

## Comment

---

# Optimism and calibration in risk assessment: a note in reply to Hartz and Elrod

J.S. Busby

*International Ecotechnology Research Centre,  
Cranfield University, Bedfordshire, MK43 0AL, UK  
Tel.: +44 1234 754 017; Fax: +44 1234 750 163;  
E-mail: j.s.busby@cranfield.ac.uk*

By way of a reply to Hartz and Elrod's discussion of optimism in risk assessment, we describe some findings of a study into the process of predicting the engineering resources associated with complex, technologically advanced systems in the aerospace industry. These suggest that biases in risk prediction stem both from limitations in individuals' cognitive abilities and from incentive effects. These incentive effects in turn arise partly from conditions in the organisation and partly from conditions in the organisation's environment. Our findings also suggest that incentive effects can lead both to under-estimation of future outcomes and over-estimation. We address the extent to which under-estimation, or optimism, is functional in the organisations in question.

Keywords: Risk, estimation, complexity, engineering, bias



**J.S. Busby** is a lecturer at the International Ecotechnology Research Centre, Cranfield University. He works in the field of engineering design, performance and prediction and currently holds an EPSRC advanced fellowship. He has an MA from Cambridge University, an MSc from London University and a PhD from Lancaster University.

### 1. Introduction

In a recent article in *Human Systems Management* [2] Hartz and Elrod drew attention to the important matter of optimism in human risk assessment. They argued that, potentially, emotion was as impor-

tant as cognition in leading to biased assessments, and that the interaction of the two had been neglected in prior research. They suggested that quantitative research was needed to find out how far emotion explained human choice in the face of risk, and to enhance the prediction of risk.

The article was important both because it pointed to a topic which has been under-emphasised in past research, and because it has considerable practical significance. In our own recent work on estimating the engineering resources needed to develop highly complex, technologically advanced systems, estimates of high calibration and resolution were needed by organisations for two basic functions:

1. *For contract bidding.* To ensure commercial survival, the organisations in question had to avoid loss of business by bidding too high for contracts, and avoid loss of earnings by bidding too low. Good calibration was needed to ensure that, in aggregate, engineering cost estimates were close to outcomes. High resolution was also needed, because contracts were often large and irregular, and significant discrepancies between outcome and estimate on just one occasion could be highly problematic for the businesses.
2. *For giving diagnostic feedback to engineers and engineering managers.* Feedback was given as the discrepancy between outcome and estimate in order to remove the variance in cost outcomes due to factors beyond the control of the engineers (such as product novelty and complexity), but this feedback then confounded the performance of the estimation and engineering tasks. It was important therefore, in order to diagnose problems with engineering tasks, that estimates were based on correct models of the historical relationship between cues such as novelty and complexity, and the criterion variable of outcome cost.

We cannot, unfortunately, offer quantitative evidence of the effects of cognition and emotion on risk estimates in this setting. But we can recount (in outline) some of the qualitative evidence from a piece of research that we conducted into the judgmental aspects of engineering resource prediction in three aerospace firms. All three firms produce complex aerospace systems which are essentially designed to order after winning contracts in a competitive bidding process. The resource predictions are carried out by estimators whose task, in outline, is to:

1. Assign expected values to the resources (mainly cost) consumed in various engineering tasks such as design, technical analysis, systems integration and trials.
2. Assign a confidence interval to these estimated values to reflect the risk that is associated with the activities in question. In two cases this took the form of a three-point estimate of the parameters of a triangular distribution, while in the third case it took the form of assigning probability masses to outcomes that were labelled as 'worst-case', 'work-around case' and 'no-risk' outcomes.

The question of calibration concerns the consistency of these confidence intervals with the distribution of actual outcomes of engineering cost.

## 2. Procedure

The evidence came from the retrospective protocols of cost estimators in the three firms: four individuals in the first, two in the second, and one in the third. In each case the estimator was responsible for the estimate as a whole, but relied on specialists in the engineering disciplines and support functions (such as system trials and technical publications) for predictions of the resources required in specific activities.

The estimators were asked to describe the process by which their estimates had been built up during a recent project, and were asked to use logbooks, project files and schematic drawings to assist their recall. They were prompted before and during their descriptions to explain, for each main estimating judgment and its associated risk assessment, the following:

- their outputs, goals and consequences;
- the informal and documentary information inputs to their judgments;
- their relevant experience and training;

- the feedback they received or expected to receive;
- the rules of thumb they used; and
- the expected performance of the judgment.

The protocols were analysed qualitatively by searching for evidence of bias in the *process* that the estimators followed. We did not have access to the quantitative outcomes of either the estimating or the engineering process (which were confidential) so we had to rely on assessing the behaviours and strategies that people claimed to use.

