Biases in forecasts and being certain about uncertainty – why we need social scientists in meteorology

helen.roberts@metoffice.gov.uk, Helen Roberts

When I joined the Met Office twenty years ago as a trainee weather forecaster, the prerequisites were a strong background in physics and maths, and the in-house training gave us a good understanding of the dynamics of the atmosphere. The emphasis was very much on the physical and natural sciences. My career in operational meteorology took to me to a variety of locations, working with a multitude of customers from numerous industry sectors, including military, retail, aviation, sport, energy, construction and media.

The question we were aiming to answer for our customers was 'what is the forecast?' However, as the years progressed so the question subtly changed, what people really wanted to know is what the weather will *do*, rather than what it will *be*. This required a shift from forecasting the weather to forecasting the impact of the weather. But even this does not give the full picture, what we really want to do is neatly summarised in our Met Office purpose, which is 'helping you make better decisions to stay safe and thrive'. This statement of intent puts *people* at the heart of our raison d'etre. When we produce a weather forecast, we're not doing it just for the sake of predicting a future state of the atmosphere, we're doing it because weather impacts us. It determines when we put our washing out, whether we wear a coat, our hobbies, our businesses and their operations, and our safety.

So, we need to use the science of people, something we have been rather slow to acknowledge. But there is now a growing recognition that the social and behavioural sciences are as important as the physical sciences that are more traditionally associated with meteorology. Social and behavioural science can offer insight all the way through the forecast process. Our operational meteorologists have vast amounts of data at their fingertips. How do they decide which data to look at, how to interpret probabilistic data, and what to do when data sources are conflicting? Are there human biases in the process? Answering these questions requires an understanding of decision science and cognition.

Cognitive biases explain ways in which human behaviour differs from rationalism, often in common and predictable ways. Why don't people take preparatory action, for example if they live in a flood risk area? This is called hyperbolic discounting, where people tend to prioritise immediate benefits over bigger future gains. Hick's Law tells us that more options lead to harder decisions, so weather warning advisory action statements must be few, clear and easy. People tune out to things they are repeatedly exposed to, named banner blindness, this is the danger of over-warning. People are more likely to take an action when the effort is small, termed the spark effect or principle of least effort, so in a weather warning we can highlight the easy and free actions first. Social norms mean people adapt their behaviour based on what others do, so if people around them are not following an evacuation order, then they likely won't either. And availability heuristic whereby people favour recent and available information over past information, so someone who has been recently flooded is more likely to take heed of a flood warning.

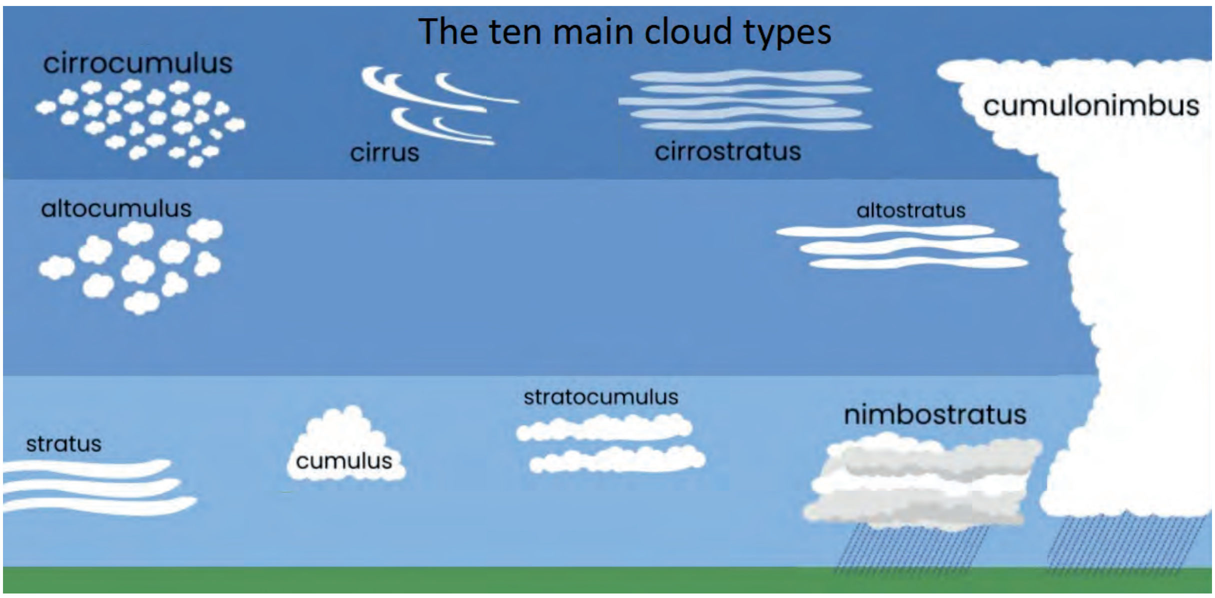
This summer, I took some experiments on the road. This included five Met Office Services Roadshows and the British Science Festival, at which I ran interactive experiments with various audiences to test contextual cognitive biases. Experiment #1 investigates *hyperbolic discounting, or present bias.* Participants are shown **figure 1**, they are told they have (hypothetically) won a holiday and are leaving today. They are at the airport and given the choice of two holiday locations. Both are five-star resorts. Holiday A will look much like the picture on the left when they arrive, with heavy rain today but becoming increasingly sunny through the week. Holiday B will look much like the right-hand picture on arrival, with lots of sunshine, but will gradually turn cloudier with rain by days 4 and 5. They are then asked which holiday they would choose?

