FORECASTING DECISIONS IN CONFLICTS:
ANALOGY, GAME THEORY, UNAIDED JUDGEMENT, AND SIMULATION COMPARED

by

Kesten Charles Green

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the

requirements for the degree of

Doctor of Philosophy

in Management

Victoria University of Wellington

4 September, 2003

Abstract

There has been surprisingly little research on how best to predict decisions in conflicts. Managers commonly use their unaided judgement for the task. Game theory and a disciplined use of analogies have been recommended. When tested, experts using their unaided judgement and game theorists performed no better than chance. Experts using structured analogies performed better than chance, but the most accurate forecasts were provided by simulated interaction using student role players. Twenty-one game theorists made 98 forecasts for eight diverse conflicts. Forty-one experts in conflicts made 60 solo forecasts using structured analogies and 96 solo forecasts using unaided judgement (a further seven provided collaborative forecasts only) while 492 participants made 105 forecasts in simulated interactions. Overall, one-in-three forecasts by game theorists and by experts who did not use a formal method were correct. Forecasters who used structured analogies were correct for 45 percent and forecasts from simulated interactions were correct for 62 percent of forecasts. Analysis using alternative measures of accuracy does not affect the findings. Neither expertise nor collaboration appear to affect accuracy. The findings are at odds with the opinions of experts, who expected experts to be more accurate than students regardless of the method used.

Keywords: accuracy, analogy, conflict, expert opinion, forecasting, game theory, unaided judgement, role playing, simulated interaction, simulation, structured analogies.

Contents

Appendices

Lists of tables, figures, and formulae

Tables

Figures

Formulae

1.   Introduction

If you can look into the seeds of time,
And say which grain will grow and which will not,
Speak then to me, who neither beg nor fear
Your favours nor your hate.
                    Shakespeare (1606), Banquo to the witches.

Like Banquo, who consulted the "weird sisters", modern managers often wish to know
how a conflict will unfold. Whether a conflict is industrial, commercial, civil, political,
diplomatic, or military, predicting the decisions of others can be difficult. Yet it is
important that managers plan for likely eventualities and seek effective strategies. Errors
in predicting the decisions of others can lead to needless strikes, losses, protests,
reversals, wars, and defeats. This research addresses the problem of choosing the best
method for forecasting decisions made in conflicts: particular conflicts that involve
interaction between few parties.

Conflicts are complex and hence decisions in conflict situations can be difficult to
predict. The complexity of conflicts is highlighted by, for example, the number and
variety of aspects Raiffa (1982) described in his attempt to characterise them.
Paraphrased, Raiffa's conflict characteristics are: number of parties, cohesion of parties,
likelihood of iteration, possibility of linkage, number of issues, need for agreement, need
for ratification, possibility of threats, constraints on time, binding of agreement, arena of
negotiation, norms of parties, and possibility of intervention. Raiffa himself describes
the characteristics as a "partial classification" (p. 11).

1.1   Outline

This thesis replicates and extends the research described in Armstrong (2001a).
Armstrong presented evidence on the accuracy of forecasts of decisions in conflicts from
two methods: unaided judgement and role playing. The participants in the research were
primarily university students. Armstrong sought evidence on the relative accuracy of a
third method, forecasts by game theorists, but was unable to find any.

In my work, I have followed Armstrong's (2001e) recommendations on evaluating forecasting methods. First, I describe my search for evidence on the relative accuracy of forecasts from reasonable alternative methods for forecasting decisions in conflicts. Second, I describe my research and present my findings. Third, I assess the generalisability of the findings. Finally, I draw on my findings to make recommendations for managers that, if adopted, will lead to improvements in the accuracy of forecasts of decisions in conflicts.

Document structure

There are five chapters in this document. In this, the first chapter, I describe the methods that are used or have been recommended for forecasting decisions in conflicts. I then describe the objectives of my research and their motivation, and address the implications of the objectives for my research programme.

In chapter 2, I describe my search for empirical evidence on the accuracy of forecasts from four conflict forecasting methods and present the findings of my search.

In chapter 3, I discuss the methodology of my empirical research, and describe my research programme in detail. The chapter includes detailed descriptions of the four forecasting methods that I compared, the conflicts that I used and how I chose them, and how I collected forecasts and opinions from participants.

In chapter 4, I present my findings on the relative performance of the four methods. I examine the effect on forecast accuracy of forecaster expertise and of collaboration between forecasters. I also present my findings on the likely appeal to managers of the methods I examined.

Finally, in chapter 5, I draw conclusions about the relative performance of the different conflict forecasting methods and about other influences on forecast accuracy. Some of these conclusions will be surprising to experts and managers. The chapter includes discussion of implications and limitations of the research, as well as suggestions on further research and recommendations to managers on choosing and implementing forecasting methods for conflicts.

1.2   Conflict forecasting methods

Forecasting methods are often chosen because of popularity (frequency of use by practitioners), or on the basis of expert judgement. Armstrong, Brodie, and McIntyre (1987) surveyed forecasting practitioners on their use of six methods for forecasting decisions in conflicts and for their assessment of the usefulness of these methods. The authors intended these to be an exhaustive list of methods for forecasting decisions in conflicts (personal communication from J. S. Armstrong, 29 August 2001).

The methods that were included in the Armstrong et al. (1987) survey were: unaided judgement, intentions of other parties, game theory, statistical analysis of analogies, role-playing, and field experiments. Singer and Brodie (1990) evaluated the face validity of theories of and approaches to analysing business competition. The authors suggested that the findings of their evaluation were broadly in accord with the stated forecasting method preferences of respondents to the Armstrong et al. (1987) survey. They concluded that expert judgement and role playing were associated with superior approaches, and that game theory extensions appeared worthy of further research.

In this section, I consider unaided judgement and the intentions of other parties under the heading of "unaided judgement", game theory under the heading of "game theory", statistical analysis of analogies under the heading of "structured analogies" and role playing and field experiments under the heading of "simulated interaction". I suggest that incorporating the avowed intentions of others is a common aspect of unaided judgement, and treat the two as one method. I have not included field experiments in my research. I present findings from the Armstrong et al. (1987) survey, together with other opinions on the usefulness of the methods.

1.2.1   Unaided judgement

The term "unaided judgement" is intended to be self-explanatory – it is judgement without recourse to a formal forecasting method. For the purposes of my research, unaided judgement is what managers or forecasters use when they are asked to forecast decisions in real conflicts, but do not use a particular method.

It is clear from the literature that managers mostly rely on their own judgement for forecasting decisions in conflicts; either entirely or in conjunction with the judgemental predictions of others who know about the situation. In some situations it may be practical to ascertain the judgements of the other party or parties to a conflict, in the form of their avowed intentions. For example, in his political manifesto, *Mein Kampf*, Hitler outlined the policies he would later pursue as German dictator (Drabble, Ed., 1995). Managers can incorporate such information into their own judgemental forecasts of the behaviour of another party.

Expert judgement was used for forecasting competitive action by 85 percent of organisations in the Armstrong et al. (1987) practitioner survey. More than 90 percent of the forecasting and marketing experts surveyed endorsed expert judgement for this purpose. Singer and Brodie (1990) observed that expert judgement "plays a major role as a forecasting technique because there is no comprehensive unified theory from which formal or analytic techniques might be derived" (p. 86). Unaided judgement is thus a benchmark against which other conflict forecasting methods must be judged.

1.2.2   Game theory

Hargreaves Heap and Varoufakis (1995) describe game theory as being underpinned by three key assumptions about the parties in conflict. These assumptions are that the parties are (a) instrumentally rational, (b) know this, and (c) know the rules. In order to forecast decisions that will be made in a real conflict, a game theorist might (1) develop a new model (or adapt on old one) based on rules and utilities deduced from knowledge of the conflict, (2) use judgement informed by knowledge of game theory, or (3) use some combination of modelling and judgement. Experts are employed for their expertise and so, for the purpose of my research, game theory is the method used by game-theory experts when they are asked to forecast decisions made in real conflicts.

It seems reasonable to suppose that game theory could help practitioners to forecast more accurately than they would if they relied on unaided judgement because, for example, the discipline of the approach should help to counter judgemental biases. Nalebuff and Brandenburger (1996, p. 8) wrote:

> By presenting a more complete picture of each ... situation, game theory makes it possible to see aspects of the situation that would otherwise have been ignored. In these neglected aspects, some of the greatest opportunities ... are to be found.

McAfee and McMillan (1996) made the bolder statement that game theory "is to show how people behave in various circumstances" (p. 172). The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel was awarded in 1994 to three game theorists: John C. Harsanyi, John F. Nash, and Reinhard Selton. A press release from Kungliga Vetenskapsakademien The Royal Swedish Academy of Sciences (1994) stated:

> … non-cooperative game theory… has had a great impact on economic research. The principal aspect of this theory is the concept of equilibrium, which is used to make predictions about the outcome of strategic interaction.

Game theorists "hope to produce a complete theory and explanation of the social world" (Bullock and Trombley, Eds., 1999). Goodwin (2002) found the authors of two of a convenience sample of six introductory game theory textbooks (Dixit and Skeath, 1999; and Hargreaves Heap and Varoufakis, 1995) claimed that the method has value for prediction or explanation. Binmore (1990) puts prediction first in a list of the aims of game theory. The authors of a recent edition of a textbook on corporate strategy (Johnson and Scholes, 2002) stated "Game theory provides a basis for thinking through competitors' strategic moves in such a way as to pre-empt or counter them" (p. 354).

Game theory was recommended by some experts in the Armstrong et al. (1987) survey. It was used in nearly 10 percent of the surveyed organisations.

While Nalebuff and Brandenburger (1996) and Bullock and Trombley (Eds.) (1999), for example, made optimistic claims for game theory, Shubik (1975, p. xi) described as "peculiarly rationalistic" the assumptions behind formal game theory:

> It is assumed that the individuals are capable of accurate and virtually costless computations. Furthermore, they are assumed to be completely informed about their environment. They are presumed to have perfect perceptions. They are regarded as possessing well-defined goals. It is assumed that these goals do not change over the period of time during which the game is played.

Shubik suggested that while game theory may be applicable to actual games (such as backgammon or chess), and may even be useful for constructing a model to approximate an economic structure such as a market, "it is much harder to consider being able to trap the subtleties of a family quarrel or an international treaty bargaining session" (1975, p. 14).

The claims made for game theory by some authors, the recommendations of experts to use game theory, the evidence of game theory's use by forecasting practitioners, and controversy over the usefulness of game theory are all reasons to ask whether the method can provide managers with useful predictions for real conflicts.

1.2.3   Structured analogies

The entry on "analogy" in the *Forecasting Dictionary* (Armstrong, 2001g) stated: "A resemblance between situations as assessed by domain experts. A forecaster can think of how similar situations turned out when making a forecast for a given situation". The structured-analogies method is described in the online version of the *Forecasting Dictionary*[1] as involving

> …domain experts selecting situations that are similar to a target situation, describing the similarities and differences, and providing an overall similarity rating for each similar (analogous) situation. The outcomes of the analogous situations are then used to forecast the outcome of the target situation. The analogous situations' outcomes can be weighted to forecast a target situation decision or used to assign probabilities to possible decisions.

This is the approach that I adopted.

Analogous information has been shown to improve forecast accuracy in forecasting tasks other than forecasting decisions in conflicts. For example, Efron and Morris (1977) show that forecasts of an individual baseball player's final batting average are more accurate when the player's early-season average is heavily weighted by the league average than are forecasts based on the individual's early-season average alone. Kahneman and Tversky (1982) recommend a similar procedure for adjusting "intuitive"

---

1 http://morris.wharton.upenn.edu/forecast/dictionary, 12 August, 2002.

numerical forecasts (they use the example of sales of a book) towards the average for a reference class (say, cookbooks by television celebrity cooks).

People who are asked to use their judgement to make a prediction for a situation may think of analogous situations. Neustadt and May (1986) provided examples of analogies being used by decision-makers to forecast the decisions of others in conflicts such as the Cuban missile crisis. The authors suggested the use of analogies may in many situations have led to *in*accurate predictions with serious consequences. They attributed instances of forecast inaccuracy, in part, to an ill-disciplined or uncritical use of analogies, and recommended a more formal use of analogies to improve accuracy.

