Getting better judgement working party: bias, guess and expert judgement, in actuarial work

Abstract of the London Discussion

[Institute and Faculty of Actuaries, Sessional Research Event, London, 18 January 2016]

This abstract relates to the following paper: Tredger, E. R. W., Lo, J. T. H., Haria, S., Lau, H. H. K., Bonello, N., Hlavka, B., Scullion, C. A report by the Getting Better Judgement Working Party. *British Actuarial Journal*, doi: 10.1017/S1357321716000155.

The Chairman (Mr T. P. Overton, F.I.A.): Tonight's paper tackles a fascinating area of research that really does have relevance right across actuarial work. It had the distinction of winning the Brian Hey prize at GIRO 2015. Let me first introduce some of the authors.

Ed Tredger is a Fellow of the Institute. He completed a PhD in climate modelling for insurers at the London School of Economics. He has since worked in the London market for Beezley, UMACS and currently at Novae, as a Senior Pricing Actuary. He has a strong interest in actuarial research, having given talks at several conferences, including "R in insurance", GIRO, London Market Actuaries' Group (LMAG) and having chaired the Getting Better Judgement Working Party.

Sitting next to Ed, we have Catherine Scullion from Lloyd's. She fell into actuarial work following a natural sciences degree, which had a focus on genetics. On joining EMB in 2009, she was involved in a wide range of general insurance work across personal and commercial lines. In 2014, Catherine joined the Government Actuary's Department and managed to infiltrate some familiarity with actuarial work into 17 of the 24 ministerial departments. Following that she has been at Lloyd's. She has taken on the actuarial lead in reporting and strategy.

Sejal Haria is a technical specialist in London Markets Supervision at the Bank of England. She leads the PRA's response on HM Treasury's initiative to attract insurance-linked securities to the UK and is leading the review of Flood Re's application for authorisation. She has represented the UK in Europe on a broad spectrum of issues, including EU/US dialogue and discussions on catastrophe risk. Prior to joining the PRA, Sejal was with PwC, advising Lloyd's and London market players with a specialism in mergers and acquisitions. She has sat on a number of working parties and spoken on topics including Enterprise Risk Management (ERM) and non-life insurance strategies.

Jo Lo heads Actuarial R&D activities at Aspen, which has taken him into a variety of areas, including actuarial risk and catastrophe risk. He is an enthusiastic supporter of research in the profession. He has co-authored several research papers and chaired a number of working parties and is currently serving as a member of the GIRO committee.

Mr E. R. W. Tredger, F.I.A. (co-author, introducing the paper): I am going to give a relatively high-level overview of the background to the working party and the processes we went through to come up with the paper which we are presenting tonight.

The working party began in late 2013 with an article that we published in The Actuary magazine. It was catchily entitled "Sitting in Judgement". In that we tried to raise awareness of the issue of expert judgement and to see what issues we could throw out there for the profession to take on board.

Based on that, we formed a working party and we then conducted a survey, kindly sponsored by the IFoA, in order to try to gauge whether this really was an issue which a lot of people across the actuarial profession were interested in us looking at.

We received 220 responses, which I understand is a very good response for an IFoA survey. Needless to say, many people who did respond were quite positive about us taking on expert judgement as a subject.

We then carried out a literature review, trying to see what work had been done in other fields on expert judgement. We were aware of a few high-profile publications that have come out in recent years, most prominently Daniel Kahneman's book, *Thinking Fast and Slow*, which captured the public's imagination about the propensity for the human mind to make mistakes in all kinds of different situations.

Having read some of the literature, we felt that the case studies were quite fascinating in some regards. We thought about how they could be taken over into actuarial work and we considered what lessons we could learn.

I am going to give some examples from the literature which I hope will point you in the same direction as our thinking. One study was carried out where the respondents were asked what proportion of UN countries come from African states. The snag is that before being asked the question, they were shown a spin from a roulette wheel. The roulette wheel was rigged so that it would produce an answer of 10 or 90. Those people who saw the number 10 gave an answer of 25%. Those people who saw the number 90 on the roulette wheel gave an answer of 40%. The true answer is somewhere in the middle.

What that demonstrates is that there is what we call anchoring bias which is that people, even when they see relevant information, will somehow be mentally anchored towards a higher or lower level. We then thought about how actuaries, claims handlers, underwriters and management are influenced by seeing a set of numbers, either relevant or irrelevant, before then being asked for a key piece of expert judgement.

Another example: people were asked about whether they were happy to drive 10 miles in order to save $10 on a $40 toy. The majority of people were happy to make that 25% saving. However, most people were not happy to drive 10 miles to save $100 on a $10,000 car. Whilst a higher absolute amount, this is obviously much smaller in proportion. This is what we call framing bias, which is when the context of the question leads to a different outcome on people's behaviour, even though it would have been logical to drive to make a discount on the car.

A third and final example: people were asked in studies, which is the more common in the English language, words that begin with the letter "R", or words where the third letter is "R". Most people would guess that it was words that begin with "R" because we can all think of words such as road, ran, ring, and rung. There are many words which come to mind easily. But, of course, the answer is

that it is words where "R" is the third letter which are in fact the most common: for, word, car, per cent, etc. This is an example of what we call availability bias, which means that things that come more readily to mind achieve a higher probability in the way that we see the world.

When you are thinking about this in terms of work, you need to consider what is front-page news, what are the high-profile risks and what are the things on which people are focussing? Then you should consider what are the less available things which people may miss out?

The literature gave us a kind of vocabulary to talk about these different biases and heuristics, as they are called, when bringing the subject into actuarial work: anchoring, framing, small sample bias, availability bias and substitution bias. We found these quite useful concepts in order to pinpoint and specify what we mean by the different biases and errors of judgement.

We went on to deep dive into certain, what we thought were key, areas of actuarial expert judgement, armed with this collection of vocabulary and knowledge of the literature. Where do these biases come up in actuarial work? Can we pinpoint certain examples of these biases and then can we find ways to do things differently in order to reduce those biases?

The result of the work, through these case studies, and through some empirical studies which were carried out at LMAG and a capital modelling seminar run by the IFoA, at GIRO and also in-house, is a set of proposals about how we might get better judgement.

We do not feel that these are necessarily yet full gold standards. But they are useful ways of thinking about where bias can arise, and tools to think about how it might be potentially reduced.

We hope that people will take away something about the different biases that exist, the prevalence of bias within actuarial expert judgement and how this can add commercial value to many companies on top of the regulatory angle.