## 3. Outcomes

On the specific question of how well the estimating process led to well-calibrated judgments, there was evidence both of bias arising from incentive effects and bias arising from cognitive effects. The *incentive* effects were these:

1. There was a highly asymmetrical treatment by the organisation of discrepancies between outcome and estimate. Engineers believed that significant negative consequences were incurred when a cost outcome was substantially greater than the estimate, whereas very mild negative consequences, or even positive ones, were incurred by the individual when the outcome was substantially less than the estimate. It was perceived that it was better to be responsible for a project which ended up within-budget and within-schedule than one that did not. There was thus an incentive for the specialists contributing cost predictions to over-estimate resource consumption: that is, to demonstrate apparent pessimism. In one of the firms, the estimator who gave the protocol believed that different specialists were pessimistic to widely varying degrees, and he contrasted someone he had 'won over' to giving genuine estimates with someone else whom he had 'considerable problems with'.
2. There was a countervailing incentive to underestimate on the part of discipline specialists. This was essentially to do with the desire among specialists such as designers to undertake challenging tasks, and challenging tasks were typically those that involved either considerable complexity (demanding high reasoning abilities in engineers) or novelty (demanding high levels of imagination and broad prior experience). Both complexity and novelty are correlated with

greater cost, however, and specialists therefore had to try to overcome the scepticism of their managers by playing down such costs. The desire for challenging work therefore translates into an incentive to under-estimate costs: that is, to demonstrate apparent optimism.

The notable aspects of these incentive effects were these:

1. They do not have direct origins in individuals' emotions. It is more likely that they stem from individuals' beliefs about the inconsistencies that exist between their personal goals and those of the organisation, and their reasoning about how this inconsistency should influence the estimate they give.
2. Both effects arise because the individual making the prediction of how much resource will be consumed in a task is also the one who will be engaged in the task. In one of the organisations, the estimator argued that this was beneficial, since, if an overspend did occur, the engineer responsible could not attribute it to someone else's under-estimate.
3. In the estimators' protocols, the pessimistic effect (to do with penalties associated with overspending) was in fact much the more salient of the two, suggesting that incentive effects lead mainly to over-estimation, or pessimism, in this particular setting.

There were also biases in the process that seemed to arise from cognitive effects, and these pointed to *under-estimation*:

1. In very much the way that Kahneman and Lovullo [3] argued, there was a striking neglect of distributional evidence and a strong attachment to prediction using singular modes of thinking. The basic procedure that all the companies followed was to break down a project as a whole into work packages, and then plan a series of activities extending across a series of engineering disciplines against each work package. It was at this activity level that expected values and confidence intervals were judged. The problem with this approach, the 'inside view' in Kahneman and Lovullo's terminology, is that it omits activities which cannot specifically be anticipated. Examples of activities that are inherently difficult to anticipate are dealing with unexpected interactions between complex components like micro-

electronic devices, or dealing with phenomena that emerge when aerospace systems enter regions of performance for which there is no prior engineering knowledge. What a distributional or outside view would have demonstrated is that one can predict unanticipated activity of some sort, even if one cannot predict the exact form it will take.

2. The result of this singular mode of estimation was a consistent under-estimation of expected values and excessively tight confidence intervals. In other words, the estimates exhibited optimism in both outcome estimates and probability estimates: outcomes, as engineering costs, were almost always greater than estimates, and fell outside the confidence intervals too often. There appeared to be three underlying explanations, although we had no objective evidence as to how important each of these was.
3. The first explanation was commercial pressure. One of the estimators knew perfectly well that, given historical evidence of unanticipated activity consistently consuming 20% of the outcome cost, new estimates should be inflated by this much to recognise that unanticipated engineering is a fact of life with complex, novel systems. Unfortunately, sales directors saw such statistical evidence as being inadequate evidence. In negotiation with customers, they would remove the 20% factor that the estimators had added.
4. This looks like evidence that sales directors are optimistic, even if estimators are not, but this may not be the sort of chronic, inadvertent optimism that is sometimes ascribed to sales people. The sales director has an incentive to win contracts which are large in revenue, not in profit margin, since his prestige and possibly his remuneration are determined by revenue rather than profitability. For the sales director, bidding low to take on a loss-making contract may be a better outcome than loosing the contract as a result of bidding high. Whether they believe in the distributional evidence or not, sales people thus have an incentive to act as though distributional evidence is invalid. This means acting as though the new project is in some way different from its predecessors, and uniquely able to avoid the problem of unanticipated activity. (One factor that made it easier for sales people to act in this way was the use of the term 'contingencies' for the 20% unanticipated activity reserve that the esti-

mators attempted to build in. The term suggests that the reserve is there just in case something unforeseen happens. Of course, the distributional evidence said that it was virtually *inevitable* that something unforeseen would happen.)