Analogies have been used in a formal way to forecast the distant future. For example, in "The Railroad and the Space Program: An Exploration in Historical Analogy" (Mazlish (Ed.), 1965) the authors used a single historical analogy "as a device to assist us in forecasting... the impact of the space program on society" (p. v). Glantz (1991) explored the use of analogies for, *inter alia*, forecasting societal responses to climate change.

Khong (1992) examined the evidence for and against the view implicit in Neustadt and May (1986) that analogies are used by policymakers for analysis, and not solely for advocacy and justification. Khong argued that the Neustadt and May view was supported by the evidence. In particular, Khong argued that the favoured analogies of decision-makers and advisors provided the best explanation of the decisions made by the US administration early in the Vietnam war. He also suggested that analogies are not used very well because policymakers tend to cling to readily accessible analogies and to reject disconfirming evidence, rather than because they do not adhere to formal processes. Nevertheless, the use of formal procedures for forecasting has been shown to increase experts' accuracy (for example Armstrong, 2001b; Collopy, Adya and Armstrong, 2001; Harvey, 2001; MacGregor, 2001; Rowe and Wright, 2001; and Stewart, 2001).

The use of analogies was recommended by Armstrong (2001c) for forecasting problems where similar situations can be identified. Armstrong has also suggested (2001a) that extrapolating from analogies may be useful for forecasting decisions in conflicts, but pointed out that novel situations and novel strategies will lack obvious analogies – that is, similar situations cannot be identified.

More than half of the experts surveyed by Armstrong et al. (1987) agreed that a formal analysis of analogies should be useful. Statistical analysis of analogous situations was the second most popular method for forecasting competitor actions – being used by 58 percent of organisations.

The common use of analogies for forecasting decisions in conflicts is sufficient reason to ask whether the method can help provide managers with useful predictions for conflicts.

1.2.4   Simulated interaction

Experiments in the field can be used to predict decisions in conflicts. Although 40 percent of experts in the Armstrong et al. (1987) survey recommended experimentation, the method was not popular with practitioners. I do not examine field experiments in this research.

Laboratory experiments, in the form of role playing, can substitute for field experiments by simulating a conflict using people who are not party to the conflict. Role playing is likely to be cheaper than field experiments, and the risk of alerting rivals is reduced.

Role playing is described in the online *Forecasting Dictionary* (op. cit., 12 August 2002) as "a technique whereby people play roles to understand or predict behavior". As the entry suggests, role playing is a technique that is applicable to problems beyond those considered here. To avoid confusion, the use of role playing to simulate the interactions of small numbers of parties whose roles are likely to lead to conflict is referred to as "simulated interaction" (online *Forecasting Dictionary*, op. cit., 12 August 2002). I have used the term "simulated interaction" in the balance of this document in preference to the term "role playing", except in cases of direct quotations or where the term "role playing" is more appropriate.

Discussions of the usefulness and realism of simulated interaction are a feature of the game-theory literature. The method is often contrasted with the limitations of game theory in this context. Nalebuff and Brandenburger (1996, p. 62), for example, noted that it is both important and difficult to appreciate the perceptions of other parties. They

suggested that managers might "ask a colleague to role-play by stepping into [another] player's shoes" (p. 63) in order to gain a better appreciation. The role-play outcomes of contrived situations are commonly used by game-theory researchers as the behavioural benchmark against which their hypotheses are tested. Pioneer of experimental economics, Vernon Smith, wrote "Theories based upon abstract conditions make no predictions… I see no way for game theory to advance independently of experimental (or other) observations" (1994, p. 121).

Shubik (1975) covered similar ground when he wrote of simulated interaction that "an extremely valuable aspect of operational gaming is the perspective gained by viewing a conflict of interests from the other side. Experience gained in playing roles foreign to one's own interests may provide insights hard to obtain in any other manner" (p. 9). He also pointed out game theory's lack of realism relative to simulated interaction (gaming):

> In summary we should suggest that many of the uses of gaming are not concerned with problems which can be clearly and narrowly defined as belonging to game theory. Environment-poor experimental games come closest to being strict game theory problems. Yet even here, features such as learning, searching, organising, are best explained by psychology, social-psychology, management science, and other disciplines more relevant than game theory (p. 17).

And on the same topic, Schelling (1961, p. 47) observed that

> Part of the rationale of game organization [simulated-interaction experiments] is that no straightforward analytical process will generate a 'solution' to the problem, predict an outcome, or produce a comprehensive map of the alternative routes, processes, and outcomes that are latent in the problem.

In contrast, simulated interactions

> …do generate these complexities and, by most reports, do it in a fruitful and stimulating way.

Surprisingly, although simulated interaction was recommended by most of the forecasting experts in the Armstrong et al. (1987) survey, and there is evidence available that the method provides more accurate forecasts than can be obtained from unaided judgement, it is not often used in practice.

Armstrong's (2001a) evidence on the accuracy of students' simulated-interaction decisions relative to the accuracy of students' unaided-judgement forecasts and the

recommendations of experts are good reasons to include the method of simulated interaction in a comparison of conflict forecasting methods.

1.2.5 Survey of experts' accuracy expectations

In order to obtain formal data on experts' expectations of the accuracy of different conflict forecasting methods examined in this research, Professor Armstrong (personal communication, 2002) surveyed academics and students attending a talk at Lancaster University on 24 April 2002. He obtained responses from 27 people. Before asking participants for their expectations, Armstrong described the forecasting methods and their implementation in my research. He also described five of the conflicts I had used and told the audience that by choosing at random from the decision options provided for these conflicts, one could expect to be correct 28 percent of the time. Armstrong repeated this procedure in a talk to Harvard alumni (responses from 18 business executives) on 7 May 2002. I followed the same procedure in a talk to practitioners, which was organised by the New Zealand Centre for Conflict Resolution, on 17 July 2002 (responses from 12 people). I repeated the procedure in a talk to educators at the Royal New Zealand Police College on 19 July 2002 (responses from five people).

Overall, these various experts expected the unaided judgement of novices to be no better than chance (Table 1). They expected a modest improvement in accuracy if novices were used as role players, rather than as forecasters. Finally, the experts expected experts to be more accurate than novices, regardless of the methods used.

Table 1
Experts' expectations of forecasting methods' accuracy [a]
Percent correct (number of responses)

Method	Actual [b]	Expectation [c]	Difference
Unaided judgement (by novices)	27 (139)	30 (60)	3
Simulated interaction (using novices)	61 (75)	40 (60)	-21
Unaided judgement (by experts)		50 (62)	
Game theory (by experts)		50 (60)	
Structured analogies (by experts)		50 (61)	
Simulated interaction (using experts)		50 (57)	

a Forecasts for conflicts: Artists Protest, Distribution Channel, 55% Pay Plan, Nurses Dispute, and Zenith Investment.
b Findings from Armstrong (2001a) for Artists Protest, Distribution Channel, and 55% Pay Plan except for 13 unaided judgement findings from Green (2002a): Artists Protest (1 correct / n=8); Distribution Channel (1/5). Findings for Nurses Dispute and Zenith Investment from Green (2002a).
c Median expectation for the five conflicts listed in note "a".

There is evidence available on the accuracy of novices' forecasts and novices' simulated-interaction decisions from Armstrong (2001a). As my findings for these are similar to Armstrong's, I have provided aggregated figures in Table 1. On the basis of this evidence, experts were right in their expectations that novices would be no better than chance when they forecast using unaided judgement. They were wrong, however, in supposing that simulated interaction using novice role-players would offer little gain in accuracy over unaided judgement by novices. It is interesting that the experts' expectations were wrong on this, as findings of dramatic improvements in accuracy when novices simulated, rather than predicted, were published fifteen years ago (Armstrong, 1987).

The survey of expectations supports the need for empirical research by showing expert opinion to be a poor guide on the relative accuracy of forecasts from different conflict forecasting methods.

1.3   Objectives: motivation and implications

1.3.1   Overview

Evaluating forecasting methods

Singer and Brodie (1990) wrote that forecasting competitors' behaviour "has not received much attention in the forecasting literature… there is little guidance to practitioners as to which forecasting methods to use" (p. 75). The purpose of my research was to make useful recommendations to managers who face the problem of forecasting decisions made in real conflicts. In order to achieve this purpose, it was necessary to remedy the lack of evidence, identified by Singer and Brodie (1990), by conducting research.

The purpose dictated the research task, which was to evaluate reasonable alternative forecasting methods for conflicts. Principles for evaluating forecasting methods were provided by Armstrong (2001e) (Table 2). I used these principles to guide the selection and framing of research objectives, and the design of the research programme and of this document.

Table 2
Forecasting method evaluation principles [a]

A/ Using reasonable alternatives

  1 Compare reasonable forecasting methods

B/ Testing assumptions

  1 Use objective tests of assumptions

  2 Test assumptions for construct validity

  3 Describe conditions for generalisation

  4 Match tests to the problem

  5 Tailor analysis to the decision

C/ Testing data and methods

  1 Describe potential biases

  2 Assess reliability and validity of data

  3 Provide easy access to data

  4 Disclose details of methods

  5 Do clients understand [and accept] the methods?

D/ Replicating outputs

  1 Use direct replication to identify mistakes

  2 Replicate studies to assess reliability

  3 Extend studies to assess generalisability

  4 Conduct extensions in realistic situations

  5 Compare with forecasts from different methods

E/ Assessing outputs

  1 Examine all important criteria

  2 Specify criteria in advance

  3 Assess face validity of methods & forecasts

  4 Adjust error measures for scale

  5 Ensure error measures are valid

  6 Ensure error measures *in*sensitive to difficulty

  7 Ensure error measures are *un*biased

  8 Ensure error measures are *in*sensitive to outliers

  9 Do *not* use R^2 to compare models

10 Do *not* use RMSE

11 Use multiple error measures

12 Use *ex ante* tests for accuracy

13 Use statistical significance to test only reasonable models

14 Use *ex post* tests for policy effects

15 Obtain large samples of independent forecast errors

16 Conduct an explicit cost-benefit analysis

a Based on Exhibit 10 "Evaluation principles checklist" Armstrong (2001e, p. 465)

Appendix 1 provides a summary of how I addressed each of the principles in my research.

Estimate relative performance of methods

Accuracy is generally rated the most important criterion for selecting a forecasting method (Yokum and Armstrong, 1995) and is the principal criterion I consider. Consequently, the primary objective of my research was to estimate the relative accuracy of forecasts from reasonable forecasting methods – ones that were in use or recommended by experts. In this context, an accurate forecast of a decision made in a conflict is one that matches the decision actually made. For example, a decision may be made to reject a pay offer, resist a take-over bid, disrupt a community, change an allegiance, support a rebellion, or plan an invasion.

Assess generalisability and appeal to managers

My secondary objectives were to assess (a) the generalisability of the ranking of forecasting methods by relative accuracy and (b) the likely appeal to managers of the forecasting methods.

In order to assess generalisability, I investigated whether forecaster collaboration and forecaster expertise affected the principal findings on the relative accuracy of forecasts from the four methods.

Assessing the different methods' appeal to managers is critical to the purpose of making practical recommendations to managers. There would be no point in making recommendations on forecasting method selection if the recommendations were not accepted because of some overlooked selection criterion used by managers.

The data collection methods I used to address my research objectives are described in chapter 3.

1.3.2 Estimate relative performance of methods

Estimate effect of forecasting method on forecast accuracy

Armstrong (2001f) has conjectured that the accuracy of forecasts from a method is related to the realism with which that forecasting method allows forecasters to model the target situation (Principle 7.2, p. 695). I suggest that unaided judgement allows the least realism and simulated interaction the most. Game theory models conflicts using abstract mathematical analogies, whereas the structured-analogies method uses real analogous conflicts. I suggest that structured analogies will, therefore, allow more realistic modelling than will game theory. A distinction could be made between forecasting methods that rely on thinking and analysis by a forecaster, and those that rely on simulating. It seems reasonable to assume that simulation will tend to result in greater realism than thinking and analysing, particularly when forecasting a conflict that may involve several rounds of direct interaction between two or more parties (Armstrong, 2001a). These conjectures are, however, contrary to the opinions of experts (Table 1). In order to test the hypothesis of realism, I compare the accuracy of forecasts from the four methods I consider using percent correct, and other measures.

