

Value of expertise for forecasting decisions in conflicts

Kesten C. Green,* Department of Econometrics and Business Statistics
Monash University, VIC 3800, Australia
E-mail: kesten@kestengreen.com
Phone +61 3 990 52489
Fax +61 3 990 55474

J. Scott Armstrong, The Wharton School, University of Pennsylvania
Philadelphia, PA 19104
E-mail: armstrong@wharton.upenn.edu
Phone 610-622-6480
Fax 215-898-2534

January 2, 2006

Abstract

In important conflicts such as wars and labor–management disputes, people typically rely on experts' judgments to predict the decisions that adversaries will make. We compared the accuracy of 106 forecasts by experts and 169 forecasts by novices about eight real conflicts. The forecasts of experts who used their unaided judgment were little better than those of novices, and neither group's forecasts were much better than simply guessing. The forecasts of experts with more experience were no more accurate than those with less. The experts were nevertheless confident in the accuracy of their forecasts. Speculating that consideration of the relative frequency of decisions across similar conflicts might improve accuracy, we obtained 89 sets of frequencies from novices instructed to assume there were 100 similar situations. Forecasts based on the frequencies were no more accurate than 96 forecasts from novices asked to pick the single most likely decision. We conclude that expert judgment should not be used for predicting decisions that people will make in conflicts. When decision makers ask experts for their opinions they are likely to overlook other, more useful, approaches.

Keywords: bad faith, framing, hindsight bias, methods, overconfidence, politics.

*Corresponding author.

Asking an expert to predict what will happen in a conflict within his domain seems a reasonable thing to do. For example, the media find professors and politicians to tell us what will happen when discussing conflicts such as the war on terrorism. In business, the CEO might ask the marketing manager to predict how competitors will respond to a new product launch or ask the human resources manager whether the offer of a 2% wage increase will deter a threatened strike. In the military, a general might ask his intelligence officer whether the enemy is likely to defend an outpost.

Evidence from surveys suggests that forecasts of decisions in conflicts are typically based on experts' unaided judgments (Armstrong, Brodie, and McIntyre 1987). Informal evidence that this is so abounds in everyday life and in the news. Winston Churchill observed that a politician should have "The ability to foretell what is going to happen... And to have the ability afterwards to explain why it didn't happen" (Adler 1965, p. 4). The same observation might be made of executives in business, the public sector, and the armed services.

While it is attractive to think that if we can find the right expert we can know what will happen, Armstrong (1980) in a review of evidence from diverse subject areas was unable to find evidence that expertise, beyond a modest level, improves experts' ability to forecast accurately.

Some beliefs about the value of expertise

What do people think about the value of expertise when forecasting decisions in conflict situations? Prior to giving talks about forecasting, we asked attendees for their opinions on the likely accuracy of experts' and novices' (university students') forecasts of decisions in conflicts. We told respondents that, for the purpose of our survey, they should assume that people who were asked to make predictions were presented with descriptions of several different conflicts and were asked to choose from between three and six possible decisions such that the expected accuracy from choosing randomly across the full set of conflicts was 28%. The figure of 28% is the average chance of a correct prediction for the eight conflicts we used in our research, or $[\frac{1}{6} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}] / 8 * 100$. By asking respondents to adopt the 28% figure for chance when they made their assessments we are able to make meaningful comparisons between our research findings and their accuracy expectations.

The talks in which we conducted our survey were to academics and students at Lancaster University (19 usable responses), Manchester Business School (18), Melbourne Business School (6), Royal New Zealand Police College educators (4), Harvard Business School alumni (8), conflict management practitioners in New Zealand (7), and attendees at the International Conference on Organizational Foresight in Glasgow (15). A copy of the questionnaire we used is available at www.conflictforecasting.com. [It is included at the end of this paper, for the purpose of review only, as Reviewer Appendix 1]. We excluded 27 responses from those who expected accuracy to be less than 28% for any method as it seemed implausible to us that the forecasts of any method would on average be worse than chance. If a method really were worse than chance, the decision predicted by the method could be eliminated and another one chosen at random; one would thereby obtain forecasts that were more accurate than chance.

