Heuristic unrealistic

Consultant and trainer Felix Redmill explains how we unconsciously bias our judgements

Risk can provide an effective basis for testing, but gaining an understanding of the risks can be problematic. It is deceptively easy to arrive at an intuitive assessment, but not only is this likely to be far from correct, it also carries no means of verification of its results – or of revealing them to be off the mark. A methodical approach is essential. But even a procedural analysis of risks does not guarantee objective results. It depends on judgemental decision-making, which, in turn, is open to subjectivity introduced by those who carry it out.

Psychological research, much of it initiated by two great experimenters, Kahneman and Tversky, and subsequently substantiated by others, has revealed that our decision-making is subject to biases resulting from the ways in which the human brain handles information. This article offers an overview of some relevant results, which provide insights into what we, as test planners and testers and, indeed, decision-makers in general, might usefully look out for, not only in identifying and analysing risks, but whenever we apply our judgement. It is written in the context of testing and the analysis of risks, but it is relevant in all contexts.

The large volume of information that we receive from disparate sources is almost certain to be too great for a thorough analysis in the time available for most decision-making. Moreover, not only is a thorough analysis likely to be beyond the competence of most brains in all but the simplest situations, but also there are certain to be gaps in the information. So, in order to arrive at decisions at all, humans take mental shortcuts and make approximations. We use mental rules and strategies, referred to as heuristics. Our status on the planet suggests that heuristics have served us well in evolution, but in complex situations they can induce unconscious biases that distort our decision-making processes.

The 'availability heuristic' causes judgements about the likelihood of an event to be based on how readily information that is relevant, or thought to be relevant, comes to mind. Its effect is to increase the judged likelihood if the event is easy to recall or imagine. When the information to hand consists of an individual's own experience, in a purely local context, it can be an accurate basis of judgement. For example, without carrying out formal analysis, a villager may make an accurate estimate of the likely frequency of visitors to the village or of earthquake, storm or flooding, based on observations or recollections of past frequency. But when received information is unrelated to frequency, it can distort judgement. For example, a vivid news item, a recently seen film, or wide media coverage is likely to condition the brain to judge the event's occurrence to be more frequent or likely than it is. Studies show that the public believes murder, which is frequently and dramatically reported in the media, to be more frequent than suicide, which seldom is. In fact the reverse is very much the case.

The availability bias has been explored in the context of egocentricity, and spouses were found to overemphasise their own responsibilities for occurrences (including negative ones such as arguments) within their marriages. Sportsmen were found to ascribe responsibility for the outcomes of games to the actions or inactions of their team-mates rather than to those of members of the opposing team.

Not only those analysing risk, but also software engineers, project managers, test planners and testers, must continually make 'expert' judgements concerning both the probabilities or frequencies of events in absolute terms and the relative likelihood of one event compared with another. The availability bias affects both types of judgement.

Then there is the 'overconfidence bias'. Lack of awareness of the assumptions on which judgements are based, and a failure to appreciate what is not known, lead to this. In one experiment subjects were asked to place odds on their being correct in their judgements of which of two lethal events was more frequent, and, although odds of 100:1 or greater were named more than 25% of the time, about one in eight of the judgements was incorrect.

Experimenters have shown that experts are as prone to overconfidence as lay people, and have cited a number of instances where the bias contributed to accidents. In a study of scientific work they found significant underestimation of errors and the likelihood of errors. In another experiment it was found that psychologists' confidence in their supposed knowledge and understanding of their patients, as information about them increased, was not a sure sign of increasing accuracy of their conclusions. Their certainty about their own decisions became entirely out of proportion to the actual correctness of the decisions themselves. Experiments have shown too that the majority of drivers believe themselves to be above average. And most of us will have recognised that designers are over confident of the correctness and efficacy of their designs and programmers of the faultlessness of their code. No doubt test planners are over confident of the effectiveness of their plans.

A feature of overconfidence is that it encourages a belief that the overconfident person's judgement is infallible and that advice from another is unnecessary. But experts' trust in their estimates is often a function of how much information they have gathered rather than of its accuracy or predictive success or suitability. We make judgements about the likelihood of bugs in a system, and of failures of the system in operation, based on test results, often without consideration of the nature or coverage of the testing or of whether or not the test environment was representative of operational conditions. In planning testing or interpreting its results, it would be useful to be aware of our biases and how they could affect our judgements. But then, the overconfidence bias mitigates against our doing this!

Another heuristic is 'representativeness', which is defined as a subjective judgement of the extent to which something (eg an event, the traits of a person) reflects the salient features of its parent population or the process by which it is generated. For example, a small, statistically unrepresentative sample may be assumed to be representative of its large parent population. Statistics suggest, and experiment confirms, that if a large number of balls is drawn from a well-shaken barrel of red and white balls, the proportion of red to white in the balls drawn is likely to be highly representative of that in the original population in the barrel. But we intuitively make the same assumption – of representativeness (possession of all the attributes) of its parent population – on the basis of a small sample. But, as many gamblers know, or should know, the law of large numbers does not hold for small numbers. Experiment has shown that when provided with information (the input) about, for example, a person or situation, we make predictions about the person or situation by selecting outcomes that appear to be most representative of the input – even though the input may only be partially descriptive of the person or situation. In this respect our intuition is biased and the assumption can lead to considerable inaccuracy.