Accuracy is likely not only to be a function of the forecasting method employed, but also of how well the method is implemented. I describe how the methods are implemented, but do not attempt to quantify how well they were implemented, as it would be impossible to distinguish quality of implementation from the effects of the methods without a more extensive research programme.

Estimate effect of forecasting method on forecast usefulness

In carrying out analysis on forecast accuracy, I use measures of accuracy that assign no or negative value to forecasts that do not match the actual outcome. It is not necessarily the case, however, that such forecasts are valueless. From a manager's point of view, a forecast of a decision that is *similar* to the decision that actually occurs will be more valuable (useful) than a forecast of a decision that turns out to be *substantially different*

to the actual decision. I investigate the possibility that analysis based on forecast usefulness may lead to different conclusions about the relative merits of the forecasting methods.

1.3.3   Assess generalisability of findings

I extended prior research to assess the effects on forecasting accuracy of (1) collaboration between forecasters for methods other than simulated interaction, and (2) the expertise of forecasters.

Estimate the effect of collaboration on forecasting accuracy

The simulated-interaction method involves several participants generating each forecast, and Armstrong's (2001a) accuracy data for unaided-judgement forecasts were from pairs of participants. Collaboration requires forecasters to justify their forecasts to their fellows and allows forecasts to be combined. Both justification and combining tend to increase the accuracy of judgemental forecasts (Stewart, 2001). I examine the effect of collaboration on conflict forecasting accuracy for the methods unaided judgement and structured analogies.

Estimate the effect of expertise on forecast accuracy

Armstrong's (2001a) unaided-judgement forecast accuracy data were largely obtained from novices (mostly students) and he obtained simulated-interaction forecast accuracy data largely using student role players. Experts expect experts to be more accurate forecasters than novices (Table 1), although research by Armstrong (1980; 1991) and Tetlock (1992) suggest that this may not be so. I obtained game-theoretic and structured-analogies forecasts solely from experts as game-theoretic forecasting requires a knowledge of game theory and structured-analogies forecasting requires a knowledge of conflicts similar to a target conflict. On the other hand, it is feasible to ask novices to use their unaided judgement to forecast decisions in conflicts. Thus my first test of the effect of expertise was to assess whether non-game theorists who were experts in conflicts,

forecasting, judgement, or decision making tended to provide unaided-judgement forecasts that were more accurate than those provided by students.

Non-game theorists might or might not be domain experts in regard to particular conflicts. For example, a conflict expert might have industrial relations expertise and provide forecasts for conflicts in the industrial relations arena. My second test of the effect of expertise was to assess whether experts who had more experience with conflicts similar to a target conflict were more accurate than those who had less experience with similar conflicts.

My third test of the effect of expertise was to assess whether experts who had more years of conflict management experience were more accurate than those who had fewer years of such experience.

The source of analogies might have a bearing on their usefulness for forecasting using structured analogies. For example, analogies from direct experience may tend to lead to forecasts that are more accurate than those from indirect experience such as those from informal accounts, current affairs, history, or literature. If this were not the case, reading the newspaper and studying history are likely to be good substitutes for direct experience with this approach to forecasting. In my fourth test of the effect of expertise, I examined the effect of analogy source on forecast accuracy. I also examined the effect of the quantity and quality of analogies provided by forecasters on forecast accuracy.

Simulated interactions using role players who are similar to the real protagonists in a conflict may provide more accurate forecasts than those that use student role players. Surprisingly, however, the limited evidence that is available suggests that casting has little effect on simulated-interaction forecast accuracy (Armstrong, 2001a). Moreover, in practical applications of simulated-interaction forecasting, as in this research, the cost in time and money of obtaining representative role players is likely to limit their use. For these reasons, I have not examined the effect of casting on simulated-interaction forecast accuracy.

Those who complete several conflict forecasting tasks may be better or worse than they would have been had they forecast a single conflict. The data I have collected does not

allow me to distinguish between such an effect and any self-selection bias that might exist, and so I have not examined this matter.

1.3.4   Assess appeal to managers

The accuracy of forecasts from a method, although it is the most important, is not the only criterion used by managers to select a forecasting method. Yokum and Armstrong (1995) summarised the importance given by researchers, educators, practitioners, and decision-makers to thirteen forecasting method selection criteria. In order of importance to the Yokum and Armstrong participants, these are (slightly paraphrased from Armstrong, 2001c, p. 369):

1. Accuracy
2. Timeliness in providing forecasts
3. Cost savings resulting from improved decisions
4. Ease of interpretation
5. Flexibility
6. Ease in using available data
7. Ease of use
8. Ease of implementation
9. Ability to incorporate judgemental input
10. Reliability of confidence intervals
11. Development cost (computer, human resources)
12. Maintenance cost (data storage, modifications)
13. Theoretical relevance

Armstrong (2001c) suggested three additional criteria. These are:

1. Ability to compare alternative policies
2. Ability to examine alternative environments
3. Ability to learn (experience leads forecasters to improve procedures)

It seemed plausible that managers might weight the criteria differently when selecting a method for one particular forecasting purpose rather than another. In order to allow for this possibility, I obtained from experts criteria weights for the specific task of selecting methods for forecasting decisions in conflicts. I also obtained ratings for the four methods against the criteria.

1.3.5 Summary of objectives

The five objectives of my research on forecasting decisions in conflicts are summarised in Table 3 under the three broad headings of performance, generalisability, and appeal.

Table 3
Research objectives:
For reasonable methods, investigate the effect of…

Performance	❶ method on relative forecast accuracy
	❷ method on relative forecast usefulness
Generalisability	❸ collaboration on relative forecast accuracy
	❹ expertise on relative forecast accuracy
Appeal	❺ method characteristics on appeal to managers

2.   Prior evidence on methods[2]

I sought empirical evidence of the relative accuracy of forecasts of decisions in real conflicts for each of the four forecasting methods I examine in this research.

For the methods of game theory and structured analogies, I searched the Social Science Citation Index (SSCI) and the internet. I also sent appeals to relevant email lists asking for evidence and I communicated with leading researchers. The findings of these searches are presented at the end of the relevant sections.

2.1   Unaided judgement

Decision-makers can be subject to serious biases, or "blind spots". For example, decision-makers who are involved in a conflict tend to give "insufficient consideration of the contingent decisions of others", as is evidenced by phenomena such as winner's curse and non-rational escalation of commitment (Zajac and Bazerman, 1991, p. 50).

Researchers have demonstrated that unaided judgements can be biased by the role of the person making the judgement. Babcock, Loewenstein, Issacharoff, and Camerer (1995) asked participants to estimate a "fair" judgement in a dispute between two parties. Participants were given the role of lawyer for the complainant or for the defendant *before* being presented with identical briefing material. The estimates of "complainant lawyers" were, on average, higher than those of "defendant lawyers". The researchers found that the two groups had interpreted the same briefing material in different and self-serving ways. Similarly, participants who took on the roles of either "cost analyst" or "sales analyst" in research by Cyert, March, and Starbuck (1961) produced divergent forecasts from identical sets of numbers, depending on their role. Statman and Tyebjee (1985) replicated this research with consistent results.

The foregoing evidence suggests that a manager wanting a forecast for a conflict may benefit from asking people who are *not* involved in the conflict for their judgement on

---

2 A version of this literature review was published in Green (2002a). The article did not include evidence on analogies, nor did it contain the findings of SSCI and Internet searches conducted in 2002.

the likely outcome. Independent judges are also, however, subject to influences that lead to inaccurate forecasts. Experts, in particular, may be subject to overconfidence (Arkes, 2001), for example, or to biases resulting from the use of common and well-documented judgemental heuristics (for example, Bazerman, 1998).

Armstrong (2001a) provided evidence on the accuracy of independent judges' forecasts of decisions in conflicts. He found that student research participants performed no better than chance when exercising their unaided judgement to predict decisions made in conflicts in which they were not involved. Tetlock (1999) found that experts' (area specialists') predictions of the outcomes of political conflicts in the Soviet Union, South Africa, Kazakhstan, the European Monetary Union, Canada, the US presidential race of 1992, and the Persian Gulf crisis of 1990-91 were "only slightly more accurate than chance" (p. 351).

Overall, the evidence suggests that unaided judgement is unlikely to be a valid and reliable method for predicting decisions in conflicts.

2.2   Game theory

2.2.1   Others' reviews

Despite more than half a century of research, there is little evidence on the predictive validity of game theory for decisions made in real conflicts. In a review of all game theory articles published in the leading US operations research and management science journals, Reisman, Kumar, and Motwani (2001) found an average of less than one article per year involved a real-world application. In a review of Nalebuff and Brandenburger's book *Co-opetition* (1996), Armstrong (1997) wrote "I have reviewed the literature on the effectiveness of game theory for predictions and have been unable to find any evidence to directly support the belief that game theory would aid predictive ability" (p. 94).

Evidence that is available tends to be indirect and incomplete, typically comparing game-theoretic predictions with the outcomes of context-poor experiments using role players rather than with real-world conflicts. In a search for evidence on the accuracy of game-theoretic predictions, Armstrong and Hutcherson (1989) found two studies

(Eliashberg, LaTour, Rangaswamy, and Stern, 1986; and Neslin and Greenhalgh, 1983) that used game theory to predict the outcomes of negotiations. In both studies role-play negotiations, rather than actual negotiations, were the benchmark – the implication being that the outcomes of the role-play experiments (which were imperfectly predicted by game theory) were equivalent to actual negotiation outcomes.

Experiments such as those described above have been widely used by game-theory researchers. For example, Shubik (1975) noted that "experimental gaming" is employed to examine the "validity of various solution concepts" developed by game theorists (p. 20). Rapoport and Orwant (1962), in a comprehensive review of the use of experimental games to test game-theory hypotheses, concluded that "game theory is not descriptive and will not predict human behavior, especially in games with imperfect information about the payoff matrices".

Bennett (1995, p. 27) suggested that classic game-theory models lack four aspects that are present in real conflicts, namely: differing perceptions, dynamics, combinatorial complexity, and linked issues. Attempts have been made to extend game theory in order to address its shortcomings. Two of these extensions are "hypergame analysis" (Bennett and Huxham, 1982), and "drama theory" (Howard, 1994a and b). Hypergames are intended to account for players' divergent perceptions by describing and analysing a set of subjective but linked games. Drama theory seeks to incorporate emotion into the analysis of conflict situations. While these developments may facilitate greater realism than classic game theory – and perhaps, therefore, greater predictive accuracy – in order to test drama theory originated hypotheses about behaviour, Bennett and McQuade (1996), for example, used role-play experiments for their benchmark, rather than real conflicts.

2.2.2   Social Science Citation Index search

Searches of the SSCI were conducted for the period 1978 to 7 July 2001, and on 4 December 2002 for the intervening period. Eleven articles were found using the phrases "game theory" and "forecasting". A further 33 articles were found when the term "prediction" was substituted for "forecasting" – 44 articles in total. Six turned up in two of the searches. Eight were about animal, rather than human, behaviour and three were

concerned with artificial intelligence or similar. Two were of work by me that is described in this document (Green, 2002a; Green, 2002b) and five were commentaries on my work.

The 20 remaining articles were: Austen-Smith and Banks (1998); Batson and Ahmad (2001); Blume, DeJong, Kim, Sprinkle (2001); Carayannis and Alexander (2001); Diekmann (1993); Ghemawat and McGahan (1998); Ghosh and John (2000); Gibbons and Van Boven (2001); Gruca, Kumar, and Sudharshan (1992); Henderson (1998); Jehiel (1998); Keser and Gardner (1999); McCabe and Smith (2000); McCarthy (2002); Sandholm (1998); Scharlemann, Eckel, Kacelnik, Wilson (2001); Schwenk (1995); Sonnegard (1996); Sugiyama, Tooby, and Cosmides (2002); Suleiman (1996).