Our practitioners, forecasting experts, and miscellaneous academics had little faith in the predictions of novices, expecting their predictions to be accurate only 30% of the time—little better than chance. The respondents had greater confidence in experts: 66% expected experts to be more accurate than novices whereas only 9% expected novices to be more accurate. Despite their greater faith in experts, respondents expected only 45% of experts' forecasts to be accurate¹.

We suggest that accurate prediction is difficult because conflicts tend to be too complex for people to think through in ways that realistically represent their actual progress. Parties in conflict often act and react many times, and change as a result of their interactions. There may be interactions within each party and more than two parties involved.

Tversky and Kahneman (1982) suggested that when people are faced with complex situations, they are likely to resort to the heuristic of availability in order to judge the likelihood of outcomes. That is, they test their memories and judge an outcome likely when a similar outcome is easily recalled or imagined. For example, some people tend to think it likely that new wars will end badly because the unceremonious withdrawal of US and allied troops from Vietnam is such a vivid memory for them (Kagan 2005). There is, however, ample reason to be skeptical about whether the availability heuristic will lead to accurate predictions. For example, salient outcomes and the situations that gave rise to them are unlikely to be representative; quite the opposite. Unstructured reviews of the past are likely to offer poor guidance for the future (Fischhoff 1982, Harvey 2001).

Information processing is problematic. If we take Bayes's theorem as the standard, people tend to adjust their predictions less than they should when they receive new information (Edwards 1982). When they consider the likelihood of an outcome from a multistage process (Hitler invades Belgium, he succeeds, Britain declares war, Hitler attacks Britain) people have the opposite tendency: they act as though their best guesses of what will happen at early stages are certainties (Gettys, Kelly, and Peterson 1982).

Stewart (2001) found that judgmental forecasts are likely to be unreliable when (1) the task is complex, (2) there is uncertainty about the environment, (3) information acquisition is subjective, or (4) information processing is subjective. Stewart's four conditions for unreliability are likely to be met with the type of problem we are considering.

It is difficult for people to become better at predicting decisions in conflicts using unaided judgment because basic conditions for learning are typically absent. Timely and unambiguous feedback is uncommon, and opportunities for practise are rare (Arkes 2001). Feedback may be in the form of deliberately misleading information leaked by an adversary or the unreliable accounts of witnesses. Accurate feedback may be misinterpreted because experts misunderstand the situation (Einhorn 1982). Decision-makers may take action aimed at avoiding a predicted outcome thereby confounding feedback. Conflicts often occur over long periods of time and those responsible for predicting an outcome may no longer be present when the actual outcome emerges. Many experts will be faced with important conflicts only rarely and, in any case,

¹ If the excluded responses were included, the average expectations would be 30% for novices and 42% for experts instead of 30% and 45% respectively.

conflicts are typically diverse and each one may appear more–or–less novel. Spurious correlations that support experts’ theories can be readily constructed (Chapman and Chapman 1982; Jennings, Amabile, and Ross 1982).

Finally, Tetlock (1999) found that experts have excellent defenses against evidence that their forecasts were wrong, so that even in situations where conditions for learning are good, experts may still fail to learn.

Robert McNamara (Morris 2003), Secretary of Defense under Presidents Kennedy and Johnson, referred to the “fog of war” in relation to conflicts in which he was involved. We suggest that this term, which appears to have originated in the writings of Napoleonic wars veteran Prussian Major General Carl von Clausewitz², might reasonably be applied to most conflict situations where unaided judgment is applied.

Research method

We recruited domain experts, conflict experts, and forecasting experts to predict the decisions made in eight diverse conflicts. The conflicts were real situations for which accurate forecasts might reasonably have been expected to save money or lives. Each was either obscure or was disguised in order to make recognition of the real situation unlikely. The specific conflicts were chosen for their diversity and because information about them was readily obtainable. The conflicts involved nurses striking for pay parity, football players wanting a bigger share of revenues, an employee resisting the down–grading of her job, artists demanding public financial support, a novel distribution arrangement proposed by a manufacturer to retailers, a hostile takeover attempt, a controversial investment proposal, and nations preparing for war. Each involved two or more interacting parties. The materials used in our research are available on conflictforecasting.com. [They are included at the end of this paper, for the purpose of review only, as Reviewer Appendix 2 and 3]

We allocated the conflicts to expert participants on the basis of their expertise. For example, we sent conflicts between employers and employees to industrial relations specialists, and we sent all eight conflicts to conflict management experts. Contact with participants was via email messages, and hence we had no control over the time they spent on the task or whether they referred to other materials or other people.