5. Another reason why the organisations preferred singular modes of thinking was the feeling of control it gave their members. The singular approach seems to be more predictive because it allows people to break down the overall task to a level where the elements are within their compass (the control or execution of single individuals). It also gets down to a level where individuals can visualise the process of undertaking these tasks, and in fact the estimators often spoke of picturing how the product would look, or of picturing the operations someone would have to perform during the production of the product. This visualisation in detail seems to contribute to feelings that the task is within the individual's control: that it is insensitive to uncertainties in the task environment. Of course, decomposition is a genuinely helpful strategy in estimation tasks [5], but not if it leads to a bias against distributional evidence. Moreover, in reaching a level at which the individual can visualise concrete actions on his or her part, the estimates are probably also subject to illusion of control effects [4]. The estimator is likely to under-estimate the probability of outcomes significantly different from the expected outcome because he or she under-estimates the role of chance factors in the task environment.
6. A final contributor to this preference for singular thinking seemed to be another bias: that of over-estimating in hindsight what could have been known at the time of the event [1]. In one of the estimators' protocols the individual pointed to a failure to foresee an event that (by the time of the interview) had already transpired and had cost the firm a considerable amount of money. The estimator's diagnosis was that he had not planned carefully enough. He believed that he could have avoided the under-estimate by thinking through the activity in more detail. From what we saw of the problem, however, this was a mis-diagnosis. The unanticipated activity in question had not been encountered before, and in fact the estimator was using an estimating strategy that was new to the company. There was really nothing to suggest that he could have avoided problems

by more detailed planning, except hindsight bias. As a result of hindsight bias, the estimator inferred that the evidence of outcome exceeding estimate was evidence that next time he should plan in still more detail. We would argue that it points to the opposite conclusion. The estimator should plan in less detail and pay more attention to distributional evidence to do with the level of unanticipated activity in historical projects.

To summarise, there was a marked preference for singular modes of estimation, involving detailed planning and a failure to recognise from historical outcomes the virtual inevitability of unanticipated activity. This appeared to have cognitive origins in hindsight bias and an illusion of control, and was reinforced by environmental factors (incentives to win business). The result was optimism in estimates of expected value and confidence intervals, and a resistance to correcting this optimism in the light of experience.

There were in fact some other limitations in the estimators' prediction process. For example, they seemed to be insensitive to sample size when estimating new projects by comparison with old projects. (It did not seem to make a difference to their confidence intervals, for example, whether the set of old, anchor projects consisted of ten projects or just one.) Nonetheless, the most significant effects were those we have just described.

Given that costs were under-estimated, and confidence intervals were too tight, it appeared that overall the organisations' risk assessments were optimistic. The main incentive effect which led to over-estimation (that is, pessimism) was outweighed by the cognitive effect of bias towards the inside view of these complex projects.

#### 4. Conclusion

We have presented some evidence of both optimism and pessimism in the risk assessments of engineering organisations. Some of the effects are due to individuals' cognitions, others to the nature of organisational incentives.

Hartz and Elrod also referred to the question of whether apparent optimism was functional for an organism in adverse conditions. In our study you could argue that optimism *was* functional, at least as far as the businesses were concerned, when trading was

difficult. Winning contracts, even loss making contracts, can tide the firm over short term problems, so under-estimation in such conditions would increase the probability of winning contracts with an acceptable penalty. Of course, this only postpones profitability problems, and ultimately exacerbates them, but this may be tolerable in a cyclical industry. The fact was that the organisations we studied were surviving in difficult conditions, despite consistent under-estimation. Unfortunately, the type of optimism that the companies exhibited was also likely to lead to a bias towards especially novel projects (since these are the projects that the singular approach to estimation will serve worst). From a commercial standpoint this does not seem at all functional.

Our findings are very limited, partly because they are qualitative in nature, partly because they rely on the reports of estimators, partly because they involve small numbers of people. But one has to remember that the estimators we spoke to were, by profession, evaluators of the predictions other people made, and had a great deal of experience in the process of making predictions and the subsequent process of supporting them during contract bidding. One also needs to bear in mind that the effects stem from the setting in which people worked, as much as individuals' dis-

positions, so experimental tests might not be effective in explaining how engineering risk estimates are determined in reality.

What we hope to have demonstrated, however, is that the process of predicting how complex technological systems will perform is one of great practical importance. The extent to which humans making such predictions are influenced by emotion, and by cognition, is therefore something well worth exploring.

## References

- [1] B. Fischhoff, Hindsight not equal to foresight: the effect of outcome knowledge on judgment under uncertainty, *Journal of Experimental Psychology: Human Perception and Performance* **1** (1975), 288–299.
- [2] C. Hartz and R. Elrod, The role of optimism in the probabilistic assessment of risk: a second look at calibration of probabilities, *Human Systems Management* **15** (1996), 79–83.
- [3] D. Kahneman and D. Lovallo, Timid choices and bold forecasts: a cognitive perspective on risk taking, *Management Science* **39** (1993), 17–31.
- [4] E.J. Langer, The illusion of control, *Journal of Personality and Social Psychology* **32** (1975), 311–328.
- [5] D. MacGregor, S. Lichtenstein and P. Slovic, Structuring knowledge retrieval: an analysis of decomposed quantitative judgments, *Organizational Behavior and Human Decision Processes* **42** (1988), 303–323.