▲ *Figure 1: Image used to test hyperbolic discounting.*

My hypothesis was that more people would choose holiday B, taking the sunshine sooner option, despite the fact that they get more sunshine with the first option.

Experiment #2 *tests availability* bias, whereby people draw on recently or easily available information to make decisions. Participants are divided into two groups; group one are asked to draw a cloud, group two are given a graphic showing all main cloud types (**figure 2**) and also asked to draw a cloud. My hypothesis was that group one would tend to draw a cumulus, with group two showing a wider variety of cloud types in their drawings.

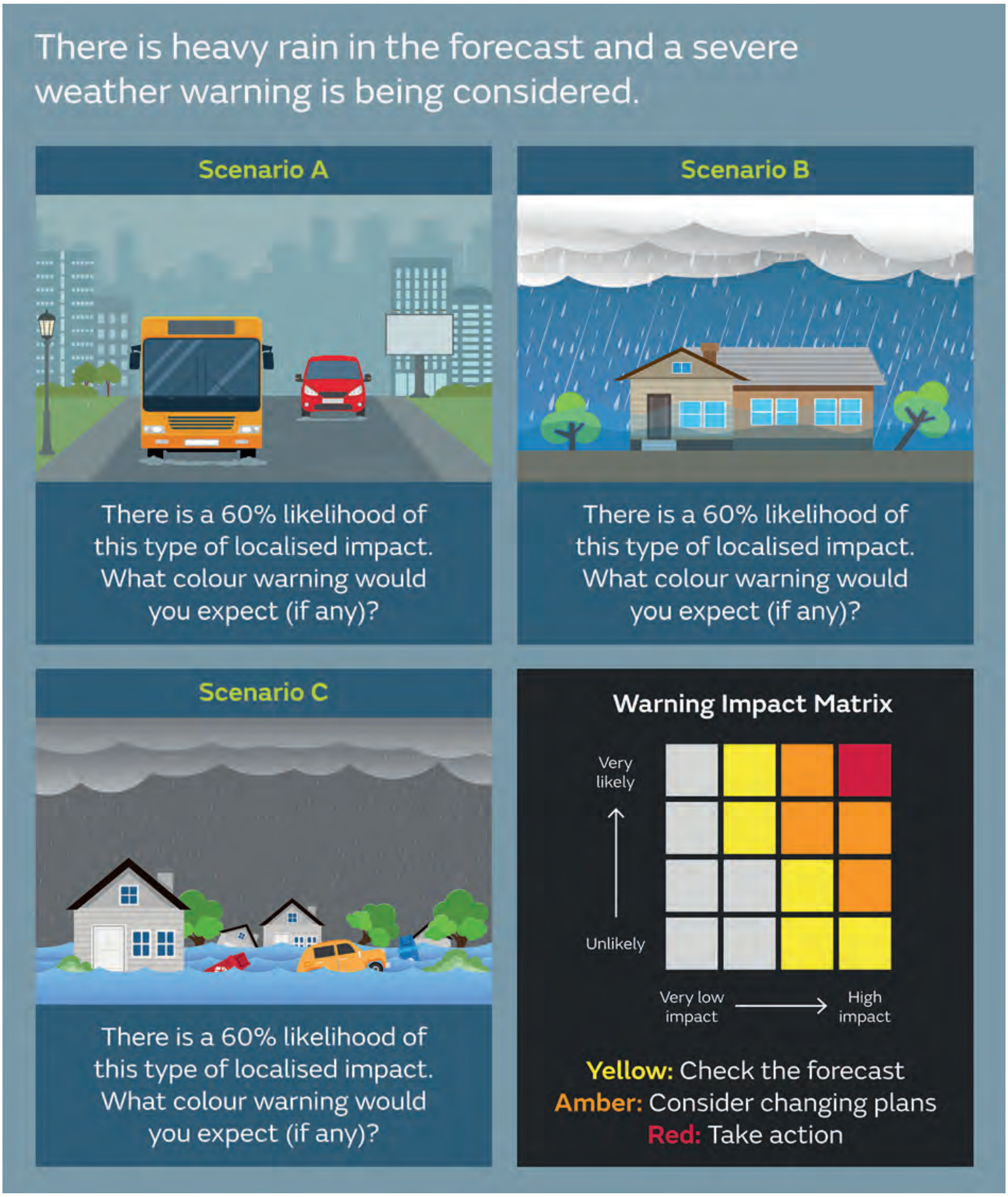
The *severity effect* describes the tendency for people to implicitly interpret probability expressions as more likely when they describe more severe or undesirable outcomes. Someone who