We recruited novices to make predictions for the same situations (Green 2005). Materials were the same as for the experts but, instead of receiving the material by email, the students were paid to sit in lecture theatres and make their predictions. No attempts were made to match the backgrounds of the students with the subject matter of the conflicts and, unlike the experts who had discretion over which if any of the conflicts they made predictions for, the students were paid \$20 only when they had provided forecasts for all their allocated conflicts.

² First published in 1832, Clausewitz’s writings have been republished in an English language edition as Clausewitz (1993).

Obtaining the forecasts

For each conflict, we provided participants with a set of between three and six decision options. We gave no instructions to participants on how they should make their predictions.

The way in which a problem is posed often affects judgmental predictions. One important distinction is whether a problem is framed as specific instance or a class of situations. For example, one might ask “How probable is it that the US will sign the Kyoto Protocol?” Alternatively, one could frame the problem as “In what proportion of cases would the US sign a treaty that would cause certain harm to the nation’s interests in return for uncertain benefits?” Kahneman and Tversky (1982a, 1982b) proposed that whereas people tend to think of situations as being “singular” when they assess the likelihood of outcomes (Kyoto Protocol signature), their predictions would be more accurate if they used a “distributional” approach (international treaty signatures) to assess likelihood. Kahneman and Lovallo (1993) presented evidence on the superiority of a distributional approach using the term “outside view.” Tversky and Koehler (1994) postulated that the greater accuracy is a result of peoples’ tendency to consider alternatives in more detail. They suggest that people are prompted to think more about different ways that an outcome might occur when a problem is framed as a class of similar situations than when it is framed as a singular instance. Cosmides and Tooby (1996) found evidence for the proposition that people have innate mechanisms for storing and manipulating frequency information.

We conducted an experiment to compare the accuracy of unaided–judgment forecasts collected using a singular format with those collected by asking for frequencies of different decisions across a set of hypothetical similar situations. We hypothesized that participants who were asked for frequencies might provide forecasts that were more accurate than those who were not.

Fifty–two participants, all university students, were paid the equivalent of US\$20 to take part in the experiment. We allocated them randomly between the singular and frequencies treatments. Each singular–treatment participant received a different sequence of four of the eight conflicts we used in our research and matching sequences were given to frequencies–treatment participants. For each conflict, participants were given approximately 30 minutes to read the material and answer the questions.

Four participants each claimed to recognize a situation, and we excluded their responses. Aside from the following forecasting questions, the treatments were identical.

Singular treatment question:

How was the stand-off between Localville and Expander resolved?

(check one ✓ or %)

- a. Expander's takeover bid failed completely
- b. Expander purchased Localville's mobile operation only
- c. Expander's takeover succeeded at, or close to, their August 14 offer price of \$43-per-share
- d. Expander's takeover succeeded at a substantial premium over the August 14 offer price

Frequencies treatment question:

Assume there are 100 situations similar to the one described, in how many of these situations would...

- a. The takeover bid fail completely? out of 100
- b. The mobile operation alone be purchased? out of 100
- c. The takeover succeed at, or close to, the offer price? out of 100
- d. The takeover succeed at a substantial premium over the offer price? out of 100

Findings

Expert versus novice judgment

Recall that our survey respondents expected experts' unaided-judgment forecasts to be substantially more accurate (45%) than those of novices (30%): This did not prove to be the case. The unaided experts' accuracy averaged only 32% across the conflicts used in our studies, little better than the average accuracy of 29% for novices' forecasts (Table 1). These results are consistent with evidence summarized in Armstrong (1985, pp. 91 – 96); there was little relationship between expertise and forecast accuracy. Neither group did appreciably better than chance.

