, the most likely probability that all three structures exist is *c*. 25%, with an upper limit of ~60%. Figure 5c shows the variation between experts before the elicitation session, with experts E1 and E2 producing distributions that are mutually exclusive.

Discussion

The results of the elicitation demonstrate clearly for the first time the dynamics of uncertainty in the interpretation of geological data by teams of multiple experts. The distributions of individual experts at pre-session and at start of session highlight the potential for experts to reach contradictory conclusions from the same data, even given their own estimation for their potential for error (e.g. experts E1 and E3 in Figs 2a and 3a). This reflects the subjective nature of the prior knowledge and experience of experts, which influences their interpretation.

In real life lone experts will regularly be required to make judgements in situations of uncertainty, whereas having more than three experts involved in such an assessment is rare. However, the results of this study highlight the risk associated with using single experts, becuse of the potential for disagreement as highlighted in Figure 3a. Had only expert E1 been asked to assess the probability of the existence of the reservoir, the result would be a confident assessment of a high probability of existence. However, the other two experts elicited show far less certainty about the reservoir's existence and are far less confident in their answers, as is expert E1 by the end of the process (Fig. 3c).

Of particular concern is the potential for misunderstandings and incorrect assumption, which may remain undetected. Misunderstandings affected two of the three pre-session assessments for one expert despite every effort having been made to explain the task. Other studies (e.g. Phillips 1999) have also shown that when required to make an assessment, experts may make different prior assumptions resulting in radically different assessments. It is only through the structured elicitation process that these assumptions and misunderstanding were detected. It is possible that in real-life interpretations, assumptions and misunderstandings commonly occur in expert judgements but remain undetected.

Group elicitation allows knowledge and expertise to be shared amongst experts and can reduce bias. The results from this study clearly demonstrate pooling of information, resulting in changes in individual opinions. This process is observed for the fault, reservoir and seal, suggesting a systematic process that results in a convergence of individual positions towards a group consensus. The first two stages of the structured elicitation process measure the existing beliefs of the participants. Through the later stages the elicitation process itself acts to create new ideas and insights, meaning that the process itself actually creates beliefs. As the session progressed, individual experts became less certain of their assessment, resulting in wider distributions. This highlights the potential for the elicitation process to reduce well-known cognitive biases, in this case expert overconfidence.

For the seal and reservoir, where there was agreement amongst the group and consensus was relatively easy to reach, the final average individual assessments (Figs 2d and 3d respectively) closely matched the consensus. The consensus distribution also bounded or nearly bounded the range of individual opinions, as shown in Figures 2c and 3c for the seal and reservoir respectively, encapsulating the combined knowledge of the group.

For the fault, agreement was not as strong and the consensus was partially forced (as is typical in a decision-making scenario). The final consensus assessment does not reflect the combined knowledge and opinions of all individuals in the group. In this case 'herding', where the group has been led by or followed one individual (Keynes 1921, 1937), is clearly observed and for the first time quantified: the group has moved towards a single member's opinion, in this case expert E1, with the consensus distribution primarily reflecting the views of this one individual as shown in Figure 4c. To reach a consensus, individual members of the group have agreed to a distribution in which they did not truly believe. One individual in particular, expert E4, gave a final individual distribution that was significantly outside the group consensus to which they had previously agreed.

Where one expert disagrees with the majority of the group, care must be taken to ensure that the potentially important views and information held by this individual are not neglected. It is therefore of significant concern when the final consensus does not span such opinions and in such cases, where a consensus is not easily forthcoming, mathematical aggregation of individual assessments may better represent the full scope of the group knowledge and expertise. Nevertheless, the combination of a consensus p.d. with post-consensus, individual p.d.s is demonstrably a key tool to identify such herding behaviour.

Because of issues with herding and the consensus distribution for the fault, it is likely that the results for the seal and reservoir are relatively more reliable. However, without the structured elicitation process, which allowed both misunderstandings and herding to be identified, and herding to be differentiated from mere convergence of views, we would not have been able to reach this conclusion. It is only by tracking the evolution of individual beliefs and comparing such dynamics with final consensus distributions that we were able to identify herding behaviour and were alerted to problems with the very nature or meaning of a 'consensus' distribution. This method should therefore be applied in a wide variety of interpretation tasks in future.