Catherine (Scullion) is now going to take you through some of the case studies that we have carried out.

Ms C. A. Scullion, F.I.A. (co-author, introducing the paper): Following on from those general remarks about biases and the literature outside of the actuarial context, I will take you through some of the detail that we considered when we translated those into the actuarial context. The working party members have a general insurance background. That is from where our case studies arose, but we would be very interested in examples, and particularly any techniques that those from other practice areas have addressed with similar issues. We feel quite confident that they do arise in most areas of actuarial work.

In any situation requiring elicitation of an expert judgement, we first of all consider some general principles that would apply before taking those into a particular situation. I will start on those general principles which are covered in the paper. They all stem from a need to consider some of the more qualitative aspects of the work with which we are involved as actuaries: the need to make sure that within those aspects, we avoid any sort of "trust me, I am an actuary" mentality; and that we acknowledge the uncertainty around some of the areas that we are discussing.

The first aspect that we considered was that of preparation. Preparation for any discussion that is going to be used to elicit a judgement from an expert is really key. It definitely should not be

underestimated as part of the process, both in terms of preparing for what approach will be taken and also looking at what data will be used: what is available; how can it be presented; and how should it be used to structure discussion with the experts?

Moving onto the discussion, it is very important to manage the potential for any bias with careful structuring of the questioning. That is where preparation has been key. The case studies will provide a bit more flavour around what we mean. Ed (Tredger) mentioned Kahneman's book, *Thinking Fast and Slow*, where it is key to slow down the thinking of the experts, to avoid obtaining a gut feeling or instantaneous response, rather making sure that there is clear rationale so that it can be considered carefully to derive the judgement. Part of that is making sure that that aspect is documented. The document can be referred to again and checked for consistency with future judgements when that is next required.

After obtaining judgement from an expert, we have thinking around how we validate that judgement. It is not only validation of the judgement itself, since there are quantitative ways of doing that most of the time, but also validation of how the judgement was elicited, how the discussion took place and how it was documented.

There are other techniques that we can use to validate the judgement in a more qualitative way. Can we check the coherence of the judgement by using a different set of questions and check that we get a consistent result? Can we check it against any independent information? More widely, should we have peer review of that process? We typically have peer review of the numbers, but the process itself is often taken for granted. We need to consider how we could improve on that aspect.

We also had a few debates in the working party on where the actuary sits in this process. Which part of the actuarial skillset is relevant to doing this activity and doing it well?

The first case study which illustrates some of these points is one around high impact, low-probability events. These are a particular focus for general insurance actuaries. The question is: "What do you expect your 1 in 200 loss to be?"

Under Solvency II, the 1-in-200 event becomes the centre of such a focus. What we are looking for, in this question, is the worst outcome in 200 realisations of the situation.

It might be something, typically, that an actuary is discussing with an underwriter for a certain class of business and a certain outcome. Essentially, what we are asking for is quite a difficult question: we are asking for an estimate from a distribution that is beyond the data which are available, almost certainly, and requires significant judgement to be applied to something quite technical.

This is a simplistic way of asking the question. It is typically not undertaken, but by looking at it in this context you can break it down and see how some of the issues may arise.

To add a little bit of colour, I would highlight the word "your" in the question. That makes it quite personal to the expert. By doing so, you naturally make someone's judgement less objective. They become defensive. They may become quite protective.

So, if they are an underwriter, writing for a class of business, they may feel as if it is almost their loss, and that will affect the objectivity of their judgement. They may also have over or under confidence

given that situation, or suffer from availability bias. Especially in this context, people tend to live in temporal bubbles. They will be able to recall events that they have personally experienced, but they may not look further back, or to other areas, for equally relevant events for their judgement.

Also, ownership might give them the impression that they should know what you are asking for, so while it is a very challenging question on which to provide judgement, if you are asking it to them personally, they may feel less likely to ignore the uncertainty associated with the situation, and more likely to be over-confident.

The next word that I highlight is the word "loss". Again, the language may cause a bit of defensiveness. Although losses are inevitable in insurance, they can still touch a nerve. In addition, to any emotional response in this sense, there is also an issue with meaning. It is a bit ambiguous. Do we mean an aggregate loss over the year? Do we mean a single claim? Do we mean gross or net of any reinsurance? It is not clear what is being asked for. If the experts do not feel they can have that easily clarified, they will jump to their own conclusion and that might not be the one required by the questioner.

The next word to consider to be of potential difficulty is the word "expect". Again, there are issues with ambiguity. The questioner might have a particular statistical interpretation in mind, but that might not be how the expert sees it. The subtlety of the mode and the mean being the expectation is quite important in skew distributions, and if you do not make that clear then you may not obtain what you are asking for from the experts.

The final point is 1 in 200. It is quite a challenging concept, and difficult to think through. To many people, it is not clear whether you want 200 previous years or 200 future years. In most cases what you want is 200 realisations of the next year, but that is probably the most difficult to picture – it is certainly not the most intuitive way of thinking about the topic, even for people with quite a lot of statistical training.

What we are asking for is quite a remote point. In reality, if we ask for the 1-in-100 event, how different an answer would we obtain? People's minds cut off at a certain level of remoteness, if you like. Everything beyond that may become grouped into one place. So, this is a particularly difficult area to address from a standing start.

In the paper, we discuss ways that we could improve. We have taken a simplistic question, broken it down and you can see quite quickly that there is a minefield of biases and issues when asking an expert for a judgement. As Ed said, we have tried to make suggestions to move towards improvements: we are certainly not suggesting that we have solved the issue. But we have offered some ideas to move the reasoning in the right direction.

The first, as I have pointed out, is really quite a simplistic way of approaching this problem. It would be much better to have a series of questions building up a picture to reach this type of judgement. As I said, the 1-in-200 event is very remote. Most people would struggle to reach that point immediately, so building it up from lower percentiles and other aspects of the distribution that are, perhaps, more familiar would result in a much more coherent judgement that could then potentially be extrapolated to the extremes where experience is known to be very sparse.

583

Second, in order to avoid availability bias or people using their own reference examples, pre-defined scenarios could be prepared. That would avoid the experience of the experts dominating their judgements. It would slow down their thinking to make them consider these scenarios.