The evidence provided by these articles is discussed in subsection 2.2.6.

2.2.3   Internet search

A search of the internet on 10 December 2002 using the Google™ search engine and the terms that were used for the SSCI search produced an unusable number of hits. A search using the phrases "comparative", "forecasting accuracy", and "game theory" yielded 25 unique hits. One was a copy of a draft of my own paper (Green, 2002a). Another site required a password to gain access. Nineteen were of lists of various types with the search terms dispersed among different items.

The balance of four hits were of: Leeflang and Wittink (2000); Schrodt (2002); Smith (1994); and Tesfatsion (2003 – forthcoming).

The evidence provided by these articles is discussed in subsection 2.2.6.

2.2.4   Appeal for evidence

I sent an email message to 474 email addresses of game theorists asking for empirical evidence on the predictive validity of game theory for real conflicts. I received 18 responses. One response was an "invalid address" message and five were automatic

messages stating that the addressees were on leave, or similar. A respondent claimed to have no relevant expertise and four asked for more information about the research or stated that they would look at the material later.

Seven respondents commented on my preliminary research findings or provided information on game theory. Four of the seven did not address predictive validity. One of the seven referred to the "huge literature on experimental economics that shows under what conditions game theory with the 'rational actor assumptions' works, and where it does not", but did not suggest specific works. Another suggested his game-theoretic analysis of the civil conflict in Northern Ireland was evidence of predictive validity for a game theory variation (Brams and Togman, 2000). He stated "I think our predictions… have for the most part been borne out by events"[3]. The same volume also contains claims of accuracy for a model using game-theoretic and decision-theoretic analysis for forecasting aspects of the civil conflict over Jerusalem (Organski, 2000). Finally, one of the seven respondents suggested I look at his work (Walker and Wooders, 2001) for evidence.

The evidence provided by these articles is discussed in subsection 2.2.6.

2.2.5 Personal communications

As a result of a request for information on the use of analogies for forecasting (described below) I was informed about the use of the game-theoretic models of Bruce Bueno de Mesquita for forecasting conflicts. I contacted Bueno de Mesquita[4] and asked him for evidence on the relative accuracy of his expected utility models for forecasting conflicts. He referred me to Stanley Feder of Policy Futures, Frans Stokman of the University of Groningen, and to Bueno de Mesquita and Stokman (1994). I sent email messages to both Feder and Stokman on 4 December 2001. Stokman did not reply to my request for information. Feder referred me to the work of Fraser and Hipel (1984) on the use of game theory to forecast conflict outcomes[5] and to his own work (Feder, 1987).

---

3 Personal communication with Steven Brams, 13 June, 2001.

4 Personal communication with Bruce Bueno de Mesquita, 4 December, 2001.

5 Personal communication with Stanley Feder, 5 December, 2001.

Fraser did not respond to an email request for information sent on 5 December. An inspection of Fraser's publication list[6] found one title which included the word "forecast": Fraser (1986). Both Fraser (1986) and Fraser and Hipel (1984) promoted the game-theoretic technique "conflict analysis" as a forecasting method.

The evidence provided by these articles is discussed in the next subsection.

---

6 From http://www.openoptions.com/publications.htm, 5 December 2001.

2.2.6 Search findings

As a result of my SSCI and internet searches, together with my requests for information and communication with researchers, I found a total of 31 unique articles that either included the search terms and met basic relevance criteria or were recommended as offering evidence on the relative accuracy of game-theoretic forecasts of decisions in real conflicts. The content of these articles that is relevant to my search for evidence is summarised in Table 4.

Table 4
**Content of articles found in searches for evidence on
the relative accuracy of game-theoretic forecasts
of decisions in real conflicts**

Study	Game theory	Empirical findings	Real conflict(s)	Specific forecasts	Other methods
Ghosh & John (2000)	✗				
Schwenk (1995)	✗				
Tesfatsion (2003 – forthcoming)	✗				
Austen-Smith & Banks (1998)	✓	✗			
Carayannis & Alexander (2001)	✓	✗			
Henderson (1998)	✓	✗			
Jehiel (1998)	✓	✗			
Leeflang & Wittink (2000)	✓	✗			
McCarthy (2002)	✓	✗			
Sandholm (1998)	✓	✗			
Schrodt (2002)	✓	✗			
Batson & Ahmad (2001)	✓	✓	✗		
Blume, et al. (2001)	✓	✓	✗		
Diekmann (1993)	✓	✓	✗		
Gibbons & Van Boven (2001)	✓	✓	✗		
Keser & Gardener (1999)	✓	✓	✗		
McCabe & Smith (2000)	✓	✓	✗		
Organski (2000)	✓	✓	✗		
Scharlemann, et al. (2001)	✓	✓	✗		
Sonnegard (1996)	✓	✓	✗		
Smith (1994)	✓	✓	✗		
Sugiyama et al. (2002)	✓	✓	✗		
Suleiman (1996)	✓	✓	✗		
Walker & Wooders (2001)	✓	✓	✗		
Brams & Togman (2000)	✓	✓	✓	✗	
B. d. Mesquita & Stokman (1994)	✓	✓	✓	✓	✗
Feder (1987)	✓	✓	✓	✓	✗
Fraser (1986)	✓	✓	✓	✓	✗
Fraser & Hipel (1984)	✓	✓	✓	✓	✗
Ghemawat & McGahan (1998)	✓	✓	✓	✓	✗
Gruca, et al. (1992)	✓	✓	✓	✓	✗

Three of the articles (Ghosh and John, 2000; Schwenk, 1995; Tesfatsion, 2003 – forthcoming) turned out not to be about game theory. Eight of the articles (Austen-Smith and Banks, 1998; Carayannis and Alexander, 2001; Henderson, 1998; Jehiel, 1998; Leeflang and Wittink, 2000; McCarthy, 2002; Sandholm, 1998; Schrodt, 2002) while

being concerned with game theory, did not provide empirical findings. One of these authors, in his article about alternative approaches to understanding and predicting crime, wrote "Game theory models provide considerable insights into the complexities of illegal decisions but these are mostly untested" (p. 437, McCarthy, 2002). Further, "…although several studies… reach the same conclusions as game theory research, there is little overlap between game theory models of crime and empirical studies of offending" (p. 437). In his review of forecasting for foreign policy, Schrodt (2002) found no evidence on the relative accuracy of game-theoretic forecasts, or of forecasts from any other method.

Thirteen of the 31 articles provided empirical findings on game-theoretic predictions. The predictions, in the case of 11 of the articles, were not of decisions made in real conflicts but of the outcomes of context-poor experiments. These were: Batson and Ahmad (2001); Blume et al. (2001); Diekmann (1993); Gibbons and Van Boven (2001); Keser and Gardener (1999); McCabe and Smith (2000); Sonnegard (1996); Scharlemann et al. (2001); Smith (1994); Sugiyama et al. (2002); Suleiman (1996). Organski (2000) compared game-theoretic predictions for a single conflict with what *might* have happened had Israeli Prime Minister Rabin not been assassinated. Walker and Wooders (2001) compared their game-theoretic predictions with behaviour in a competitive sport – professional tennis.

One article (Brams and Togman, 2000), was concerned with the conflict between the Irish Republican Army and the British government over Northern Ireland. The authors' used their game-theoretic analysis to explain what had happened, but did not provide evidence of having made predictions in ignorance of the events that later transpired.

Six articles included evidence of specific game-theoretic forecasts of decisions in real conflicts. None of the six, however, made legitimate comparisons between the accuracy of the game-theoretic forecasts and the accuracy of forecasts from other plausible methods. Fraser (1986), and Fraser and Hipel (1984) did not compare the accuracy of their game-theoretic forecasts with the accuracy of any other forecasts. Bueno de Mesquita and Stokman (1994), and Gruca et al. (1992) compared the accuracy of forecasts from their game-theoretic models only with the accuracy of forecasts from other game-theoretic models. Ghemawat and McGahan (1998) compared their forecasts with naïve forecasts, which they called non-strategic analysis. Finally, Feder (1987)

compared the accuracy of forecasts from a game-theoretic model with the forecasts of experts. On the face of it, the experts were as accurate as the game-theoretic model. Feder pointed out that their forecasts were not so specific but, as the experts had not been asked to be more specific than they were, comparisons are not legitimate.

In sum, prior to the research presented here, there was no coherent body of evidence available on the accuracy of game-theory forecasts relative to that of forecasts from other methods that are in use or are recommended.

2.3   Structured analogies

Despite the recommendations made by various authors (subsection 1.2.3) I was unable to find direct evidence on the predictive validity of using analogies to forecast decisions made in real conflicts. My search for evidence, and the findings of the search, are described in the next five subsections.

2.3.1   Social Science Citation Index search

Searches of the SSCI were conducted for the period 1978 to 7 July 2001, and on 4 December 2002 for the intervening period. Six articles that concerned human behaviour were found using the phrases "analogies" and "forecasting". A further six articles concerned with human behaviour were found when the term "prediction" was substituted for "forecasting" – 12 articles in total.

The articles were: Adam and Moodley (1993); Bernstein, Lebow, Stein, and Weber (2000); Bowander, Muralidharan, and Miyake (1999); Castro, Lubker, Bryant, and Skinner (2002); Glantz (1991); Graham (1991); Kadoda, Cartwright, and Shepperd (2001); Liu, Pham, and Holyoak (1997); Mildenhall and Williams (2001); Pfister and Konerding (1996); Souder and Thomas (1994); Tetlock (1992).

The evidence provided by these articles is discussed in subsection 2.3.5.

2.3.2   Internet search

A search of the internet on 11 December 2002 using the Google™ search engine and the phrases "comparative", "forecasting accuracy", and "analogies" yielded 10 unique hits. A search with the word "analogy" substituted for "analogies" yielded a further 21 unique hits. Of the 31 hits, seven were of lists of various types with the search terms dispersed among different items.

Another four hits were of websites with subjects that were unrelated to the purpose of the search. One of these was a company website (www.theplanningbusiness.com/compgdp.htm) that referred to the "famous 'butterfly' analogy". A second was a summary of an oral submission on industrial relations policy that used the word "analogy", but not in relation to forecasting: www.workplace.gov.au/WP/CDA/files/WP/WR/appendix_H_98_99.pdf. A third was titled "Historical overview of transportation planning model development" that referred to the "gravitational analogy": www.ctre.iastate.edu/Research/multimod/Phase1/II.htm. A fourth was a US Department of Transportation "Manual for regional transportation modelling practice for air quality analysis" that also referred to the "gravitational analogy": http://tmip.fhwa.gov/clearinghouse/docs/airquality/mrtm/ch3.stm.

Finally, I could not gain access to another of the hits: www.tiff.org/pub/library/QR_Archive/ 1996/1996_3Q_MUT_FUND.pdf.

The balance of 19 items were articles (some hits were pre-publication versions of the items listed) or unpublished documents. They were: Anastasakis and Mort (2001); Armstrong (2001g); Armstrong and Brodie (1999); Armstrong and Collopy (1998); Beatty, Riffe, and Thompson (1999); Beer (c2000); Cannon and Reed Consulting Group (1999); Chambers, Mullick, and Smith (1971); Chang (1999); Gonzalez (2000); Lawson (1998); Mentzas (1997); Rey (2000); Sanders and Ritzman (2001); Schrodt (2002); Tesfatsion (2002); Tesfatsion (2003 – forthcoming); Winkler (1983); Wong and Tan (c1992).

The evidence provided by these articles is discussed in subsection 2.3.5.

2.3.3   Appeal for evidence

On 28 November 2001 I sent emails to the International Institute of Forecasters listserver[7] and to the Judgement and Decision Making mailing list[8] including the same request. I received three relevant replies. One respondent suggested that Gentner, Holyoak, and Kokinov (Eds.) (2001) might include some useful material. An inspection of the contents pages and index failed to find any reference to "forecasting" or "prediction". Another of the respondents suggested contacting Bruce Bueno de Mesquita, founder of Decision Insights Inc. and senior fellow at the Hoover Institution, about his conflict forecasting models. These models are game-theoretic in nature, and this line of enquiry was discussed previously. Finally, the third respondent suggested that the field of case-based reasoning might provide some evidence.