▲ *Figure 2: Image used to test availability heuristic.*

interprets a 'slight chance' of showers to mean a 1%–5% chance will likely interpret a 'slight chance' of a hurricane to mean something closer to a 10%–15% chance (Ripberger et al., 2022). This was tested by showing pictures of impacts varying in severity, accompanied with a probability to keep the base rate the same, and asking people what colour warning (if any) they would expect (figure 3). My hypothesis was that people would conflate severity with likelihood, creating a skew towards amber/red for impactful images regardless of likelihood. (Results will be analysed late 2023).

These are just a few examples of psychological biases, which aren't a problem in themselves, and fortunately we humans are quite consistent in our divergences from rationalism. So as long as we are aware of the biases, we can work around them and even use them to our advantage.



▲ *Figure 3: Image used to test severity bias.*

Behavioural insights research shows that advice should be listed easiest and cheapest first and be specific and actionable. It also tells us that in the case of heat warnings (as opposed to other weather parameters such as wind, snow, rain, fog etc) that people are more likely to act on behalf of someone else, so framing the advice in terms of helping vulnerable relatives and neighbours is most efficacious. This has the additional benefit that once people have taken this action, such as ensuring an elderly relative keeps their curtains closed during the day, drinks plenty of water, and doesn't go outside during peak daytime heating, they are then more likely to take that advice on board themselves.

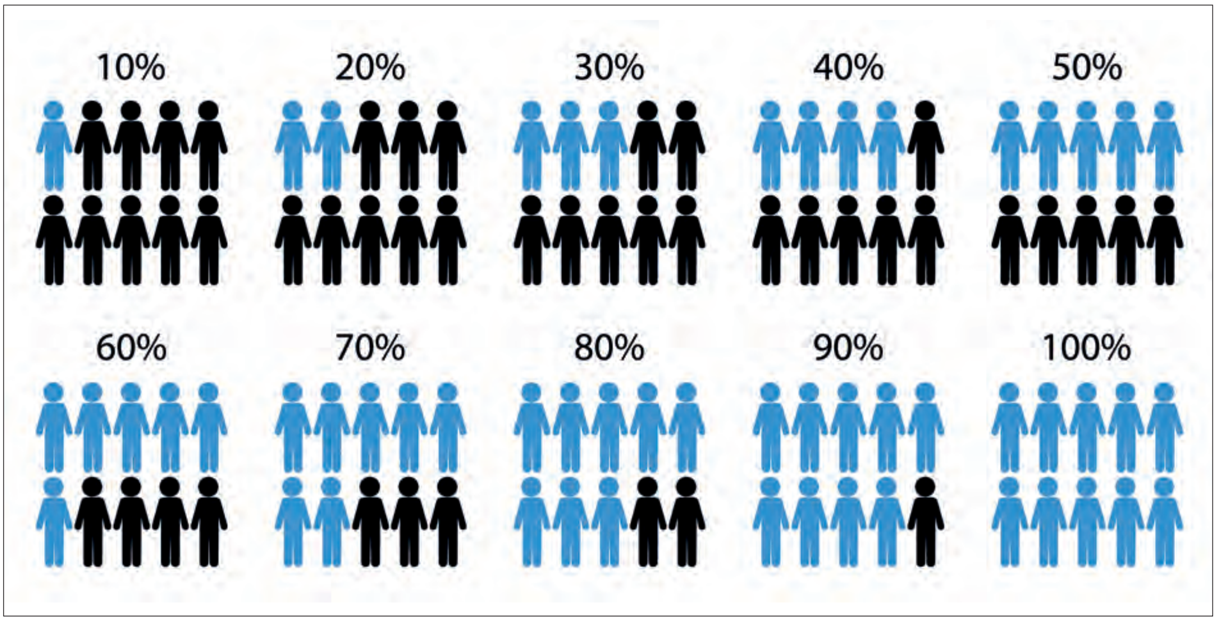
As the meteorological community continues to shift from deterministic numerical weather prediction to probabilistic solutions using ensembles, we continue to grapple with the most effective ways to communicate and visualise uncertainty. Evidence shows that meteorologists hugely underestimate the public's ability to understand and use probabilistic information. In fact, most people intuitively infer uncertainty even when given a deterministic forecast (Savelli and Joslyn, 2012), and as long as the information is presented in an effective way probabilistic information greatly improves decision-making, leads to greater trust and more understanding of forecast information (Ripberger et al., 2020). So it is a win-win. However, those caveats around the method of communication and visualisation are important.