Then, on the validation side, which I touched on as being one of the general principles in trying to reduce these biases, is to play back the answer that they have built up to allow them to take some considered time to make any improvements. That can also incorporate any independent information, for example if there is any existing modelling, particularly around catastrophes, or information from reinsurance performance. If they provide a particular event, you can check how reinsurance would perform, and that would make sense in the context of their judgement.

The second case study that we looked at is around parameters for dependency structures, typically in a capital model or another variability model.

In the first case, the question was quite simple to obtain a point estimate. In this case, we are asking for something quite complex, most likely subject to even greater uncertainty and even more intuitively difficult for someone without statistical training. Again, this might be something that is discussed between actuaries and underwriters in order to establish which classes of business are expected to be strongly dependent and which are less so in the modelling. We might be aiming to populate our matrix to look a bit like Figure 1.

We have various classes of business and how related each one is to each other. We, as a working party, discussed a number of levels to the judgements in this regard: the dependency structure itself; what form that would take; and the parameters for how it is correlated.

In the paper, we focussed on the parameters themselves and how those might be judged. Equally, there are issues with the structure.

In this case, we probably have even less data and information than we did in the estimation of an extreme scenario. It is very difficult to use that data to imply strongly supported parameters for dependency. Significant judgement is involved and it is quite difficult to validate.

Ed touched on anchoring bias. That is a real danger in this case. The bounds are naturally 0 and 1. So the expert might be overly anchored to either one of those. They will adjust either upwards from 0 or downwards from 1. They may not do so as much as they would if they were just

Judgement in practice – case study 2
How do we select parameters for a dependency

	CoB 1	CoB 2	CoB 3	CoB 4	CoB 5
CoB 1					
CoB 2	20%				
CoB 3	30%	20%			
CoB 4	20%	20%	30%		
CoB 5	40%	20%	30%	20%	

Figure 1. Judgement in practice - case study 2

584

asked to estimate the numbers. Equally, because of the uncertainty involved, there might also be strong incentives to anchor to any existing benchmark or previous exercises because the judgement is so uncertain.

Again, as in the previous case, there might be availability bias, where connections seem obvious once they are known. In the financial crisis some liability classes deteriorated at the same time. That is usually now commonly recognised in models. If such events have not yet occurred, they may be explored to a much lesser degree.

In this context as well, framing of the questions is another area on which Ed touched. This has significant technical challenges. If you want to estimate the dependency based on extremes, you might try to question around: 'If class of business one is beyond the 99th percentile, what is the chance that class of business two is also beyond the 99th percentile? Or vice versa? Those are the technical demanding questions to frame, but framing is very important to elicit the expert response.

Similarly to the other case, preparation and data are key, especially if there are any data that you can use to show potential effects. Again, predefining scenarios where it is discussed how effects could happen is better than relying on the examples that the experts can recall.

Also, avoiding anchoring, potentially by not using numerical ratings but by using some kind of qualitative rating, will improve the possibility of too much anchoring to 0 or 1.

Again, feedback to the experts will be very useful. Once their judgements are input into the model, it will look very different, and feeding back what that implies about various likelihoods across different types of business should be very informative and allow the experts to assess their judgements and have more of an understanding of the implications.

Having pointed out the issues, Sejal (Haria) will provide more detail on judgement quality.

Ms S. Haria, F.I.A. (co-author, introducing the paper): From the case studies we have the idea that expert judgement can be of varying quality. I think this is something of which we are all aware. It is not something that is often explicitly acknowledged. We also have the idea that it is the elicitation process that can be fundamental to the quality of the end judgement. This is really interesting because it is nothing to do with the experts.

The aim of our paper is to stimulate debate on this topic and to share some ideas for rethinking expert judgement. Figure 2 shows the concept of the guess universe.

Conceptually, expert judgement and guesses are not so far apart. Expert judgement is a subset of the guess universe. So it is a view formed in the presence of incomplete, and often imperfect, information.

The fundamental questions resulting from this are: How do we distinguish between guesses and expert judgements? How do you avoid being blinded by experts? How can non-experts, including boards, properly challenge expert views and understand embedded assumptions?

In the paper, we develop this thinking further and we share some ideas about how we propose it should be tackled. For example, we talk about how we often need to seek extra views that are not
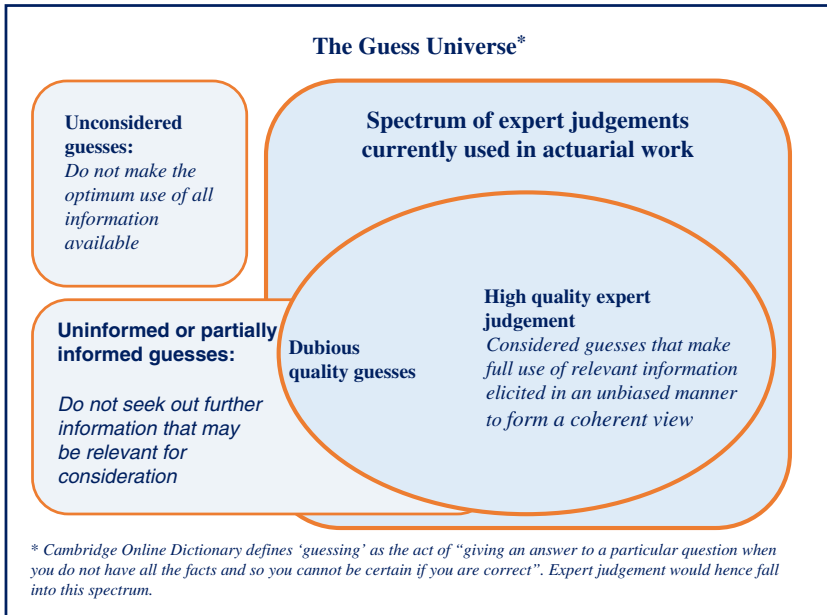
585

The Guess Universe*

Unconsidered guesses:
Do not make the optimum use of all information available

Spectrum of expert judgements currently used in actuarial work

Uninformed or partially informed guesses:

Do not seek out further information that may be relevant for consideration

Dubious quality guesses

High quality expert judgement
Considered guesses that make full use of relevant information elicited in an unbiased manner to form a coherent view

* Cambridge Online Dictionary defines 'guessing' as the act of "giving an answer to a particular question when you do not have all the facts and so you cannot be certain if you are correct". Expert judgement would hence fall into this spectrum.

Figure 2. The guess universe

necessarily in the line of sight of our experts. This does not come free: it comes at a price. So, we need to weigh up the cost/benefit, and we need to attach credibility to different views.