Case-based reasoning

In the online Forecasting Dictionary (op. cit., 8 January 2003), case-based reasoning (CBR) is described as follows

> Information on situations (cases) is stored with the purpose of recalling cases that are similar to a target problem in order to help solve the problem. People commonly use this approach informally in problem solving and forecasting (See analogy). It can also be used as the basis for designing expert systems by starting with examples rather than with the process. CBR is a term used in the fields of cognitive science and artificial intelligence. The forecasting method of structured analogies could be viewed as one type of CBR. We have been unable to locate any tests of the predictive validity of CBR.

The respondent who suggested looking for evidence in the case-based reasoning literature referred me to his course outline (http://www.cs.rpi.edu/courses/fall01/soft-computing/) for a review of the field. I found no evidence in the course outline on predictive accuracy for decisions in conflicts. He also referred me to a listserver (members@ai-cbr.org). On 8 May 2002, I sent an appeal to the listserver asking "Are any of you aware of any empirical evidence on the accuracy of forecasts of decisions in conflicts from using analogies, relative to the accuracy of forecasts from other

---

7 278 addresses subscribed as at 20 July, 2001.

8 579 addresses subscribed as at 20 December, 2001.

methods?". I received two replies. One of the respondents wrote seeking, rather than offering, information. The other referred me to "Loui et al [on] defeasible reasoning" and suggested that there is "a lot of literature in AI and law… that would be directly relevant". I found a list of Loui's publications on the internet[9]. From the titles of the publications, it was clear that they were concerned with explaining and modelling reasoning (particularly legal reasoning using AI) rather than with forecasting in conflicts. Moreover, publications by this author had not turned-up in my earlier internet or SSCI searches for evidence on forecasting with analogies. I did not pursue this line of enquiry further.

Finally, I examined the contents of a widely used case-based reasoning textbook for evidence (Kolodner, 1993), and could find none.

2.3.4   Personal communications

On 27 November 2001 I sent an email letter to Professors Neustadt and May – the authors of "Thinking in Time: The uses of history for decision makers" (1986) – asking if they were aware of any evidence on the accuracy of forecasts of decisions in conflicts that were derived from analogies, relative to the accuracy other forecasting methods. Neustadt and May teach the use of analogies for forecasting conflicts. I received no reply.

---

9 http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/l/Loui:Ronald_Prescott.html on 9 January 2003.

2.3.5 Search findings

I found a total of 31 unique articles from my searches of SSCI and the internet, and from requests for information to listservers and communications with researchers. The items either included the search terms and met basic relevance criteria, or were recommended as offering evidence on the relative accuracy of analogical forecasts of decisions in real conflicts. The content of these articles that is relevant to my search for evidence is summarised in Table 5.

Table 5
**Content of articles found in searches for evidence on
the relative accuracy of analogical forecasts
of decisions in real conflicts**

Study	Real conflict(s)	Analogies	Empirical findings	Specific forecasts	Other methods
Anastasakis & Mort (2001)	✕				
Armstrong (2001g)	✕				
Armstrong & Brodie (1999)	✕				
Armstrong & Collopy (1998)	✕				
Beatty, et al. (1999)	✕				
Beer (c2000)	✕				
Bernstein, et al. (2000)	✕				
Bowander, et al. (1999)	✕				
Cannon & Reed Consulting (1999)	✕				
Castro, et al. (2002)	✕				
Chambers, et al. (1971)	✕				
Chang (1999)	✕				
Glantz (1991)	✕				
Gonzalez (2000)	✕				
Graham (1991)	✕				
Kadoda, et al. (2001)	✕				
Lawson (1998)	✕				
Liu, et al. (1997)	✕				
Mentzas (1997)	✕				
Mildenhall & Williams (2001)	✕				
Pfister & Konerding (1996)	✕				
Rey (2000)	✕				
Sanders & Ritzman (2001)	✕				
Souder & Thomas (1994)	✕				
Tesfatsion (2002)	✕				
Tesfatsion (2003 – forthcoming)	✕				
Winkler (1983)	✕				
Wong & Tan (c1992)	✕				
Adam and Moodley (1993)	✓	✓	✕		
Schrodt (2002)	✓	✓	✕		
Tetlock (1992)	✓	✓	✓	✓	✕

Unlike game theory, which is concerned with conflict, the term "analogy" is used widely. It is not surprising then that the great majority of the items I identified from my searches (28) were not concerned with real human conflicts. There seemed little point in

including the term "conflict" among the search terms. The term is not consistently used when conflicts are being discussed and might have led to evidence being overlooked.

The balance of three items were concerned with real conflicts and with the use of analogies for forecasting their outcomes. One of these, Adam and Moodley (1993), was concerned with the internal political situation in South Africa. The authors used analogies to forecast the likely outcomes of policies that might be adopted by the South African government. They did not, however, present evidence on the accuracy of those forecasts.

Schrodt (2002) looked for empirical evidence on the relative accuracy of analogies for forecasting decisions in foreign policy conflicts, as he had done for game-theoretic forecasts, and found none.

Finally, Tetlock (1992) found that, on average, experts' forecasts of the outcomes of conflicts in international politics were no more accurate than chance. He showed, after further analysis, that both experts who attempted to avoid common human biases (including reliance on "facile historical analogies") and those who referred to fundamental processes, tended to be somewhat more accurate than those experts who appeared to do neither of these things. The small sample of 20 forecasts means, however, that there can be limited confidence in this finding.

In sum, prior to the research presented here, there was no coherent body of evidence available on the accuracy of forecasts from a structured use of analogies relative to the accuracy of forecasts from other methods that are in use or are recommended.

2.4   Simulated interaction

2.4.1   A review

Compared to the research efforts on game theory and unaided judgement, there has been little research on simulated interaction as a method for forecasting decisions. Armstrong (2001a) recently presented the findings of a comprehensive review of the evidence on the relative accuracy of forecasts from role playing (simulated interaction). He found

that simulated interaction provided more accurate forecasts (52 percent correct) across a sample of four actual conflicts involving interaction than did the unaided judgement of students (11 percent) and than could be expected by chance (25 percent). The four conflicts were diverse in their characteristics (Table 6).

Table 6
Armstrong's (2001a) evidence on the accuracy of simulated-interaction decisions and unaided judgement forecasts by students
Percent correct forecasts (number of forecasts)

	Chance	Unaided judgement		Simulated interaction	
Distribution Channel	33	3	(37)	**75**	(12)
Artists Protest	17	3	(31)	**29**	(14)
Journal Royalties	25	12	(25)	**42**	(24)
55% Pay Plan [a]	25	27	(15)	**60**	(10)
Totals (unweighted) [b]	25	11	(108)	**52**	(60)

a Chance for 55% Pay Plan was misreported in Armstrong and Walker (1983) as 33% and this error was perpetuated in subsequent reports of the study's findings (personal communication with J. Scott Armstrong, 11 July 2000)
b Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

As the data in Table 6 show, simulated-interaction forecasts were more accurate than students' unaided-judgement forecasts and chance for each conflict.

Armstrong (2001a) also presented informal evidence, from others' accounts, on the accuracy of individual simulated-interaction forecasts relative to that of forecasts by experts using their unaided judgement. The simulated-interaction forecasts were more accurate.

The research summarised in Armstrong (2001a) and presented in Table 6 represents the only coherent body of evidence available, prior to the research presented here, on the accuracy of simulated-interaction forecasts relative to that of forecasts from other methods that are in use or are recommended.

3.   Research programme

This chapter describes my research programme in four sections. Section 3.2 describes the four forecasting methods as they could conveniently be implemented to address real problems of forecasting decisions in conflicts. The descriptions each assume that a manager (such as a corporate chief executive, political party leader, ambassador, or general) is embroiled in a conflict and wishes to choose a strategy that is likely to result in the best outcome for his or her party. Section 3.3 describes how I selected eight real conflicts for this research. Section 3.4 describes how the forecasting methods were implemented in order to fulfil the objectives of the research. Finally, section 3.5 describes how I collected experts' opinions on the appeal of the methods.

Before describing the research programme I describe, in section 3.1, my reasons for adopting the approach that I did.

3.1   Approach

I designed this research programme with the objective of developing useful recommendations, or principles, for managers who face the problem of forecasting decisions in conflicts. In order to do this, I endeavoured to match the research design to managers' conflict forecasting problems and to feasible methods for addressing those problems. That is, I sought to use actual conflicts and to use forecasting methods that the managers who faced them would have been able to employ.

In essence, the research task was to evaluate contending forecasting methods for conflicts (subsection 1.3.1). In doing so I drew on the resources of the Principles of Forecasting project (www.forecastingprinciples.com). In particular, I sought to build on the research described in "Role playing: a method to forecast decisions" (Armstrong, 2001a), the first chapter in *Principles of forecasting: A handbook for researchers and practitioners*. I was guided by the principles described in the chapter "Evaluating forecasting methods" (Armstrong, 2001e). The former summarised the state of knowledge on forecasting in conflicts and the latter described best practice for evaluating forecasting methods.

3.2   Conflict forecasting methods described

3.2.1   Unaided judgement

Unaided judgement could be used in various ways for forecasting the decision that will be made in a conflict. A manager may, for example, simply think hard about a conflict and make a prediction; discuss the conflict with colleagues and then make a prediction; or reach a consensus forecast after discussion with colleagues. A manager may go to greater lengths and have a paper prepared that describes the conflict situation and the parties involved, have a meeting to discuss the problem, and then make a prediction. Finally, a manager may seek the help of people who have experience with, or knowledge of, similar problems – for example, people within the manager's organisation, consultants, or management school professors – and ask them to make predictions using their judgement. These approaches can all be regarded as forecasting using unaided judgement. That is, using judgement without recourse to a formal method.

3.2.2   Game theory

Recall that, for the purpose of this research, game theory is what game-theory experts do when asked to use their expertise to predict decisions in real conflicts. Although few managers are likely to be game-theory experts themselves, some will be aware of the promises made of game theory from, for example, business school courses, the media, or popular books such as Nalebuff and Brandenburger (1996).

A search of the Businessweek Online archive (businessweek.com, 14 January 2003) supported this contention. The search, using the term "game theory", resulted in 26 hits dated from 29 July 1996 to 6 January 2003. Five of the 10 were accounts of business school experiences. One was a set of book recommendations. Another was a commentary on conflict in the Middle East, and a third was a commentary on the US Electoral College. Finally, two were subscriber-only articles whose subject matter was not clear from the title and text provided by the search engine.

Managers who are not game-theory experts themselves, but who wish to employ game theory, would be likely to use the services of people who are. A manager may have

experts (consultants) briefed on the situation and ask them to make a prediction. Alternatively, managers may find a game-theoretic approach is recommended to them by consultants. For example, London (2002) interviewed people from various management consulting firms about their use of game theory in their advice to clients. In either case, experts would use their professional judgement to determine the approach they would take to forecasting a decision in the target conflict. In particular, the experts may or may not develop a formal game-theory model.

3.2.3   Structured analogies

Analogous situations may or may not be recalled in the course of using unaided judgement to derive forecasts for decisions in conflicts. The structured-analogies method, however, involves managers adopting the following, or a similar, formal approach:

- First, recall and describe conflicts that are analogous to the target conflict
- Second, for each analogous conflict, list similarities and differences vis-à-vis the target conflict and rate the analogy for quality – the extent to which it is similar to the target
- Third, determine the decision that is implied for the target conflict by each of the analogies
- Finally, combine the individual forecasts (implied decisions) using either judgement or arithmetic in order to choose a single decision or to estimate the probability of potential decisions.

In the case of a probabilistic forecast, the preferred forecast would be the most probable decision, although the manager may take account of other, less probable, decisions. The manager may choose to involve colleagues in this process or ask experts (people who are likely to be familiar with analogous situations) to use this method to forecast the conflict. There seems little point in asking anyone who is unlikely to be familiar with relevant analogies (who is not an expert) to use this method.