The three biases already mentioned affect decision-making in general, and in particular they affect judgements of likelihood and frequency. They are all likely to be prejudicial in our estimates of, for example, failure rates and bug densities. A consideration of risks also depends crucially on the identification of the things that could go wrong (the hazards) and give rise to them. A factor influential in the effective identification of hazards is our tendency to take the 'inside view'. This causes us to treat the case in hand as unique and to neglect or reject the aspects that relate it to other cases of the same type.

As a project manager I knew that I should seek information on the successes and failures of other projects, so as to derive lessons from them, but often I was too busy with my own project to take the 'outside view'. Taking the outside view leads us to ask such questions as: 'What happened on the last occasion that we did something like this?' and 'Has anyone else done something like this and, if so, what happened?' By taking the inside view, we fail to consider, or even to recognise, relevant external information, so we neglect the lessons that we might have learned and the experience that could be useful. We are likely to be over-confident in our plans and overemphasise their virtues, thus overlooking their weaknesses. We tend to perceive our own problems as special and unrepresentative when we would benefit by seeing them as examples of a wider class.

A procedural way of neutralising the inside view (and other biases) is for a team rather than an individual to engage in the activity in question (eg hazard identification or Fagan's inspection). However, the team needs to be carefully chosen. Members must have different experience, responsibilities, and perspectives, for they need to complement rather than reinforce each other's perspectives. Ironically, when its members have similar outlooks, a team is likely to suffer from 'groupthink' and is likely to strengthen the conviction that its inside view is both correct and good.

Kahneman and Tversky proposed that, rather than leaping directly from problem to decision, humans have a two-stage decision-making process. The first stage is concerned with analysing, editing, and perhaps reframing the problem and the possible acts and outcomes, and the second stage with evaluating the options and arriving at a decision. Decisions depend heavily on how problem information is expressed, or 'framed', in the first stage. Experimenters framed the possible outcomes of two options for a public health programme, first in terms of lives saved, and then in terms of lives lost, and, by doing so, achieved a complete reversal of opinion on what action should be taken.

In another experiment, in which frames were considered to be communicative rather than cognitive constructs, records of situations in which councillors sought to achieve conflict resolution between disputants were examined. It was found that a disputant would make a choice of language (frame the issue, though not necessarily consciously) so as to highlight certain aspects of the communication, while ignoring others. The other disputant would then do the same, perhaps ignoring the first disputant's frame. In each case the choice of language provided a verbal cue to the other participant, who could respond by converging or diverging from the other's chosen frame. The researchers found a positive relationship between the convergence of frames and the reaching of agreement.

In another experiment, two groups, selected to be different, had to choose which of two election candidates to vote for, the information presented to the groups consisting of simulated news stories about the candidates' views, using 'ethical' and 'material' frames. The framing was found to have a pronounced influence on the interpretation of the issue in hand, and also a 'priming effect' on how the voters perceived other issues. If ethical framing was activated, it fostered an ethical interpretation of other issues - for both groups of subjects. Material framing had a corresponding effect. Further, voters were likely to put the frame at the centre of their evaluation for decision-making. Thus, framing was not only influential on the subject's interpretation of the particular issue, but also on subsequent judgements.

Framing is not merely a theoretical concept. Advertisers understand it well. And the CORE (controlled requirements expression) technique recognised that requirements capture demands not one but several frames of reference (perspectives), and provided a defined set of procedures for identifying relevant stakeholders and acquiring information from them.

Framing is crucial in the presentation and communication of risk information, and the framing is defined, consciously or unconsciously, by the information's source. For example, someone who stands to gain from the acceptance of a risk is likely to frame the risk information differently from a person who opposes the risk. We need to be alert to the possibility of such biases. Similarly, in preparing risk information, or any information, for decision makers, we should be aware of how influential our structuring and communication of it can be.

In summary, because of the shortcuts that the brain takes in order to prevent information overload and cognitive strain, intuitive decision-making can be severely flawed in all but the simplest cases. Mental heuristics are a normal aspect of human reasoning processes, and serve us well, but they can reduce problems to unrealistic oversimplifications and thus turn into biases that distort our judgement.

Experts, like lay people, depend on heuristics and are therefore subject to biases, particularly when they function outside of their fields of specialist knowledge. Testers, test planners, project managers, and other software-based professionals step outside of their specialities when they seek to identify and analyse risks and to use the results to inform their activities. It would be useful if they not only took tuition in risk and its analysis, but also became aware of the normal human biases that can, and in most cases do, affect their decision-making. With awareness comes the intention to mitigate the effects of the biases. It makes sense to introduce not only means of counteracting known biases but also procedures to check and ensure that they have been applied. An example is the use of a carefully chosen team for assessing relevant information and checking its sources and reviewing decisions. The rub, then, is in translating intention into reality. PT