As Catherine (Scullion) has already mentioned, we look for biases in the formation of our judgement and the processes that we use. We want robust group discussions and to consider how different views are amalgamated in order to form a coherent whole. Also, we want to allow for appropriate validation and feedback to the extent that this is possible. This is really good because it allows a way to improve learning and to think about refining any prior preconceptions which may or may not be correct.

Next, I wanted to touch briefly on Groupthink. Groupthink was a phrase first coined by the psychologist Irving Janis in the early 1970s.

I think we all intuitively have an understanding of what we want out of a widely engaging group discussion. The idea is that we obtain different views from different information roles that are present in the company – there are some quite deep views; some more shallow views – and amalgamate them and land at what is supposed to be, in theory, a more robust view.

We also now have an idea of Groupthink, when group discussions go wrong. We discussed some of these features in the paper, such as when the discussion is not really a discussion but it is more of a play act, or when there is insufficient challenge behind any prevailing view.

The paper goes on to explore some possible solutions. It touches on the fundamental role of the chair in controlling the group discussion and making sure alternative views get sufficient air time. It talks about how discussions are structured and how views are presented so that views that may seem to be in the minority receive sufficient discussion time. It also talks about how we try to prevent our natural tendency for conflict avoidance.

586

So, we do not want to reach consensus too soon and we need sometimes to be in the uncomfortable position that we allow an argument to take place between two quite different and conflicting opinions in order to form a more robust view.

In this meeting, I want to look at successful examples of Groupthink and contrast this against how we currently arrive at group judgements and see what we can learn.

This may be the first time in an actuarial lecture where we have a study of the behaviour of ants. The question is: how do the ants find a suitable location for their new nest when their colony is destroyed? In the problem set up we deal with similar issues to those which we encounter in the actuarial sphere. We have a deluge of potential information sources. We have significant cognitive demands placed on a single ant. So how does it deal with it all?

How do the ants organise themselves, and why are they not overwhelmed or paralysed by different choices? Two features to note immediately. First, no single ant is in charge. Second, they distribute this task among their colony members.

The ants send out scouts, and these scouts check out potential home sites. They look for things such as the size of the entrance and how big is the cavity. If the ant likes what it sees, they return to the colony. Then the ants send out a pheromone message saying "follow me". Then another ant checks out the potential site. If she likes what she sees, she goes back and repeats the process and that brings in other ants. If she does not like what she sees, she just returns to the colony and it stops.

If enough ants like a site, the colony reaches a quorum and chooses the new home.

So what are the potential learning points? There are some set rules around how the ants arrive at judgements and these are specified in advance. There is a democracy of opinion. No opinion is given more favourable voting than another at the outset. The process of feedback from the other ants is actually very quick and they are being judged purely on their result.

If we contrast this to our model, we place more emphasis on experts so we give quite a lot of weight to expert opinion. But how do we make sure that those are challenged robustly and we do not reach consensus too early? Some expert might have a pet view. Maybe if we are discussing a tail outcome and we need them to come up with a broader set of alternatives, we somehow need to stretch the possibilities available to them.

Often we have a much greater time lag between the setting of any judgements that we make and the eventual outcomes. In order to test the quality of our judgement formation, we need to keep a record not only of the decision but the method by which we arrived at the decision and the information source upon which it was reliant. Then we would be in a better position to see what can be improved next time.

Each ant only processes a very small amount of information. This is really interesting. The parallel here would be how to reduce unnecessary cognitive demands on any expert so that they are doing just what is in their expertise. How do we structure the elicitation session so that the facilitator or chairperson can weed out any superfluous information? For example, when the underwriter is trying to form a view of the 1-in-200 event, it is not something that, conceptually, they are used to dealing

587

with on a day-to-day basis. It is outside of their field of expertise. Should we be asking them more relevant questions and then trying to do the extrapolation?

Some thoughts that I want to leave you with: there are methods that are more closely aligned to some of the ant principles. One method in particular that we mention is the Delphi method. It uses some of these principles. It introduces the concept of preserving anonymity of opinion, almost forcing a democratic process in opinion selection. Information flow is processed in a structured way rather than in the ad hoc way which group discussions can sometimes flow unless they are properly managed.

Finally, there is regular feedback of forecasts so earlier statements can be more easily revised and there is not the same stickiness because people are no longer so attached to certain views. It is anonymised at the outset.

Dr J. T. H. Lo, F.I.A. (co-author, introducing the paper): Before our concluding comments, it would be useful to say a few words about our work in blending judgement and data. When blending judgement and data to form an answer, more weight is intuitively given to data if data are plentiful, if experience is not volatile and if judgement is uncertain. We highlight Bornhuetter–Ferguson as an example in which more weight is given to data for the more mature origin periods where experience is usually less volatile.

In section 9 of our paper, we highlight the well-established Bayesian theory as a more general framework to help assign weight or credibility. At the coal face actuaries do not seem to make use of Bayesian credibility formulas often, but rather prefer judgementally to base and adjust credibility weights on the three intuitive principles.

Laying aside the pros and cons of such practice, an interesting question is how close our judgemental weights are to the Bayesian ones. Towards this, an empirical experiment was done at the beginning of one of our presentations. The audience was asked to provide answers individually to a frequency estimation problem which was expected to lie in the participants' comfort zone, if you like. The results suggested that we are not really natural Bayesians. Indeed, around 60% of the respondents' ignored external information totally. Moreover, there was no evidence that estimated weights varied between those with 5 years of data and those with 15 years. We believe that training can make a difference.

On another occasion, we performed the same test on a different audience but placing the test in the middle of the presentation immediately after awareness was raised regarding cognitive heuristics and biases.

The 18 responses gave weights which were statistically closer to the Bayesian theoretical ones. The discussion to this section, section 9, concludes by urging the reader to explore Bayesian approaches more in their work and as an area of research. We highlighted the more objective nature of the framework against the seemingly subjective judgemental approach to coming up with weights, using these approaches to give enhanced transparency and better buy-in.

The actuary would, however, need to manage the risk of miscommunication in which the user of actuarial work imagines that projections are now much more accurate even if very little data was available.

Moving on to section 10 of the paper, judgement is clearly an area of general interest to actuaries. We had an unusually large response to the working party survey, 220, whose results we summarise

in the appendix of the paper. One of the questions asked for the respondent's opinion as to the most important actuarial activity. Is that to do with making sure that the end user understands your actuarial output? Is it to do with technical modelling excellence? Or is it to do with formation, communication and use of expert judgement? So, basically, communication or technical modelling excellence or expert judgement.