Using structured, expert elicitation it is possible to produce a probability distribution that a reservoir has the basic required characteristics for security based on the initial assessment of data by experts. In this case we assess the probability that the reservoir exists and that it has a seal and fault. If we assume that this is a sealing fault required for reservoir security, the joint probability gives the probability that the site has the minimum required features. This, however, would be a maximum probability based on the existing data, as other features and characteristics would no doubt also be required. This assessment can then be used to inform decisions regarding further exploratory work and data acquisition activities. The distributions produced by individual experts before the elicitation session show significant variability compared with the consensus and averaged individual end of session distributions, highlighting the risk of using individual experts and the value of using a structured elicitation process.

References

Anderson, J.L. 1998. Embracing uncertainty: the interface of Bayesian statistics and cognitive psychology. *Conservation Ecology*, **2**, 2. World Wide Web Address: http://www.consecol.org/vol2/iss1/art2.

Arnell, N.W., Tompkins, E.L. & Adger, A.N. 2005. Eliciting information from experts on the likelihood of rapid climate change. *Risk Analysis*, **25**, 1419–1431.

Baddeley, M.A., Curtis, A. & Wood, R.A. 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.

Bonano, E.J. & Apostolakis, G.E. 1991. Thoretical foundation and practical issues for using expert judgments in uncertainty analysis of high-level radioactive waste disposal. *Radioactive Waste Management and the Nuclear Fuel Cycle*, **16**, 137–159.

Bond, C.E., Gibbs, A., Shipton, Z.K. & Jones, S. 2007. What do you think this is? 'Conceptual uncertainty' in geoscience interpretation. *GSA Today*, **17**, 4–10.

Browne, M.A.E., Robins, N.S., Evans, R.B., Monro, S.K. & Robson, P.G. 1987. *The Upper Devonian and Carboniferous sandstones of the Midland Valley of Scotland. Investigation of the Geothermal Potential of the UK*. British Geological Survey, Keyworth, Nottingham.

Curtis, A. & Wood, R. 2004. Optimal elicitation of probabilistic information from experts. *In:* Curtis, A. & Wood, R. (eds) *Geological Prior Information: Informing Science and Engineering*. Geological Society, London, Special Publications, **239**, 127–145.

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

Keynes, J. 1921. *A Treatise on Probability.* Macmillan, London.

Keynes, J. 1937. The general theory of employment. *Quarterly Journal of Economics*, **51**, 209–223.

Lowe, T.D. & Lorenzoni, I. 2007. Danger is all around: Eliciting expert perceptions for managing climate change through a mental models approach. *Global Environmental Change*, **17**, 131–146.

Mitchell, G.H. & Mykura, W. 1962. *The geology of the neighbourhood of Edinburgh*, 3rd edn. Memoir of the Geological Survey, Sheet 32 (Scotland).

Morgan, M.G., Pitelka, L.F. & Shevliakova, E. 2001. Elicitation of expert judgments of climate change impacts on forest ecosystem. *Climatic Change*, **49**, 279–307.

O'Hagan, A., Buck, C.E., Daneshkhah, A., et al. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley, New York.

Phillips, L.D. 1999. Group elicitation of probability distributions: are many heads better than one? *In:* Shantaeu, J., Mellors, B. & Schum, D. (eds) *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*. Kluwer, Norwell, MA, 313–330.

Rankey, E.C. & Mitchell, J.C. 2003. That's why it's called interpretation: The role of horizon uncertainty on seismic attribute analysis. *Leading Edge*, **22**, 820–828.

Ritchie, J.D., Johnson, H., Browne, M.A.E. & Monaghan, A.A. 2003. Late Devonian–Carboniferous tectonic evolution within the Firth of Forth, Midland Valley; as revealed from 2D seismic reflection data. *Scottish Journal of Geology*, **39**, 121–134.

Sniezek, J.A. 1992. Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behaviour and Human Decision Processes*, **52**, 124–155.

Thomson, M.E. 1978. *IGS studies of the geology of the Firth of Forth and its approaches*. Report of the Institute of Geological Sciences, **77/17**.

Underhill, J.R., Monaghan, A.A. & Browne, M.A.E. 2008. Controls on structural styles, basin development and petroleum prospectivity in the Midland Valley of Scotland. *Journal of Marine and Petroleum Geology*, **25**, 1000–1022.

Wood, R. & Curtis, A. 2004. Geological prior information and its application to geoscientific problems. *In:* Curtis, A. & Wood, R. (eds) *Geological Prior Information: Informing Science and Engineering*. Geological Society, London, Special Publications, **239**, 1–14.

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.

Received 9 April 2009; revised typescript accepted 24 August 2009.
Scientific editing by Ken McCaffrey.