We used the permutation test for paired replicates (Siegel and Castellan 1988) to test the significance of the differences in accuracy between experts and chance across the eight conflicts. As a casual inspection of the data in Table 1 suggests, the differences are quite likely to have arisen by chance ($P = 0.30$, one-tail test). The test is 100% power-efficient as all the information is used (Siegel and Castellan 1988, p. 100).

Table 1
Accuracy of unaided-judgment forecasts
 Percent correct forecasts (number of forecasts)

	Chance	By novices	By experts
Artists Protest	17	5 (39)	10 (20)
Distribution Channel	33	5 (42)	38 (17)
Telco Takeover	25	10 (10)	0 (8)
55% Pay Plan	25	27 (15)	18 (11)
Zenith Investment	33	29 (21)	36 (14)
Personal Grievance	25	44 (9)	31 (13)
Water Dispute	33	45 (11)	50 (8)
Nurses Dispute	<u>33</u>	<u>68</u> (<u>22</u>)	<u>73</u> (<u>15</u>)
Averages (unweighted)	28	29 (169)	32 (106)

Expert experience and accuracy

Is it possible to identify experts who are more likely than others to make accurate judgmental forecasts? One obvious way to assess this is to compare the accuracy of forecasts from experts with more experience with those from experts with less.

We asked expert participants to record the number of years experience they had as “a conflict management specialist.” As a check, we also asked some of our novice participants the same question and the responses were not surprising. Ninety-four percent of the university student participants who answered the question gave their experience as zero years; the rest claimed one or two years of such experience.

Commonsense expectations did not prove to be correct. The 57 forecasts of experts with less than five years experience were more accurate (36%) than the 48 forecasts of experts with more experience (29%).

We also asked our expert participants to rate their experience with conflicts similar to the one they were examining on a scale from zero to ten. Those who considered they had little experience with similar conflicts (they gave themselves ratings of 0 or 1) were equally accurate at 34% (72 forecasts) as those who gave themselves higher ratings (32 forecasts).

Expert confidence and accuracy

Perhaps experts’ confidence in their individual forecasts could be used to identify accurate forecasts. On the other hand, confidence might be misplaced when the forecasting problems are difficult.

We asked our expert participants:

How likely is it that taking more time would change your forecast?
{0 = almost no chance (1/100) ... 10 = practically certain (99/100)} [____] 0-10.

While it is possible that the experts might have reasoned that they were unlikely to change a forecast given more time because they did not expect their forecast to be better than guessing in any case, the fact of their participation and our evidence on accuracy expectations suggests this was not the case. We interpret the experts' responses to this question as a measure of their confidence in the accuracy of their forecasts. We compared the accuracy of forecasts in which experts had high confidence with those in which they had less confidence. Where experts assessed the likelihood that they would change their forecasts given more time as between zero and two out of 10—i.e. no more than 0.2 probability of change—we coded the forecasts as “high confidence.” All other forecasts were coded as “low confidence.” Using unweighted averages across the conflicts, the 68 high-confidence forecasts were less accurate at 28% than the 35 low-confidence forecasts at 41%.

We also compared the confidence that the experts had in their forecasts that turned out to be accurate with their confidence in forecasts that turned out to be inaccurate. There were six conflicts for which both accurate and inaccurate forecasts were available and for which no half-right forecasts had been provided³. We found, using unweighted averages across the six conflicts, that the experts assessed the probability that they would change the 27 accurate forecasts as 0.25 and that they would change the 51 inaccurate forecasts as 0.17.

Frequency responses and accuracy

We anticipated participants would be more accurate when asked to estimate the frequencies of outcomes for many similar situations. Our university student participants who judged relative frequencies were no better at identifying the actual decision than were participants who simply chose the decision they thought most likely. Averaged across conflicts, 33% of both groups' forecasts were accurate (Table 2). Also, the accuracy figures for the two groups appear to follow the same pattern when looking across the situations (Spearman rank order correlation coefficient 0.59, $P < 0.10$; Siegel and Castellan 1988).

³ The Distribution Channel conflict offered “c. Either a or b” as an option and the nine such responses were coded as 0.5.