3.2.4 Simulated interaction

A manager who wishes to use this method may recruit colleagues to take on the roles of the main players in a target conflict, and ask them to simulate future interactions between the parties. Such role players are likely to know a lot about the situation and the people whose roles they are playing. Nevertheless, this approach to the simulated-interaction method has at least two potential drawbacks. First, role players may find it difficult to take on the role of, for example, their boss, or an enemy. Second, it is desirable to have several independent forecasts and many organisations will find it inconvenient to take the large numbers of people that are required away from their normal duties. This is particularly the case if different strategies or policies are to be tested.

The evidence shows that the outcomes of simulated interactions using university student role players can provide accurate forecasts of decisions made in conflicts (Table 6). The services of university students can be obtained cheaply and they are available in large numbers. Given the evidence on accuracy and the relative advantages of using students, it would be reasonable to do so in many circumstances. Although there is no direct evidence that the representativeness of role players has an effect on forecast accuracy, it seems sensible to implement any easily achieved increases in representativeness by, for example, allocating participants to roles using salient characteristics of the individuals or self selection. If confidentiality is important, the description of the target conflict can often be disguised and, if this is not feasible, simulated interactions could be conducted in a country where the situation is unlikely to be recognised or using role players who are sworn to secrecy.

The description of the role playing procedure provided by Armstrong (2001a) and summarised in the *Forecasting Dictionary* entry (Armstrong, 2001g) also applies to simulated interaction. That procedure was followed in this research for the simulated-interaction treatment. The simulated-interaction method involves the following steps:

- First, role players read a brief description of their role and then a description of the unresolved conflict situation they, in their roles, face
- Second, role players are told to stay in-character and to improvise as necessary, so long as their behaviour is consistent with the information

they have been given. They are told what is expected from their simulation – for example a decision or agreement. They may be provided with a list of potential decisions that have been determined by experts to be a complete list of mutually exclusive options

- Third, the role players simulate the conflict and
- Fourth, choose from the list the decision that most closely matches the outcome of their simulation.

If role players run out of time, they are asked to choose the outcome closest to the outcome they believe would have occurred had they been able to finish their simulation. Several, perhaps ten, independent groups of role players should be used to simulate the target conflict. (A group of role players is no more and no fewer than the role players who are needed to play the all of the roles specified in the conflict description). Alternative strategies can be tested using other groups of role players by varying the descriptions of the roles and the conflict.

3.3   Conflict selection and description

The conflicts I used in the research and their sources are described in subsection 3.3.1. In subsection 3.3.2, I assess the conflicts' diversity and, specifically, whether the conflicts include types of conflict over which game theorists claim expertise. I chose diverse conflicts so as to maximise the number of managers for whom at least one of the conflicts was likely to be relevant and in order to make legitimate the generalisation of findings to the class of conflicts that are the subject of this research.

In subsection 3.3.3, I describe the material that I provided to participants.

3.3.1   Conflicts selected

There were eight conflict situations used in this research. All were either unlikely to be recognised by participants, or disguise without distortion was possible. I first chose three conflicts for which forecast accuracy data were available (Armstrong, 2001a). These

conflicts are referred to in this document as Artists Protest, Distribution Channel, and 55% Pay Plan. Armstrong provided me with copies of the descriptions of these conflicts.

Artists Protest was a conflict between Dutch artists and their national government over financial support. The conflict occurred in the late-1960s. By that time, the Dutch government had for approximately 20 years bought the works of artists who met certain criteria and were not otherwise able to sell their artworks. Those artists who had been accepted for the scheme were supported in this way for up to one year. Artists had become upset at the difficulty of making a good living from their art. The artists' union expressed this sentiment by occupying a museum room containing some of the nation's major art treasures and demanding relaxed entrance requirements for the scheme together with unlimited tenure. The decision to be made was what, if any, changes to make to the scheme. Armstrong (1987) based his description on a newspaper report (Newman, 1982).

Distribution Channel involved a proposal by Philco Corporation for commercial co-operation that required decision-makers to trade-off conflicting interests. Interests such as the parties' shares of the expected return in relation to the risk involved. Philco was a major US appliance manufacturer. In 1961, after a period of recession and with increased competition, Philco was in poor financial shape. In response to this, the corporation's managers approached a supermarket chain with a novel distribution proposal: the Cash Register Tape Plan. The proposal called for Philco dealers to sell appliances out of supermarkets. Supermarket customers who purchased an appliance on instalment would receive discounts on their monthly payments proportionate to their supermarket spend. The cost of the discount was to be split between Philco and the supermarket. The decision to be made was whether the supermarket chain would adopt the Philco proposal long-term, for a short trial, or not at all. The version of the questionnaire that I used included a fourth option of *either* long-term adoption *or* a short trial. For the sake of simplicity, I scored either/or responses as 0.5 for the purpose of proportion correct calculations and defined chance as 33 percent. Armstrong (1987) based his description on one case from a book of marketing case histories (Berg, 1970).

55% Pay Plan was a conflict in the USA between National Football League owners and players over revenue shares. The collective bargaining agreement between players and owners was due to expire in June 1982 and the head of the players' union stated that the

players' "bottom line" was 55 percent of gross revenue. The executive director of the Management Council rejected the possibility of a revenue share deal, no matter what the percentage. Estimates of the proportion of gross revenue accounted for by players' remuneration at that time ranged between 25 percent and 45 percent. The decisions to be made were whether to strike or not and, if to strike, for how long. Armstrong (1987) based his description on two reports in a sports periodical published prior to the start of negotiations (Boyle, 1982; Kirshenbaum, 1982) and on a personal communication from an insurance broker whose firm offered strike insurance to players.

Armstrong (2001a) provided data on the accuracy of forecasts from unaided judgement by novices and from simulated interaction for these three conflicts. In all, there were 83 judgemental forecasts made by 166 participants, and 36 simulated-interaction forecasts from 72 participants reported for these conflicts.

Secondly, I chose five additional conflicts and wrote descriptions of them. The five are referred to in this document as: Nurses Dispute, Personal Grievance, Telco Takeover, Water Dispute, and Zenith Investment.

Nurses Dispute was a conflict over pay between the nursing staff and management of Capital Coast Health, a government-owned organisation operating hospitals in Wellington, New Zealand. Nurses went on strike angry that they were being offered a much lower pay rise than had been given to intensive care nurses and to junior doctors. A mediator was appointed by a government agency – the Mediation Service. The principal decision to be made was the size of the pay rise.

I became aware of the conflict through reports in the local newspaper. I obtained information on the dispute, firstly, from that press coverage (Langdon, 2000a; 2000b; 2000c) and from transcripts of radio interviews (Radio New Zealand Limited, 2000a; 2000b; 2000c). Secondly, and most importantly, I obtained information from interviews with the principal negotiators for the two parties. I asked the nurses' principal negotiator how the description could be improved. He did not have any suggestions and considered the material a fair and accurate representation. In order to help ensure that the description would be readily understood by participants and interpreted as I had intended, I asked a colleague to read the description and provide suggestions on the writing. I then questioned him on his understanding. The colleague was an expert in

judgement and decision making. I adopted most of his suggestions and changed the description where his understanding was at odds with my intentions. When I first tested the Nurses Dispute material I did not provide participants (theatre student role players) with decision options, but asked them to record the agreed percentage pay increase. This was not a successful approach as the role-players made settlements that were more favourable to the nurses than the nurses' own negotiation starting position. As a consequence, I rewrote the material to provide three decision options to forecasters. These options were, in essence: the nurses' claim prevails, the employer's position prevails, and a compromise is reached.

I chose to include Nurses Dispute in my research because, first, I was able conveniently to get information on the conflict from: reading newspaper reports, talking to the writer of the reports, and interviewing the principal negotiators. Second, the conflict was different from previously chosen conflicts (diversity is examined in subsection 3.3.2). Third, I was able to compile my material before the parties reached an agreement and was thereby able to avoid the possibility of bias in my description from such knowledge.

Personal Grievance was a conflict over the importance of an employee's role and the pay scale for that role. A long-serving staff member of a New Zealand university students association believed her work was undervalued in the job evaluation that was commissioned by her new manager. The evaluation was conducted by the Association President. After some negotiation, the top of the salary band for the employee's position, set by the manager, was still below her current salary. The manager did not propose reducing the employee's actual salary, but it was clear she could not expect a pay increase in the foreseeable future. A Mediation Service mediator was appointed and a meeting between the parties was arranged. Decisions to be made were whether to commission a new independent evaluation, and whether to accept the salary band.

All information about the situation came from interviews with the Association manager, the staff member, and the Mediation Service mediator. The three commented on drafts of my situation description and made suggestions as to what decisions might have been made at the meeting. In order to help ensure that the description would be readily understood by participants and interpreted as I had intended, I asked three colleagues to read the description, use their judgement to predict the decision that was made, and to provide suggestions on the writing. I then questioned them on their understanding. The

three were experts in judgement and decision making. None recognised the conflict. I adopted most of their suggestions and changed the description where their understanding was at odds with my intentions. I compiled the description and decision options for a research project on the effect of mediation and information that was commissioned by the Employment Relations Service of New Zealand's Department of Labour (Green, 2002c). Employment Relations Service officials identified the conflict as one that would meet their requirements and mine, and put me in touch with the people involved. The Employment Relations Service had previously commissioned research from me on the effect of mediation. I also used the Nurses Dispute, which I had compiled prior to being commissioned by the Employment Relations Service, in that research (Green, 2001). When I tested the Personal Grievance material, I provided participants (student role players) with 11 decision options. The options were intended to meet the analysis preferences of Employment Relations Service officials. In the event, the decisions of the participants in the tests were distributed across the options to the detriment of useful forecasting and analysis. As a consequence of this outcome, I and the officials agreed to a set of four, more distinct, decision options including a failure-to-reach-agreement option. The four options were used in the research reported here. The grievance continued as I was compiling and testing the material and, despite testing and reviewing, some ambiguity remains over which of the options most accurately represents the final outcome. Two, perhaps three, of the four options could reasonably be interpreted as at least partly accurate representations of the actual outcome. Rather than rely on my own judgement to determine which, if any, of the options could be regarded as the actual outcome for the purpose of proportion correct calculations, I adopted the option that received the highest median usefulness rating from a panel of independent experts (subsection 4.1.2). The option selected in this way coincided with my own assessment of which option was the most accurate.

I chose to include Personal Grievance in my research because, first, I was able conveniently to get the information I needed by interviewing the parties involved in the conflict. Second, the conflict was different from previously chosen conflicts. Third, I was able to compile my material before the matter in dispute was completely resolved.

Telco Takeover was a conflict for the ownership of a regional telecommunications provider (CenturyTel) that occurred in the USA during 2001. Alltel, a larger telecommunications company, had been approached by CenturyTel managers with an

offer to sell Alltel their mobile telephone business. Alltel managers declined the offer. (In the case of both corporations, senior managers dominated the boards of directors). Shortly afterwards, Alltel made an offer to pay 40 percent more than the current share price to buy all of CenturyTel. CenturyTel's long-standing chairman was a substantial shareholder of the company. Managers and staff also owned shares. The CenturyTel board was reluctant to sell and took measures to prevent an Alltel take-over. Alltel appealed directly to outside shareholders of CenturyTel. The decision to be made was how the stand-off between the management of the two companies would be resolved.

I based my description of this conflict on two articles in *Business Week Online* (Haddad, 2001; Kharif, 2001) that were written before the conflict was resolved, and on an article in *Wireless NewsFactor* (Wrolstad, 2002) that was written after a deal had been concluded. I obtained supporting information from the websites of Alltel (http://alltel.com/) and of CenturyTel (http://centurytel.com/), including copies of their 2001 annual reports. I found the conflict using the *Business Week Online* search engine and the phrase "hostile takeover". In order to help ensure that the description would be readily understood by participants and interpreted as I had intended, I asked four colleagues to read the description, use their judgement to predict the decision that was made, and to provide suggestions on the writing. I then questioned them on their understanding. The four were experts in judgement and decision making. None recognised the conflict. I adopted most of their suggestions and changed the description where their understanding was at odds with my intentions. Eric W. Orts provided clarification on the relevant US law[10]. Participants appeared to consider the four decisions options provided to be acceptable choices in that none modified them in any way.