Directionality can influence perception, positive statements that focus the probability that an event *will* happen, such as 'it is possible that the storm will affect town x' can cause people to overestimate the baseline probability of an event, whereas negative statements that focus on the probability that it *won't* happen - 'it is likely that the storm will miss town x' - can cause people to underestimate the likelihood of an event (Honda and Yamagishi, 2017). The trend effect is related to anchoring bias, whereby people are heavily weighted to the first piece of information they see. In the case of the trend effect it can mean people often interpret recent forecasts in light of past forecasts, so a 'moderate' risk, for instance, may cause more worry if it has been *upgraded* from a 'low' risk than if it has been *downgraded* from a 'high' risk (Hohle and Teigen 2015).

When it comes to expressing uncertainty using words and phrases, the literature clearly shows that it is always preferable to include a numeric 'translation' for any verbal probability expressions used, as words such as *possible, likely* etc are very subjective. And ideally, the number should be situated close to, or instead of, the verbal expression, so rather than 'thunderstorms are possible this evening', a better expression would be 'there is a 30% chance of thunderstorms this evening' (Wintle et al., 2019).

Regarding numerical representations of probabilities, the research shows that a simple percentage is most easily understood. Caution should be used against '1-in-x' formats, this is true both for this

▼ *Figure 4: Icon array showing proportion of a population*



context of a weather forecast (Ripberger et al., 2022), but also for framing events in relation to climatology such as a 1-in-100 year flood event, which has also been shown to cause confusion and possible assumptions that if a 1-in-100 year flood happened last year, it won't occur this year (or on fact for another 99 years), which as we know is incorrect. Finally, and perhaps most importantly, what does the research tell us about visualisation of probabilistic information? Unfortunately, the main takeaway is there is no 'best way', one size does not fit all (Ripberger et al., 2022). However, we do know that clarity and simplicity are key, we must not overwhelm the user with cluttered displays. Visualisations should be tested with the relevant audience before rolling out (Ripberger et al., 2022), and they must have explanatory labels and descriptions to avoid ambiguity (Okan et al. 2015).

Gist information is useful, such as icon arrays which provide more transparent representations of risk that generally promote higher comprehension (Dieckmann, Peters and Gregory, 2015). These are not often used in weather but are more common in the medical sector (**figure 4**). There is no reason why we could not find innovative ways to use icon arrays. Pie charts and bar graphs seem to be fairly well understood by the general public as well (Ripberger *et al.*, 2022).

Ensemble or simulation representations (**figure 5**) promote risk comprehension and awareness of unlikely (but possible) outcomes, but they may distract some people from scenarios that forecasters believe are most likely. Therefore, could be useful for high-impact events but less so for more routine weather (Padilla et al., 2017).

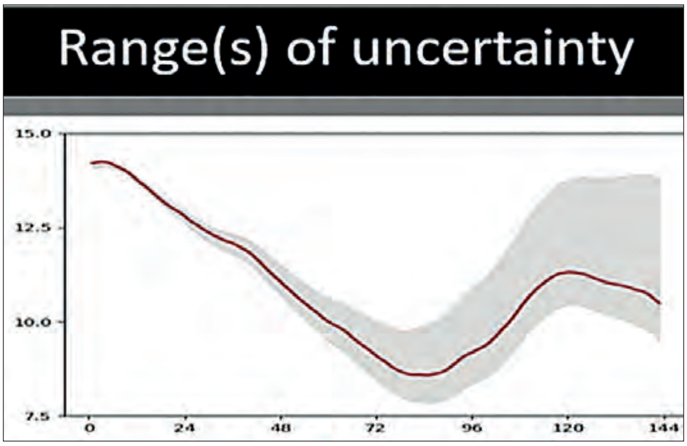
There is some limited evidence that polychromatic schemes work better than monochromatic, and warm colours indicate more risk than cool colours (Klockow-McClain et al. 2020), which makes sense at least in the western world where red is regarded as the colour of danger.