Table 2
Accuracy of novices' frequency and singular forecasts
 Percent correct forecasts (number of forecasts)

	Chance	Frequencies	Singular	Total
55% Pay Plan	25	0 (12)	9 (11)	4 (23)
Artists Protest	17	10 (10)	0 (11)	5 (21)
Distribution Channel	33	23 (13)	38 (13)	31 (26)
Personal Grievance	25	11 (9)	46 (13)	32 (22)
Telco Takeover	25	50 (12)	25 (12)	38 (24)
Zenith Investment	33	40 (10)	42 (12)	41 (22)
Water Dispute	33	67 (12)	42 (12)	54 (24)
Nurses Dispute	<u>33</u>	<u>64</u> (<u>11</u>)	<u>58</u> (<u>12</u>)	<u>61</u> (<u>23</u>)
Averages (unweighted)	28	33 (89)	33 (96)	33 (185)

Of the 89 frequencies predictions, 54% summed to the total of 100 specified in the frequencies–treatment question; 35% totaled more than 100 and 11% less. It is arguable that, despite our intentions, the decision options we provided were not entirely mutually–exclusive or exhaustive and hence the failure of some participants' responses to add to 100 is not necessarily a failure of logic on their part. On the other hand, researchers have found that even with mutually exclusive and exhaustive lists of events, responses do not consistently sum to 1.0 or 100%, as people commonly fail to interpret probability or frequency scales in ways that researchers intend (Windschitl 2002).

Nonetheless, it seems reasonable to assume that our participants, who in most cases had only three or four decision options to assess, allocated frequencies that were at least consistent with their ranking of the options' likelihoods. For our analysis, therefore, we used the decision with the highest frequency or probability, or the single decision chosen, as the forecast. We dropped ten observations where there was a tie.

When we excluded from our analysis responses that did not sum to 1.0 or 100, it made no difference to our conclusion that asking participants for frequencies did not improve accuracy. Across the conflicts, the average accuracy for frequencies responses was 29% (48 forecasts) compared to 32% (93) for singular treatment responses.

Discussion and conclusions

The various people we surveyed expected it to be difficult to forecast decisions in conflicts. Our evidence has shown that this is indeed the case. Most respondents nonetheless expected experts to be better forecasters than novices. They were wrong. Expertise did not improve accuracy. Neither experts nor novices did substantially better than guessing.

Our concerns that the wording of our forecasting tasks might have harmed accuracy proved unfounded. An analysis using only responses that conformed to the norms of probability theory led to the same conclusion: asking for an assessment of the relative frequency of decisions across

similar situations did not help. We suggest that the complexity of conflict situations means that people tend to view each one as being more–or–less unique and therefore do not store or recall frequency information in the way that they do for simpler situations such as rainy days in April or the presence of speed cameras on alternative routes home from work.

There are *no* good grounds for decision makers to rely on experts' unaided judgements for forecasting decisions in conflicts. Such reliance discourages experts and decision makers from investigating alternative approaches (Arkes 2001). While it is difficult to accurately forecast decisions in conflict situations, we have shown in Green (2005) and Green and Armstrong (2004) that it *is* possible to obtain substantially better forecasts.

Green (2005) provided evidence that simulated interaction, a type of role playing for forecasting behaviour in conflicts, reduced error by 47% compared to game theory experts' forecasts. Role players were mostly undergraduate university students. In Green and Armstrong (2004), experts were induced to recall and analyse information on similar situations from the past using a method called structured analogies. Where experts were able to think of at least two analogies, error was reduced by 39% compared to chance accuracy.

Given the methods currently used in forecasting, to accuse expert advisors and political leaders of bad faith when their predictions about conflicts prove wrong does not seem justified. Inaccurate predictions are to be expected when experts use unaided judgment to predict how people will behave in conflicts.

Acknowledgements

We are grateful to Paul Goodwin for organising the special section in which this article features and to Robyn Dawes, Don Esslemont, Jonathan J. Koehler, and Lee Ross for helpful suggestions on various drafts of this article. We are also grateful for copy–editing by Stuart Halpern and Bryan LaFrance. The article was also improved in response to probing questions from delegates at the 2003 and 2004 International Symposia on Forecasting and at the Institute of Mathematics and Its Applications' Conference on Conflict and Its Resolution, and from attendees at talks at RAND Organization, the CIA's Sherman Kent School, Warwick Business School, University College London, Monash University, and Melbourne Business School to whom we presented elements of the work reported here.