I chose to include Telco Takeover in my research because, first, I was able conveniently to locate the information I needed on the internet. Second, it was different from previously selected conflicts. Third, I was able to compile my material referring to reports that had been written before the conflict was resolved. Despite this last point, it is possible that my knowledge of how the stand-off was resolved influenced my description and my choice of decision options.

---

10 Communication from Eric W. Orts, Professor of Legal Studies and Management, The Wharton School, University of Pennsylvania received 21 May 2002.

Water Dispute was a 1975 conflict between two poor Arab nations (Iraq and Syria) over access to the water of the Euphrates River. Syria had built a dam across the river and started to fill the reservoir thereby reducing the flow into Iraq. Both are arid nations and Iraq was almost completely dependent on the Euphrates for water. The two nations, both Soviet-aligned military dictatorships, were preparing for war with their troops massing on the common border. A third Arab nation, rich and powerful Saudi Arabia, in an eleventh hour attempt to mediate a peaceful outcome, called the parties together. The decision to be made was whether Iraq would declare war or go ahead with its threat to bomb the Syrian dam, or whether Syria would release more water voluntarily.

I based my Water Dispute description on an account of the conflict in *Keesing's Contemporary Archives* (1975). I obtained additional information from Kliot (1994) and from material located using internet searches. I found out about the conflict by searching the shelves of the Victoria University of Wellington library for promising titles on international conflicts. In order to help ensure that the description would be readily understood by participants and interpreted as I had intended, I asked four colleagues to read the description, use their judgement to predict the decision that was made, and to provide suggestions on the writing. I then questioned them on their understanding. The four were experts in judgement and decision making. None recognised the conflict. I adopted most of their suggestions and changed the description where their understanding was at odds with my intentions. I provided three decision options. Participants appeared to consider these to be acceptable choices in that none modified the decision options I had provided in any way.

I chose to include Water Dispute in my research because, first, once I identified the conflict, I was able conveniently to obtain a succinct description of what had transpired in *Keesing's* and to locate additional background information (such as military and economic capacity, and historical context) by searching the internet. Second, it was different from previously chosen conflicts. As I wrote my description using material that had been written after the meeting between representatives of the three nations had occurred, it is possible that knowledge of the outcome influenced my description and my choice of decision options.

Zenith Investment was a conflict between managers of British Steel over a major investment decision. At the time (1975) British Steel had recently been re-nationalised.

The company was incurring heavy losses and management was planning to close older plants and lay-off workers as part of a plan to regain competitiveness. In 1974, the company had been unable to meet demand and enquiries were made about a new German steel-making technology. Since then, the demand for steel had eased and British Steel's planners pointed out that it would be cheaper for the company to meet demand by slowing the mothballing of existing plants than to build new-technology plants. On the other hand, the Chairman was keenly aware that building new plants in economically-depressed Scotland would appeal to the company's political masters. The decision to be made was whether to invest in expensive new technology and, if so, whether to invest in one new plant, or two.

My description of Zenith Investment was based entirely on a Grenada Television documentary (Graef, 1976). I became aware of the conflict as a result of watching the documentary during one of Professor John Brocklesby's classes at Victoria University of Wellington. Professor Brocklesby had, for many years, used the documentary in his teaching and had also worked for one of the managers who played a major part in the investment decision at British Steel. He read my description and considered it to be an accurate representation of the situation. The decision options facing the parties were clear from the documentary: the choice was between zero, one, and two new plants.

I chose to include Zenith Investment in my research because, first, I was already aware of the conflict and had convenient access to a videotape copy of the documentary about it. Second, it was different from previously chosen conflicts. Third, as the documentary presented the unfolding of the conflict in chronological order and without forward-looking commentary, I was able to base my material solely on information that was available before the decision was made. It is possible, however, that my knowledge of the decision, and how it was arrived at, influenced my description.

Partial results for five of the eight conflicts described were reported in Green (2002a). The five conflicts were: Artists Protest, Distribution Channel, 55% Pay Plan, Nurses Dispute, and Zenith Investment.

3.3.2 Conflict diversity

In this subsection, I examine the diversity of the conflicts using three measures:

- nature of the parties
- arena of the conflict
- game-theorist preference

The first two measures are intended as a framework for examining the conflicts as they might be viewed by managers looking for evidence that the research findings are relevant for the types of conflict with which they must deal. The third is for assessing whether the types of conflict that are of interest to game theorists are represented among the conflicts used in the research.

Nature of the parties

This measure refers to whether the parties to a conflict were:

- individuals
- organisations
- governments
- or some combination of these

For example, a conflict in which a disaffected *individual* employee makes a formal complaint against her employer, also involves the employee's union and the employing company (both *organisations*). An international dispute over access to water from a river that flows through several countries involves *governments*. Of course, no matter what the nature of the parties involved, it is individuals that interact with other individuals: generals, ministers, government officials, corporate CEOs, professional negotiators, and so on. Nevertheless, it has been shown that an individual's role has a substantial effect on that person's behaviour (Armstrong, 2001a) and the nature of the party that the individual represents (in the typology used here: him or herself; a government; some other organisation) is an important aspect of their role.

Four of the eight conflicts were between organisations. These are: Distribution Plan, 55% Pay Plan, Nurses Dispute, and Telco Takeover. Nurses Dispute also involved a

third party: a mediator The other four conflicts were between individuals (Zenith Investment), between governments (Water Dispute), between organisations and an individual (Personal Grievance), and between an organisation and a government (Artists Protests). Conflict between a government and an individual is the only combination of parties that is not represented among the eight conflicts (Table 7).

<div align="center">

Table 7
Classification of conflicts: Nature of the parties

</div>

	Individuals	Organisations	Governments
Individuals	Zenith Investment	—	—
Organisations	Personal Grievance	Distribution Channel 55% Pay Plan Nurses Dispute Telco Takeover	—
Governments		Artists Protest	Water Dispute

The absence of a conflict between a government and an individual is a weakness of the research in the sense that managers who are concerned with such conflicts may not believe the findings apply to their problems. This weakness may not extend to the generalisability of the findings however as, managers' beliefs aside, it seems reasonable from a research point of view to regard governments simply as a kind of organisation. Reclassifying Artists Protest and Water Dispute as conflicts between organisations, as this observation implies, results in six of the eight conflicts classified as occurring between organisations. Given that there is typically much more at stake in conflicts between organisations than there is in conflicts which involve individuals, the six to two ratio between the two is not unreasonable.

Arena of the conflict

This measure of diversity refers to whether a conflict took place in a setting that was:

- governmental
- civil
- commercial
- or industrial

My definitions are as follows. A *governmental* conflict is a conflict between national governments (for example, over security, or trade policy). A *civil* conflict is a conflict between groups within a nation (for example, religious, ethnic, or regional groups). A *commercial* conflict is a conflict between businesses, or within a business (for example, a corporate predator and a take-over target, competitors in the same market, or rival factions within a firm). Finally, an *industrial* conflict is a conflict between an employer and an employee or employees (for example, a personal grievance dispute or a strike over pay).

Three of the conflicts were industrial (employment relations) disputes of some kind. These are 55% Pay Plan, Personal Grievance, and Nurses Dispute. Nurses Dispute was a familiar type of dispute, while 55% Pay Plan was unusual. Personal Grievance essentially involved a single employee in conflict with her employer whereas the other two conflicts involved many employees represented by their respective unions. Another three were commercial conflicts: Distribution Channel, Telco Takeover, Zenith Investment. These were conflicts-of-interests between or, in the case of Zenith Investment, within commercial organisations. One of the conflicts, Water Dispute, was a conflict between national governments. Finally, another of the conflicts, Artists Protest, was a civil conflict over financial resources with artists on one side and the rest of the Dutch people (represented by government) on the other (Table 8).

Table 8
**Classification of conflicts:
Arena of the conflict**

Industrial	55% Pay Plan
	Personal Grievance
	Nurses Dispute
Commercial	Distribution Channel
	Telco Takeover
	Zenith Investment
Governmental	Water Dispute
Civil	Artists Protest

While each arena is represented by at least one conflict, there is typically much at stake in conflicts in the governmental and civil arenas, and extra conflicts of these kinds would have been likely to have strengthened the appeal of the research to those who are concerned with such problems. There are many conflicts in the industrial and commercial arenas, however, in which there is sufficient at stake to justify the expenditure of considerable sums of money in order to obtain accurate forecasts. In the light of the relative frequency of conflicts, the spread of conflicts between the arenas helps to justify the generalisation of findings.

Game-theorist preference

This measure of diversity was included to ensure that types of conflicts for which game theory has been recommended are included in the research.

Erev, Roth, Slonim, and Barron (2002), and Goodwin (2002) discussed the possibility that a subset of conflicts exists that are particularly amenable to forecasting by game theorists. Other researchers have identified types of conflict that they believe are amenable to forecasting using game theory. Brams and Togman (2000) and Organski (2000) used game theory to predict the outcomes of civil conflicts in Northern Ireland and the Middle East, respectively. Gruca, Kumar, and Sudharshan (1992) used game theory to predict incumbent responses to a new competitor. Keser and Gardner (1999) discussed the use of game theory to predict the outcomes of common pool resource

conflicts. Finally, Ghemawat and McGahan (1998) used game theory to explain the behaviour of a group of competing electricity generating companies and suggest that game theory is likely to be useful for predicting behaviour in conflicts that involve concentrated competition, mutual familiarity, and repeated interaction.

I used the explicit and implicit recommendations of these researchers as a basis for four "game-theorist preference" categories. They are:

- civil

- response to a new entrant

- common-pool resource,

- concentrated competition, mutual familiarity, repeated interaction

Note that the "civil" category is also included in the "arena" measure.

In contrast to suggestions of restricted applicability of game theory by some of the authors above, Fraser and Hipel (1984) maintained that their game-theoretic method (conflict analysis) can usefully forecast any type of conflict. This suggests that the game-theorist preferred classification may be a red herring in the sense that at least some game theorists believe it is appropriate to forecast any type of conflict using game theory. Nevertheless, I examine the extent to which the types of conflicts that game theorists have specifically identified as being suitable for forecasting using game theory are represented among the conflicts I used.

Artists Protest is the only conflict that could reasonably be classified in the civil category. In two other conflicts, one of the parties was a new entrant to a market. These were Distribution Channel, in which an appliance manufacturer sought to sell its wares through supermarkets, and Telco Takeover, in which one corporation attempted to enter a new region by acquiring the major provider. One of the conflicts, Water Dispute, was concerned with access to a common pool resource: the waters of the Euphrates River.

Five of the conflicts could reasonably be described as involving concentrated competition, mutual familiarity, and repeated interaction. In essence, Artists Protest was competition for government-controlled funds between two parties: the Dutch artists' union and the Dutch government, representing other citizens. The artists' union representatives, and government officials and politicians are likely to have had ongoing

dealings with each other. Zenith Investment was at least partly concerned with competition for resources and status within the corporation. The various managers and directors who were involved were all members of the Policy Committee or were senior advisors to members and thus had repeated dealings and were mutually familiar. The other three conflicts in this category (55% Pay Plan, Personal Grievance, and Nurses Dispute) were all employment relationship disputes and, as such, meet the three criteria for inclusion in this category (Table 9).

Table 9
**Classification of conflicts:
Game theorist preference**

Civil	Artists Protest
Response to new entrant	Distribution Channel Telco Takeover
Common pool resource	Water Dispute
Concentrated competition, mutual familiarity, repeated interaction	Artists Protest 55% Pay Plan Personal Grievance Nurses Dispute Zenith Investment

As Table 9 shows, each of the types of conflicts that game theorists have identified as suitable for forecasting using game theory *are* represented among the conflicts I used.