In our 2014 survey, a little under half of the respondents thought the most important actuarial activity was making sure that the users of actuarial work understood their outputs.

The remaining respondents were equally divided between technical modelling and expert judgement. Perhaps I should have been more mindful of the Groupthink phenomenon before I asked for the quick straw poll.

But, actuarial research outputs do not really reflect this split from our survey. Academia is giving us fantastic modelling insight. I love some of the papers in the *ASTIN Bulletin* and some in the *Annals of Actuarial Science*. I also enjoyed reading some of the IFoA's working party presentations on communication aspects. Apart from a few passing references here and there in papers and presentations, judgemental topics have yet to make regular appearances in actuarial conferences. Could the paper presented a couple of sessions ago by our life colleagues, or indeed this one, stimulate a turning point in the profession's research direction?

We singled out empirical studies as a good area for research. A bout of evidence would help to construct a solid basis on which to base best practice recommendations. Even research in communication could benefit from empirical studies. How might different decisions be made if we communicated our actuarial work differently? What judgemental biases might you receive from your experts if you framed questions in different ways?

Several specific questions are listed in the paper. There may, moreover, be opportunities to engage in useful collaborative research with other professions. Information, communication and usage of expert judgement are surely also an area of concern for underwriters, claims handlers and investment officers. Psychologists would help bring insight into the personality angle.

Many respondents in our 2014 survey said that they planned in the next 12 months to generally talk about the subject in their company, perhaps informally with their colleagues; to read more; attend conferences or workshops about the topic; and/or to facilitate regular sharing sessions of good practice. A few also put down running systematic reviews of judgements in their organisations' actuarial work and formally incorporate judgemental thinking into peer reviews. Some were contemplating organising training for actuaries, or even for non-actuaries, to produce a checklist for themselves and other colleagues. There was even one individual who said they would think about organising an internal exam for their students as part of their study requirements.

The Chairman: I will now open the discussion to the floor.

Mr J. G. Spain, F.I.A.: Groupthink is a real problem in the pensions area so I am grateful for this paper.

May I just pick up two points that Sejal (Haria) mentioned and perhaps disagree? You said I think that no ant had any more of a vote – any more power – than any other ant. I think the earlier ants did

589

have more influence than the others because once you had a critical mass, everybody else was following. I think that is what you get with Groupthink all the way.

Second, in terms of recording discussions, what is really important is not only to record what the decision was and the mechanisms used to arrive at that decision, but, most important, to record what else was considered, discarded and why those things were discarded. Then people can come back and say, "We told you so".

Ms Haria: On the second point, I would agree with you. It is very important to have a consideration and have the documentation as to why things are discarded. In the light of perfect foresight perhaps you can see the reasons for discarding those alternatives may not bear out in the future. That also actually provides a very good learning opportunity.

In terms of the ant example, yes, again, you are absolutely right. Once the second critical mass builds up, all the rest of the ants follow. The point there, though, is that in their initial foray, no one ant is in a more privileged situation than any other ant. That can be very different from a group discussion.

When your boss is in the room, how many discussions have we seen where the minors tend to follow the boss's opinion even when they have a different view entering that room than they do when they communicate in the discussion?

This is the more serious point that I was trying to reach. There are situations where once a certain decision has been reached, so long as it has been reached in a robust fashion then discussion is closed down. In the ant example, once a certain quorum has been reached, it is not beneficial for further ants to do that investigation and tackle that same question. This is more about how we reach that robust outcome in the first place.

Mr A. D. Smith, Hon. F.I.A.: I am going to pick up on a point that is raised just once in this paper, in paragraph 5.2.2, which is about the need for long-term testing. I know that Sejal (Haria) did say a bit more about feedback, and Jo (Lo) also mentioned the need for empirical studies.

It seems to me that in the GI space we have, probably, tens of thousands of data points of historic expert judgements which are then published in the PRA returns (and before that the FSA returns, and the DTI returns now) giving ultimate losses for tens of thousands of combinations of company line of business, underwriting year and year of filing.

Analysing that data is tricky because it is messy, companies merge and contracts are reclassified. But the data, in principle, is there. You have a very large number of judgements that you can analyse.

Also there, of course, are the underlying claim payments and incurred data from which you could construct mechanical estimates and ultimates using a chain ladder method or something like that.

For lines of business that may have been written more than 5 years ago, you could sensibly ask which of the mechanical method or the judgemental method is closer to the actual outturn. That work could, in principle, be done. To my knowledge, it has not been done, or at least the results have not been published, and the few calculations that I did suggested that we might not obtain the answer that we are looking for and that it might not, on a purely statistical basis, vindicate the use of expert judgement.

590

So what is it doing instead? It is serving some other objectives. For example, use of expert judgements may be smoothing the emergence of profit. The use of expert judgement by reducing volatility may be mitigating capital requirements.

It seems to me that, in those cases, the discussion we need to have is what purpose is being served by that judgement and are those purposes legitimate?

To conclude, I should like to encourage the authors, or indeed somebody else, to do analyses of the judgements that are out there. We are making a few speculations about properties of judgements that could be either verified or alternatively refuted with respect to evidence. It will be nice to see the evidence.

The Chairman: That is a very interesting point, particularly around other purposes for judgement other than purely trying to improve our estimate of the truth.

Mr Tredger: It is an interesting question. Your point about the purpose of what the judgement is trying to achieve is quite an interesting one. There certainly can be negative ways of using judgement, for example, as you were saying, to try to distort financial results to show some kind of gain in capital or whatever.

It is worth noting that, once you understand the way that expert judgement works, and bias works, you can use it to your own advantage. You can introduce positive anchoring, for example, if you know that someone is going to have a tendency to understate their reserves. You spin a roulette wheel, and give them a high result before you talk to them and hopefully that pushes them up.

You can use ways of guiding – even simple things about how you frame the question to people. If you know beforehand that there is going to be a bias, you can almost pre-empt that bias and try to cancel it out with something else. It can be useful to be aware of that dynamic and to be armed with the tools to mitigate it as you go through the world of expert judgements.

Ms Scullion: I think that also comes back to the first question on documentation and some of the aspects highlighted in the case studies. In addition to making sure you document what was not considered or what was disregarded, the reasons underlying your judgement should also be documented.

