Safety Science xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

# Forecasts or fortune-telling: When are expert judgements of safety risk valid?

Andrew Rae [a,*], Rob Alexander [b]

[a] Griffith University, Australia
[b] University of York, United Kingdom

ARTICLE INFO

*Article history:*
Received 19 November 2016
Received in revised form 13 February 2017
Accepted 25 February 2017
Available online xxxx

*Keywords:*
Expert opinion
Risk assessment
Validation
Forecasting
Ontology of risk

ABSTRACT

Safety analysis frequently relies on human estimates of the likelihood of specific events. For this purpose, the opinions of experts are given greater weight than the opinions of non-experts. Combinations of individual judgements are given greater weight than judgements made by a lone expert. Various authors advocate specific techniques for eliciting and combining these judgements. All of these factors – the use of experts, the use of multiple opinions, and the use of elicitation and combination techniques – serve to increase subjective confidence in the safety analysis. But is this confidence justified? Do the factors increase the actual validity of the analysis in proportion to the increase in subjective confidence?

In this paper, by means of a critical synthesis of evidence from multiple disciplines, we argue that it is plausible that expert judgement deserves special standing, but only for well understood local causal mechanisms. We also conclude that expert judgements can be improved by using appropriate elicitation techniques, including by combining judgement from multiple experts. There is, however, no evidence to suggest that fuzzy, neural network, or any other form of complicated processing of expert judgement has any advantage over simple combination mechanisms.



## 1. Why does expert judgement validity matter?

Would you trust a panel of government risk experts who told you that it was safe to build a nuclear waste processing plant in your neighbourhood? How about an international community of scientists predicting climate change? How about a single engineer telling you not to cross a bridge, because their calculations suggested it was unsafe?

Safety analysis has always, to a greater or lesser extent, relied on the opinions of experts. Individuals with specialist domain knowledge, or with superior understanding of risk analysis, are called upon to determine the nature, size, and acceptability of risk. Risk estimates produced by experts are more believable, but this does not necessarily make them more correct.

Our discussion in this paper is concerned with the use of experts for estimating the risk of major accident events. Unlike some risk problems such as population health, where there is a substantial body of recent data on which to base projections, major accidents occur too infrequently for past statistics to be a good indicator of risk.

It is in these situations that expert judgement is most necessary, but also most questionable. A clear understanding of the validity of expert judgements, and of how their validity is influenced by methods of elicitation and combination is essential for good safety practice. It is also important to be able to draw a clear distinction between expert estimates and value judgements. Experts should demand a role in decision-making only to the extent that they have something offer, not because their status confers special privileges.

There is an increasing trend to make use of multiple expert opinions in safety analysis, and to formalise the way these estimates are used. This involves documented methods for how opinions are elicited, how they are combined, and how they are integrated with other facets of the analysis. The trend is manifest in the academic literature - for recent illustrative examples see Zhou et al. (2017), Forteza et al. (2016), and Kokangül et al., 2017 – and in regulatory guidance (Boring et al., 2005).

Practices for forecasting using expert judgement have been heavily studied outside safety science. In particular, there has been extensive work within social psychology and management focussing on group decision-making, and within economics focussing on the mathematics of combining individual probability estimates. There is also a body of large-scale experimental work using prediction markets and competitive forecasting. A lot is known about

* Corresponding author.
*E-mail address:* d.rae@griffith.edu.au (A. Rae).

http://dx.doi.org/10.1016/j.ssci.2017.02.018
0925-7535/

expert forecasting, but little of this knowledge is employed in safety practice.

In writing this paper we have been motivated by the proliferation of complicated techniques in the academic safety literature for eliciting and combining expert judgments. Of particular concern are papers that make definitive claims about the size and nature of risk based on these methods. Such research takes an unequivocally realist position on the nature of risk, whilst making unwarranted assumptions about the validity of the methods used. For example:

- That the performance of a safety management system has improved
- That human factors make a greater contribution to coal mining accidents than other safety issues
- That there is a particular ordered ranking of risks for cargo ships
- That particular geographic locations are more dangerous than other locations
- That specific companies are safer than other companies[1]

Frankly, we would like this researcher behaviour to stop. Armstrong suggests that the Golden Rule of forecasting is *"be conservative by adhering to cumulative knowledge about the situation and about forecasting methods*" (Armstrong et al., 2015). In other words, forecasts should take into account what is known about forecasting itself, not just what is known about the problem at hand.

There are two questions that must be answered before expert opinion can be used to make definitive claims about safety risk:

1. What can be currently claimed about the validity of expert estimates as data for the purpose of safety risk estimation?
2. Under what circumstances, and to what extent, do methods for elicitation and combination of expert estimates of safety risk improve their validity?