References

- Adler, B. (Ed.) (1965). *The Churchill Wit*. New York: Coward–McCann.
- Arkes, H. R. (2001). Overconfidence in judgmental forecasting, in Armstrong, J. S. (ed.), *Principles of forecasting*. Boston, MA: Kluwer Academic Publishers.
- Armstrong, J. S., Brodie, R. J., & McIntyre, S. H. (1987). Forecasting methods for marketing: Review of empirical research. *International Journal of Forecasting*, **3**, 335 – 376.
- Armstrong, J. S. (1980). The seer–sucker theory: The value of experts in forecasting. *Technology Review*, 83 (June/July), 18 – 24. Full text available at forecastingprinciples.com.

- Armstrong, J. S. (1985). *Long-range forecasting*. New York: John Wiley. Full text available at forecastingprinciples.com.
- Chapman, L. J., & Chapman, J. (1982). Test results are what you think they are, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Clausewitz, Carl von (1993). *On war*. Howard, M. & Paret, P. (ed./trans). New York: Alfred A. Knopf, "Everyman's Library" edition.
- Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, **58**, 1 – 73.
- Edwards, W. (1982). Conservatism in human information processing, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Einhorn, H. J. (1982). Learning from experience and suboptimal rules in decision making, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Fischhoff, B. (1982). For those condemned to study the past: heuristics and biases in hindsight, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Gettys, C. F., Kelly, C., & Peterson, C. R. (1982). The best-guess hypothesis in multistage inference, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Green, K. C. (2005). Further evidence on game theory, simulated interaction, and unaided judgement for forecasting decisions in conflicts. *International Journal of Forecasting*, **21**, 463 – 472. Full text of draft paper available at conflictforecasting.com
- Green, K. C. & Armstrong, J. S. (2004). Structured analogies for forecasting. Monash University Econometrics and Business Statistics Working Paper 17/04. Full text available at conflictforecasting.com
- Harvey, N. (2001). Improving judgment in forecasting, in Armstrong, J. S. (ed.), *Principles of forecasting*. Boston, MA: Kluwer Academic Publishers.
- Jennings, D. L., Amabile, T. M., & Ross, L. (1982). Informal covariance assessment: data-based versus theory-based judgments, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Kagan, F. W. (2005). Iraq is not Vietnam. *Policy Review*, **134**. Full text available at policyreview.org.
- Kahneman, D. & Lovallo, D. (1993). Timid choices and bold forecasts: a cognitive perspective on risk taking. *Management Science*, **39**, 17 – 31.
- Kahneman, D. & Tversky, A. (1982a). Intuitive prediction: biases and corrective procedures, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Kahneman, D. & Tversky, A. (1982b). Variants of uncertainty, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

- Morris, E. (2003). *The fog of war: eleven lessons from the life of Robert S. McNamara*. USA: Sony Pictures Classics. (Clips available for viewing at <http://www.sonyclassics.com/fogofwar/>)
- Siegel, S. & Castellan, N. J. Jr. (1988). *Non-parametric Statistics for the Behavioral Sciences*, 2nd ed. Singapore: McGraw-Hill.
- Stewart, T. R. (2001). Improving reliability in judgmental forecasts, in Armstrong, J. S. (ed.), *Principles of forecasting*. Boston, MA: Kluwer Academic Publishers.
- Tetlock, P. E. (1999). Theory driven reasoning about possible pasts and probable futures: are we prisoners of our perceptions? *American Journal of Political Science*, **43**, 335 – 366.
- Tversky, A. & Kahneman, D. (1982). Availability: a heuristic for judging frequency and probability, in Kahneman, D., Slovic, P. & Tversky, A. (eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Tversky, A. & Koehler, D. J. (1994). Support theory: a nonextensional representation of subjective probability. *Psychological Review*, **101**, 547 – 567.
- Windschitl, P. D. (2002). Judging the accuracy of a likelihood judgement: the case of smoking risk. *Journal of Behavioral Decision Making*, **15**, 19 – 35.