Then, looking back at your reasons, determine which aspect was flawed or whether something else was being incorporated into your judgement. Being able to look at the judgement at that level of detail will help draw out the explanation.

Ms Haria: I think that there is a deeper point here which we do not explore in the paper. It is around some of the incentivisation of the experts forming the judgement, or the questions that the experts are actually being asked to address. Andrew (Smith), in your valid example, I do not think that the question they were targeting was "Will my loss ratio hold up in 10 years' time?" It was more "What can I live with next year, and what is the appropriate communication to the market?"

In any backwards analysis we do of judgements, we need to be conscious of how that incentivisation was working.

If we want to obtain a pure judgement, we need the incentivisation to work with us, otherwise we are working against a fast moving current and we are not going to get anywhere.

Mr M. Slee, F.I.A.: My background is life but I agree the topic applies in life as much as in GI and in pensions.

I would like to pick up on a few points. One bit in the paper talks about expert judgement as being the method of last resort. I think I would disagree with that comment. You could say that, surely, the big selling point of actuaries is we provide value, added value, by being able to make a judgement, or to analyse things and to provide the extra bit of judgement. It is not just turning the sausage machine and out comes the answer at the end of it. It is the fact that we can explain it and provide that extra bit of commentary which stands us out from the crowd.

You also talked about bias. We definitely do need to avoid that. My daughter bought me a book at Christmas entitled *How Not to Be Wrong – the Hidden Maths of Everyday Life* which talks about how you can very easily think you see a pattern in something but that pattern does not exist. It is just a bit of random noise, but you can so easily believe that something is happening. We need to be aware of that in what we are doing.

One comment on Groupthink. What can happen if you are not careful is, even if you get together a panel of experts, or experts and those with a wider knowledge, they might all have the same background and the same experience. Although they are debating it, they are debating it from a very narrow pool of knowledge. You need to make sure you have somebody else who can provide a different perspective. In some ways that is why it helps to have a turnover of staff within a company and not just have a company where everybody has been working there for 20 years.

If you have people coming in who have worked elsewhere, they have different experience and they can build that in to avoid Groupthink.

My final comment is on training. You talked about what training is necessary. That led to the thought: is that technical training or is it professional skills training? The answer is that it is probably a bit of both. The Actuaries' Code talks about competence and care. It talks about communication. There is no reference to judgement within the Actuaries' Code. Maybe there should be. It is up for review in the coming months and maybe that is one of the things at which we should be looking.

The Chairman: There is a number of things to pick up: the value of an actuary; what we add to the process; and how important it is to be aware of what might skew our judgement. There was also a point on diversity and a point on training and a comment on whether we might incorporate judgement into the Actuaries' Code.

Mr Tredger: Just to come back to the first point, that is a really interesting question. You are saying that actuaries are valued for the expert judgement that they can overlay on top of technical skills. I would certainly agree with that comment. In the area in which I work, GI and the London market, there is intensive use of expert judgement, both by non-actuaries and by actuaries. It is critical to our work.

I think that there are actuaries, perhaps more so in America than in the UK, who would see actuaries' value in providing the quantitative side, the analysis and not being as involved in overlaying expert judgement.

I think expert judgement is really critical for us. I am not sure that every actuary would necessarily see that as a key part of their work. I think that a lot of actuaries will come into the profession from a

592

background of maths, statistics or sciences, where there is an answer. You do exams, and again there is an answer. You are trained to be very good at producing the numbers. You can go through a quarter of a century of education being told that this is how you turn a handle effectively to produce the right answer from a formula or from a set of numbers and you are never asked, "But what do you think? What do you think about that equation?"

That feeds into your last point, which is that there can be a certain bias in terms of the people who become actuaries. Initially, they are interested in the technical side and then, later, they find that they have to provide judgements. I wonder how well-equipped everyone is to deal with that side of the real world, even after all their education.

Mr M. G. White, F.I.A.: I will try to give an answer to that point. Generally, on the value of actuarial judgement, I would say where it has most value is where it is based on a deep understanding of the business area concerned. So, for example, in life insurance and pensions there may be one class of people who have a decent idea across the entire business of how it works. But, we must be extremely careful to recognise that for some of the most important assumptions of all, such as future investment returns and inflation, we really have not got a clue.

Ms Haria: Diversity again: I completely agree. I think that our paper does lead to this conclusion. It is not only professionals within the insurance base, so, for example, where we are developing 1-in-200 events, let us face it, insurance professionals are not very imaginative people. If we are trying to come up with a 1-in-200 event, we need to imagine completely new scenarios that have not happened yet, but that could come to pass.

Can we see parallels in other professions? Can we see parallels in other industries that we could adapt and develop into insurance examples? We need to spend the time going out exploring these parallels. It is not just about taking information that is readily available.

This feeds into the first point, where the question arose around how actuaries add value to this process. I sense that a lot of actuaries feel expert judgement is the method of last resort because they are much more comfortable dealing with their data sets. It is not clear to them how they enmesh qualitative judgements and qualitative aspects into their ordered, statistical view of the world.

Ed (Tredger) mentioned the example of the American actuary. They are traditionally very keen on their data to the extent that they do not want to overlay any expert judgement at all because inevitably any expert judgement that you overlay will be wrong.

We need more of a mindset which does not just consider those data sets to which we have access, but also asks whether those data sets are relevant to the problem under consideration? What other things should we be bringing to the table? There is a lot that could be done in this area.

The Chairman: Jo (Lo), did you have any thoughts on training, the Actuaries' Code, and how we might change the way that we view this topic as a profession?

Dr Lo: I very much agree with the comments made earlier. I might even go into a little bit more detail. Potentially, a training programme might include some kind of facilitation training, for facilitation of judgemental elicitation.

There is also a step back that we might need to think about. There has been a flurry of big data discussion in recent years. Some actuaries have started to ponder whether we should now think about attracting computer scientists to the profession as well as mathematicians or people who are relatively numerate.

I am wondering whether we should also try to attract psychologists. It would be wonderful to have psychological input. Then, if I then take another step back, in order to attract psychologists to come to our field, we might also need to demonstrate that we are serious about it. One way to demonstrate that might be through research.

I do not know what excites you when you get things through the post. Nowadays we get bills and statements in the emailing world. Even the *Annals of Actuarial Science* and the *ASTIN Bulletin* are now in electronic form. But, I do get excited by the actuarial journals.