That is the theory, but my role is about putting theory into practice, using these insights and doing our own research to inform improved products and services for our customers. As part of our summer testbed, I was able to compare the decision processes of operational meteorologists who had access to either deterministic model data only, probabilistic (ensemble) only, or both, the results from these experiments will inform our move to ensemble only within the next few years, ensuring a smooth transition for our Operational Meteorologists as well as our customers.

I created a post-event analysis template that can be used after an impactful weather event to summarise the key lessons learned, with a particular focus on decision-making and the behavioural and societal response, which alongside our newly developed climate context documents will ensure we evolve our processes and ways of working so that we and our customers can better plan, prepare for and respond to extreme weather events.

Another key aspect of my job is to raise awareness across our organisation, and the meteorological sector more broadly, of the importance of the social sciences. I am using a multifaceted approach to this, including running a social science community of practice at the Met Office which has been in place for over two years now and has been extremely successful at encouraging engagement with the topic, as well as knowledge sharing and learning. This community is hosting a conference to promote the importance of integrating the social and physical sciences in the environmental sectors and the effective use of the social sciences in meteorology and climatology.

I could go on, but I will leave you to think about additional ways in which social and behavioural science can augment the physical sciences. Suffice to say, this is a rapidly evolving area that combines my three great passions of weather, social science and communication, and I'm delighted to be able to call myself the UK's first Socio-Meteorologist.



◄ *Figure 5: Range of uncertainty visualisation using ensemble data.*

References

Dieckmann, N. F., Peters, E. and Gregory, R. (2015) 'At home on the range? Lay interpretations of numerical uncertainty ranges'. *Risk Analysis*, 35, 1281–1295. DOI: 10.1111/risa.12358.

Hohle, S. M., and Teigen, K. H. (2015). 'Forecasting forecasts: The trend effect'. *Judgment and Decision Making, 10*(5), 416–428. DOI: 10.1017/S1930297500005568

Honda, H., and Yamagishi, K. (2017) 'Communicative functions of directional verbal probabilities: Speaker's choice, listener's inference, and reference points'. *Quarterly Journal of Experimental Psychology*, 70, 2141–2158. DOI: 10.1080/17470218.2016.1225779.

Klockow-McClain, K. E., McPherson, R. A. and Thomas, R. P. (2020) 'Cartographic design for improved decision making: Trade-offs in uncertainty visualization for tornado threats'. Annals of the *American Association of Geographers*, 110, 314–333. DOI: 10.1080/24694452.2019.1602467.

Okan, Y., Garcia-Retamero, R., Cokely, E. T. and Maldonado, A. (2015) 'Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays'. J*ournal of Experimental Psychology Applied*, 21, 178–194. DOI: 10.1037/xap0000045.

Padilla, L. M., Ruginski, I. T. and Creem-Regehr, S. H. (2017) 'Effects of ensemble and summary displays on interpretations of geospatial uncertainty data'. *Cognitive Research: Principles and Applications*, 2, 40, https://doi.org/10.1186/s41235-017-0076-1.

Ripberger, J., Bell, A., Fox, A., Forney, A., Livingston, W., Gaddie, C., Silva, C. and Jenkins-Smith, H. (2022) 'Communicating Probability Information in Weather Forecasts: Findings and Recommendations from a Living Systematic Review of the Research Literature'. *Weather, Climate and Society*, 14, 481–498, https://doi.org/10.1175/WCAS-D-21-0034.1.

Savelli, S., and Joslyn, S. (2012) 'Boater safety: Communicating weather forecast information to high-stakes end users'. *Weather, Climate and Society*, 4, 7–19. DOI: 10.1175/WCAS-D-11-00025.1.

Wintle, B. C., Fraser, H., Wills, B. C., Nicholson, A. E. and Fidler, F. (2019) 'Verbal probabilities: Very likely to be somewhat more confusing than numbers'. *PLOS ONE*, 14, e0213522, DOI: 10.1371/journal.pone.0213522.