## 2. Is there such a thing as a "risk estimation expert?

### 2.1. "Expert" is a very ambiguous term for risk assessment and analysis

Predicting the future is a fundamental element of carnival fortune telling, sports betting, religious prophecy, and financial planning. Some types of prediction can be trusted, and others are cannot. Some people are better at making predictions. What does it mean to be an "expert" at predicting the future?

There are two main definitions of experts for the purpose of forecasting.

1. An expert is someone whose judgement is accorded extra weight, due to their qualifications, experience, and other signals of authority (Farrington-Darby and Wilson, 2006).
2. An expert is someone who makes especially accurate forecasts (Mellers et al., 2015).

Each of these descriptions is, in its own way, a fair summary. Which definition applies for risk assessment and analysis depends on how exactly risk is understood.

The realist view (Smith, 2004) maintains that risk is a real, objective and quantifiable truth. The likelihood of an event in the future becomes the frequency of that event with the benefit of hindsight. Whilst we cannot know for certain how accurate risk estimates are at the time they are made, their accuracy may (at least in principle) be knowable at some point in the future.

In contrast, the phenomenological tradition, as explained by Rosa (1998), holds that even if objective risk exists as an abstract idea, there is no way to separate objective risk from our subjective and constructed experience of risk.

Very few researchers or practitioners argue that risk is entirely objective or entirely constructive – strict realism and strict phenomenology are extremes on a theoretical continuum. However, an inclination towards one paradigm or the other determines what is knowable about risk, and therefore what can be "valid".

Most practical risk assessment is conducted from a generally realist perspective, whilst acknowledging that some degree of uncertainty is inevitable. Estimating risk, under this perspective, is analogous to guessing the number of marbles in a jar. The estimate is subjective, and the true number may never be known, but it is still possible to make statements about the objective goodness of the estimate. Goodness encompasses accuracy, certainty, and calibration.

An estimate is more "accurate" if it is closer to the true number. For example, if there are 250 marbles, an estimate of 240 is more accurate than an estimate of 230.

An estimate is more "certain" if it provides a narrower range of values. For example, an estimator might say "90% of the time, the true number of marbles will be between 240 and 260". An estimator would be overconfident if statements of this type were correct less than 90% of the time, and under-confident if they were correct more than 90% of the time.

There is no accepted term for correctness of certainty. We will use "calibration"; an estimate is calibrated if it is neither under confident nor over confident. It is better for an estimate to be more certain rather than less, but only if it is also well calibrated.

The applicability of "accuracy", "certainty" and "calibration" obviously depend on how risk is described. Not all descriptions of risk involve quantification (Kristensen et al., 2006), and not all quantified risk includes separate assessment of certainty.

Not everyone agrees that risk estimate validity can be discussed in terms of accuracy, certainty and calibration at all. For those who believe that risk is primarily constructive, risk assessments and analyses are cultural artefacts. They document rather than determine decisions about risk. Validation comes from "justifying the choices made in producing statements about risk" (Goerlandt et al., 2016).

In this paper, following the "pragmatic validity" approach of Rae et al. (2012) and Goerlandt et al. (2016), we will evaluate claims about expertise in terms of the ontology used by the people who are making those claims. If risk assessments and risk analyses are being used for the sole purpose of explaining how decisions have been reached – that is, their makers are not intending to make objective statements about risk – then "accuracy" is not a meaningful dimension. These analyses should be validated based on a constructivist understanding. However, when risk estimates are attempts to describe risk as a real objective phenomenon - as they are whenever risk estimates are an input into decision making about further risk treatment – the risk estimates must provide a good description of the thing they purport to measure (Rae et al., 2012).

The combination of realist ontology and the use of experts requires a link between the two definitions of expertise. Experts should be a group of people whose opinions are deserving of extra weight because those opinions can be expected to be some combination of more accurate, more confident, and better calibrated.

Does such a group exist?

There are several plausible ways in which a potential expert could have a systematic advantage in making forecasts.

The first mechanism – private information – is that they could have access to privileged information held only by experts (Morgan, 2014). In economics, it is commonly assumed that given

[1] t is not our intention to name and shame individual authors, so we have listed here unreferenced examples of recent definitive claims about risk based on processing of expert judgement.

Please cite this article in press as: Rae, A., Alexander, R. Forecasts or fortune-telling: When are expert judgements of safety risk valid?. Safety Sci. (2017), http://dx.doi.org/10.1016/j.ssci.2017.02.018