Earlier this year, as a kind of Christmas present, I also received another journal called the *International Journal of Forecasting*. The journal contains peer-reviewed papers. Some are very technical about time series analysis. You might compare that to your usual *ASTIN Bulletin* or even the *Annals of Actuarial Science*, but I was relatively excited about a peer reviewed paper on herding behaviour in business cycle forecasting.

When Andrew (Smith) mentioned empirical studies, his proposals would consist of the painstaking work of organising data and analysing it. This is a totally worthwhile subject to think about. There could also be some psychological input.

Also, we talk about reserving cycles quite a lot in our IFoA working party. There is a lot more study that could be done in that area. I think that it is worthy of an actuarial science PhD programme, and there are a few others. For example, we asked whether we should go from a lower return period into a 1-in-200 event. A paper was thinking about the problem the other way round. Should we go to more extreme return periods and work backwards? What is the absolute maximum loss that you can think of?

There are various different ways. Research will be fascinating. When psychological students, graduates, etc., see that we are also very interested in those topics, we will attract different types of experts and it is hoped that we will be a more exciting profession as a result.

Mr M. H. Tripp, F.I.A.: At the moment, within the IFoA, there is a lot of conversation about the future of our profession and what it is that makes an actuary an actuary.

Whichever way we skin the cat, we are possibly not a very diverse group of people. Perhaps the real challenge is what it is that makes our expert judgement in the field appropriate. Martin (White) referred to the fact that we are not very good at predicting investment returns. One area of judgement in which actuaries might become involved is to do with investment and asset management. But we have never really succeeded in that space. If you talk to an investment expert, they would really pooh-pooh actuaries when it comes to giving any guidance at all on stock selection, sector selection, strategy and all the rest of it. Nevertheless, we have investment exams. Maybe we should just drop investment exams?

Another area of expert judgement is maybe in life and pensions when it comes to talking about longevity. Whereas, we can see the investment thing is slightly left field, longevity is much closer to home. You can think about everything that you have talked about. You apply that to longevity as

594

opposed to picking ultimate loss ratios. Would the thinking change? How could you help the pensions actuary get beyond the age of 88, as the ultimate age that people, on average, are going to die, or whatever?

I certainly can remember in previous existences talking with pensions actuaries about longevity. Whatever you say, you are going to be wrong. But that brings me to putting my hand up and I think the most important thing is communicating the thinking behind what we are saying so that you can help non-experts make their own judgements. That, in some sense, is the safety net. In the investment area, the more that you can communicate about your thinking and why, the more you help people make better judgements. The same applies to longevity.

I would be interested in what you think makes an actuary, an actuary when it comes to this space?

Mr Tredger: One thought relating to our paper: when you are in a very difficult situation of expert judgement, it is very important to acknowledge the uncertainties that exist. Perhaps the Bayesian framework helps. In a Bayesian framework, we are not only taking what data we have, weighting that with the expert judgement, we are also implicitly putting some kind of uncertainty around the parameters that were input.

If we are in a situation where all we can do is expert judgement on, say, longevity because that is pretty much all we have, then if we know we are uncertain, we should put larger uncertainty within the hyper parameters of our distributions and explore the different alternatives. That gives us a mechanism for building in some kind of allowance for the uncertainty in the judgement rather than saying simply we have the right point estimate or that there is a variety of different point estimates that can be explored.

Part of a good expert judgement is also knowing how sure you are of it. That is why we identify small sample bias and overconfidence bias. We need also to say we do not know what the answer is, but we reckon it is going to be 2.5% per annum with this distribution around it. If that is the best that you can get to, at least you will have built-in some allowance for uncertainty. That should somehow be reflected in the result.

Your question was what makes an actuary an actuary? I suppose we have the ability to draw on different statistical skills to be able to analyse the data, obtain what we can out of that analysis, and once we have exhausted that approach, we can look at judgement. We can blend the two. We can build in uncertainty, acknowledge uncertainties, produce our estimates and then communicate back what those uncertainties are to say that we have done all this work, and this is the process, and the three key assumptions are A, B and C. There is the joining up of the qualitative and quantitative.

Ms Scullion: It comes back to some of the other discussion around the role of the actuary. What training is needed?

One of the things that we found when we were doing the literature review was that in other fields there can be a much more structured process for obtaining judgement from experts. It involves a number of different individuals with different skillsets. There is the technical aspect, also the facilitative and the psychological.

What role should and does the actuary play? And are we too limited in focussing on the technical side?

595

Mr T. J. Llanwarne, F.I.A.: Everybody has said the same thing and I am adding to that chorus. Although the paper is specific to GI, the topic is much wider than GI.

I would urge people in GI, and potentially if there is an equivalent for the other specialisms, to see that there is an important need to follow what you are saying in the paper.

Let me expand from that into what do we mean by judgement? I am not sure that I am going to be quite in the same place as others on this question. Maybe that is because of biases. Maybe I am biased.

I have spoken to a number of mathematics professors, who are all Bayesian these days. I started reading a book by Susskind and Susskind. They spoke to Simon Carne and Ashok Gupta, and maybe some more other well-known actuaries, to obtain some feedback before they wrote it. Effectively, they are saying all professions are in decline, and ask whether any profession will exist in 50 years' time.

I pose that as a rhetorical question to which we do not know the answer. They give reasons for trends. You can see why. One of them is to do with professions having been fundamentally about judgement in the past, and judgement is being overtaken by technology and algorithms. Unless professionals do the sort of things you are saying, which is to make sure that you get rid of your biases, whether it is using Bayesian credibility work or something else, you are going to come up with answers that are not as good as those derived from algorithms. If your answers are not as good, then it does not matter how well you communicate it, there will be others in the market who can see that they can make some money out of it. That is a serious issue.

So, there are others out there who think differently, and if we are not perceived as getting the technical piece right, good enough, and without bias, I do not think it matters how well we communicate, since we will have an issue.

If you have any doubts, look at what is going on in medicine. I have read that the algorithms available now will give better diagnoses than a typical GP. It does not matter how well the GP communicates.

What about audit? There are people saying that audit is going to be totally automated. Is there a profession of audit? I think that long term it is absolutely critical for us as actuaries to position ourselves getting it right in exactly the sorts of ways about which you are talking. That is why I think that your paper is brilliant. If we do not, we are seriously prejudicing the future.

Mr White: In response to a point made by Sejal (Haria), I think that we should mention the impact made by professional and legal risk. Sejal mentioned some actuaries may be reluctant to exercise judgement at all, and mentioned the US as an example. I think fear of litigation may be at the root of this position. If we look at the way in which international accounting standards developed with rules emphasised over judgement and theoretical accuracy as opposed to prudence. What happened was we had a rule that is, now, I think they admit, discredited for recognising asset impairment. You were not allowed to make judgements so you could only recognise an asset impairment if the asset was going wrong. That meant that the concept of Incurred But Not Reported (IBNR) is completely ignored and, by definition, the accounts are all deeply imprudent. When this rule came in, introducing the incurred loss provision in the early 2000s, banks suddenly became profitable, and we know what happened soon after that.

596

So the accounting bodies are now having to change, reluctantly, to a new regime which is called expected loss regime. It is completely disgraceful.

We have to be careful as a profession not to allow ourselves to be biased.

On Michael Tripp's point about investment, it occurred to me, as the presentation was being given, that making judgements is exactly how you should go about investment analysis.

Mr Tredger: I will come back on a couple of points which have been raised about big data, algorithms and whether that means the end of professions. Is this the end of actuaries?

A couple of observations immediately sprang to mind. One was an actuary, who was infamously subject to a lot of attack in the Press following the 2008 financial crisis because he came up with an algorithm which people claimed bankrupted Freddy Mac and Fannie Mae, because they assumed things were normally distributed or in their copula. There are cautionary tales of where having no common sense can lead us far wrong.

The other point about the long term in the profession is we do know some things; we are not saying we do not. We are not saying we do not have algorithms or methods, because we do. The problem is how do you extrapolate from what you do know to what you do not know? That reminds me of a quote from the Duke of Wellington: he was saying that war, a lot like life in general, revolves around trying to take what you do know and applying it to what you do not know. What is over the next hill? That was his problem. He saw that as a problem for life in general.

We are trying to say what we know. What can we see over the next hill? We are not going to come up with an algorithm that solves that problem. The solution has to involve a certain amount of psychology and a certain amount of group consensus. The algorithms are our friends as long as we use common sense.

Mr D. Simmonds: One term that has been frequently used in this discussion, in the paper and in many other areas, is expert judgement. It is a term that has troubled me because who are the experts?

To my mind, many actuaries and many people in other professions, in order to do the job that they do, in order to provide the service that their clients want, need to make judgements about items which are highly uncertain. That is the key of the job.

In order to manage and mitigate that uncertainty, yes, you need to bring in what expertise is available to you, and, yes, you use some of the very interesting techniques that have been mentioned in the paper. But, it depends on the actuary's job. There is a big variety. I think we are fairly diverse; I think we are more diverse than some might think.

Some of us in this profession will need to make more judgements about uncertain things than others. I would say that this paper is very much aimed at them as a wider audience. But I should like us to be thinking about how we can make judgements about things that are highly uncertain and actually have as our focus that expertise is one of the tools we need to do that job.

Mr Spain: Just picking up on Ed (Tredger)'s point about looking at what we can see and then guessing what is over the hill, the trouble is sometimes what we are looking at we are not seeing.

597

We do not recognise what is there. We do not recognise what is not there. 2008 is a perfect example. Looking at assets at market value for long-term entities is another example. We cannot see what is there. We cannot see what is not there.

We should not kid ourselves that we project things into the future and we can tell people what is going to happen because we cannot do so. We can only give them some idea of what might happen and which way to hedge their bets and hope that we might get it right.

The other thing is, I cannot actually recall seeing the word "arrogance" in the paper. It might be there. But I can think of one very senior pensions actuary who, about 15 years ago, told me quite confidently that one piece of pensions tax legislation had been repealed by something completely different. He was utterly wrong. But he said it with such complete confidence that a couple of people started believing him, which was not good.

Ms Haria: I think that two very valid points have been raised. Yes, we do have competition from big data. But how much of it is information and how much of it is noise? Can we separate signal from noise?

This is where I see actuaries could add significant value if we do things properly. We are familiar with the concepts of judgement. Big data is only starting to get their heads around those concepts. We can steal a march here. It is very important that we think about how we improve robustness around that area.

I think there is a genuine point about humility. Whenever we look at something, we will inevitably have a prior preconception of how we expect things to work. But we need to be sufficiently able to react to change our minds, not to do that in a negative sense, but to be adaptable and recognise what we are actually seeing.

Mr A. J. Newman, F.I.A.: Regarding the predictions of Susskind and Susskind about all professions going by the board, I know that people have been saying that about actuaries ever since I started, which is a little while back now. "Once you have computers, what do you need actuaries for?" People have been saying that for years. What happens is that the actuaries then use the computers, and I think the same is going to be true of the algorithms. The actuaries will use the algorithms. As people have said before, they will know when to overlay that with expert judgement and, dare I say, communicate the results.

If we stick our heads in the sand and ignore them, then they might well do us in; but if we embrace them, which I think is the currently accepted political term, then I think that the future is bright.

Ms Scullion: Coming back on that point, I completely agree. Some say that big data might overtake us, when we can see this as an opportunity in this type of subject area. As we obtain more data and a greater ability to use it, then that could become quite a significant advantage.

Dr Lo: There were two examples that were in my mind when I was listening to the comments. One of them was about Galaxy Zoo. Basically, there are many pictures of galaxies that astronomers need to classify. There are different types of galaxies. I am no astronomer, so I do not know what they are. They need to classify all of them. They have been building algorithms, but none of them really work. In the end, NASA came up with the idea

598

of asking everyone, the public, to classify these galaxies for us. The result is an outstanding success. There are a few peer reviewed papers of that example. It is still ongoing because there are many, many pictures.

There is something in the human mind that can recognise patterns.

The other example, from when I used to play chess, was following Deep Blue versus Garry Kasparov. He won game after game after game versus Deep Blue. The kind of brain that we have is amazing. It might be, in the long term, that we could be hit, but I believe that long term is probably going to be very long term. We still need to add value by contributing to our subject.

I guess, adding to the discussion on where the profession will be many decades in the future, we are supposed to serve the public interest. If we believe that we are no longer adding value because we are overtaken by algorithms, etc., then what is wrong with there being no profession? That is slightly controversial, but we need to take that into account as well. If we are no longer serving the public interest, then let us do something else.

Ms Haria: Let us not go there. Let us make judgements and add value.

The Chairman: I will end by thanking everybody from the floor for your contributions, your thoughts and your questions. And thank you very much to the panel.