Conclusions

The eight conflicts are diverse when assessed against the measures I have used and this diversity helps to support claims that findings based on these conflicts are generalisable to all conflicts that involve interaction between few parties. The conflicts are likely to appear relevant to diverse managers, but more conflicts that involved governments (governments in conflict with individuals, in particular) would probably have helped to convince managers concerned with such conflicts that this research is relevant to them.

The conflicts include the types that game theorists have specifically identified as being suitable for forecasting using game theory. That this is so, and claims by Fraser and Hipel (1984) of the general applicability of game theory for forecasting decisions in conflicts, buttresses the research against any claims that game theorists' forecasts are disadvantaged by the selection of conflicts that were used.

3.3.3   Material provided to participants

Conflict descriptions

The descriptions of all eight conflicts were based largely on accounts compiled by neutral but knowledgeable observers. Apart from the description of 55% Pay Plan, which is a little longer, the descriptions all fit on one side of a sheet of "A4" or "Standard Letter" paper. This was the format of the descriptions used in the research described in Armstrong (2001a), and I followed this precedent.

The conflicts were disguised in the descriptions in order to reduce the chance of participants knowing the actual conflict and therefore the outcome. Disguise was kept to the minimum needed in order to achieve this aim. There were two exceptions to the policy of disguising conflicts. First, the 55% Pay Plan, which was originally used for forecasting research during the course of the conflict, was not disguised. Second, I used two versions of the Nurses Dispute description. The versions were the same in all respects other than the names of the people and organisations involved. The first version used actual names. In the second version the names of the people and organisations were changed after one potential simulated-interaction participant said that he was not willing

to play the role of a "real person", and withdrew from his session. For all conflicts, participants were asked if they recognised the actual conflict and, if so, to identify it. This enabled the forecasts of participants who genuinely recognised the actual conflict to be excluded.

Other than the exceptions just described, the descriptions provided to participants in the different treatments (combinations of forecasting method and participant expertise) were identical.

Role descriptions

The role descriptions were all brief and fit on one side of a sheet of paper, with the exception of Zenith Investment, which involved 10 roles and which ran to two pages. The role descriptions contained only information that could be gleaned from the corresponding conflict description or could be gained by extrapolating using common knowledge of conflicts. Participants, except those in the simulated-interaction treatment, were given descriptions of all roles associated with a conflict. A further, minor, exception to the uniform treatment, the non-provision of role information for some novice unaided-judgement participants, is discussed in the next subsection.

Questionnaires

Other than the simulated-interaction questionnaires prepared for this research, which were two pages long, questionnaires were kept to a single page. The conflict descriptions and questionnaires that were provided to game-theorist participants are attached as Appendix 2. Questionnaires for one of the conflicts, Zenith Investment, that were provided to participants using the methods of unaided judgement and analogies, and to participants in simulated interactions are attached as Appendix 3. Table 10 provides a summary of questionnaire content for each of the treatments as well as indications of where matters relevant to each of the items are examined in this document.

Table 10
Questionnaire content by treatment [a]

(✓: question present; ✘: question absent)

	Unaided judgement		Game theory	Structured analogies	Simulated interaction	Examined in	Tables
	Novice	Expert	Expert	Expert	Novice	Text	(figures)
Identify analogies, source, similarity, rate, select	✘	✘	✘	✓	✘	3.2.3, 3.4.3,Ch4,5	13,15-17, 20-27,33-41, 48, (1)
Forecast: select decision	✓	✓	✓	✓	✓	3.2.3, 3.4.2-6, Ch4,5	15,17, 21-26, 28-41
Describe how forecast was derived (decision made)	✓[b]	✓	✓	✘	✓ GNTWZ	3.2.1-4, 3.4.2,3.4.5	
Reason for non-prediction	✓[b] ADZ	✓	✓	✓	✘	not examined	
Time taken for task [c]	✓[b d] ADZ	✓	✓[d]	✓	✘	3.4.2-5, 4.2, 5.1.2	16-20, 27,41
Chance of change given more time	✘	✓	✓	✓	✘	3.4.4, 4.2, 5.1.2	27,35-37
Recognise conflict?	✓[b]	✓	✓	✓	✓ GTWZ	3.3.1,3.3.3, 3.4.2,3.4.4	
Identify conflict	✓[b]	✓	✓	✓	✓ GTWZ	3.3.3,3.4.2	
Number of collaborators	✘	✓	✓ GTW	✓	✘	1.3.3, 3.4.2-4, 4.2.1,5.1.2	15,26,27
Experience in field	✘	✓	✓	✓	✘	1.3.3,2.1, 2.3,4.2.2, Ch5	29-34
Experience with similar conflicts	✓ GNTWZ	✓	✓ GTW	✓	✓ GNTWZ	1.3.3,2.1, 2.3,4.2.2, Ch5	29-30, 33-34
Opinions on realism of simulation	✘	✘	✘	✘	✓	not examined	
Other characteristics of participants	✓[b]	✘	✘	✘	✓	not examined	

Abbreviations: **A** Artists Protest; **D** Distribution Channel;
G Personal Grievance; **N** Nurses Dispute; **T** Telco Takeover;
W Water Dispute; **Z** Zenith Investment

a Forecasting method plus expertise of participant
b In research by this author, only
c The standard question for each conflict was "Roughly, how long did you spend on this task? *{include the time you spent reading the description and instructions}*"
d Questionnaires for Artists Protest, Distribution Channel, 55% Pay Plan, and Zenith Investment provided to novice judges and game theorists asked "Roughly, how long did you spend on the task of deriving a prediction for this situation?".

Participants allocated to the structured-analogies treatment were told how to derive their forecasts (subsection 3.2.3) and hence were not asked how they did so. In hindsight, it would have been sensible to ask, as some who were instructed to use the structured-analogies method used their unaided judgement instead.

Participants in the simulated-interaction treatment were not asked how long they spent on the forecasting tasks as the simulations were conducted in supervised sessions in which the researchers recorded the times that were taken. As, in the other treatments, participants were mostly free to spend as much or as little time on the tasks as they wished, I asked them to record the time they spent in order to determine whether this influenced forecast accuracy. In addition, I asked participants how likely it was that they would change their forecasts if they had spent more time on the task. This information is used as a measure of forecaster confidence in the analysis and permitted an assessment of whether confidence is a good predictor of accuracy.

Participants' reasons for not forecasting individual conflicts is not examined in this document. Neither are simulated-interaction participants' opinions on the realism of their simulations nor the characteristics of the participants, other than their expertise as covered by the "experience in the field" and "experience with similar conflicts" items. These matters, although potentially interesting, are peripheral to the research objectives.

3.4 Data collection – forecasts

3.4.1 Data sources

I obtained data on forecasting method accuracy from previous research by others and from research I conducted for this thesis. As discussed, data from previous research (unaided-judgement and simulated-interaction forecasts for three conflicts) were taken from a summary of work by Armstrong and colleagues in Armstrong (2001a). In my research, I obtained (Table 11):

- unaided-judgement forecasts for two conflicts, thereby partly replicating the previous research
- unaided-judgement and simulated-interaction forecasts for five new conflicts
- game-theorist and structured-analogies forecasts for all eight conflicts
- unaided-judgement forecasts by *experts* for all eight conflicts
- unaided-judgement and structured-analogies forecasts by *collaborating* experts for all eight conflicts [11].

Table 11
Sources of forecast accuracy data
(●: Armstrong (2001a); ★: new; number of forecasts)

	Unaided judgement			Game theory	Structured analogies		Simulated interaction
	Novice	Expert		Expert	Expert		Novice
	Solo	Solo	Joint	Solo	Solo	Joint	Joint
Artists	●31 ★8	★20	★ 4	★18	★ 5	★ 4	●14
Distribution	●37 ★5	★19	★ 3	★13	★ 9	★ 3	●12
55% Pay	●15	★12	★ 4	★17	★ 8	★ 5	●10
Grievance	★ 9	★ 4		★ 5	★12	★ 2	★10
Nurses	★22	★15	★ 1	★14	★ 8	★ 5	★22
Telco	★10	★ 9		★ 7	★ 8	★ 2	★10
Water	★10	★ 6	★ 1	★ 6	★ 4	★ 1	★10
Zenith	★21	★14	★ 2	★18	★ 7	★ 1	★17
Total	●83 ★100	★99	★15	★98	★61	★23	●36 ★69

---

11 For the sake of brevity, in the balance of this document I refer to the forecasts of participants who collaborated with other people as "joint" and to those of participants who did not as "solo".

The data collection methods used to obtain the new data are described in the rest of this section. These methods were designed to replicate, to the extent that this was practical, the methods outlined in section 3.2 (Conflict forecasting methods described).

3.4.2   Unaided judgement – novices

Method

Participants read the description of a conflict and then selected the decision they thought most likely from a list of possible decisions. Novice unaided-judgement participants were selected on the basis of convenience – typically, they were undergraduate students – and not because they possessed any special knowledge of the situations or the class of problem they were asked to consider.

Participants were told the purpose of the research (Appendix 4, Information Sheet and Informed Consent form). They were provided with a full set of information for a single conflict: descriptions of the situation and the roles of all parties, and a questionnaire. They were told that the conflict was one that had actually occurred or was occurring. They were asked to read the material and, without referring to other sources of information, to use their judgement to predict the decision that had actually been made (would be made).

Participants were asked to describe the approach they used, how long they took, and for information about their age, education, and experience. Participants' descriptions of their approaches to the forecasting problem allowed, *inter alia*, deviations from unaided judgement to be identified and associated forecasts to be eliminated from analysis. In cases where predictions were not provided, participants were asked to provide a reason for not doing so. Participants who believed they recognised the actual conflict were asked to identify it and the responses of those who were able to do so correctly were excluded from the analysis. Participants were typically given approximately 50 minutes to complete the task.

As I have noted, 83 unaided-judgement forecasts by novices were summarised in Armstrong (2001a). Some of the findings reported in Armstrong (2001a) included collaboration in the research design, while others did not. I made no attempt, with that set of findings, to separate those that involved collaboration from those that did not. Role descriptions were not provided to all the participants in the research reported by Armstrong, but Armstrong and Hutcherson (1989) had found that providing role descriptions to judges had no effect on their forecasting accuracy. Forecasts from Armstrong (2001a) were Artists Protest (31 forecasts), Distribution Channel (37), and 55% Pay Plan (15).

Response

Forecasts from my research were: Artists Protest (8), Distribution Channel (5), Personal Grievance (9), Nurses Dispute (22), Telco Takeover (10), Water Dispute (10), and Zenith Investment (21).

Volunteers from two third-year classes provided eight Artists Protest predictions, five Distribution Channel predictions, and three Zenith Investment predictions. Ten minutes of class time was used to distribute material on the situations and to brief the students. They were instructed to return completed questionnaires to their lecturer, or to fold them so that the "Freepost" and address details, on the reverse of the questionnaire, were showing and post them. Copies of 55% Pay Plan were distributed, but no responses were received for that conflict. Fewer than 10 percent of the students in the two classes returned completed questionnaires. The respondents reported taking between two minutes and two hours to derive their predictions. The median time taken was 22.5 minutes.

Volunteers from one third-year university class provided 10 Zenith predictions. The 10 volunteers from this class were asked to leave the lecture theatre with the material they had been given and to return the completed questionnaires to the researcher no more than one hour after being given their instructions.

Five marketing students, one information technology student, one environmental planning consultant, and one medical doctor had been offered $NZ25 (about $US10) to

participate in a simulated-interaction session but were not needed for that purpose. Instead, they provided eight predictions for the Zenith Investment conflict. They were asked to adjourn to a room away from the role players with the material they had been given and to return completed questionnaires to me no more than one hour after being given their instructions. In a similar situation to the one just described, 22 students who were not needed for simulated interaction provided predictions for the Nurses situation instead.

In the cases of Personal Grievance, Telco Takeover, and Water Dispute, participants were allocated to use unaided judgement or to take roles in simulated interactions at the sessions. Most were recruited on my behalf by the Student Job Search agency. I recruited others by sending email appeals to students who had participated in earlier sessions. Both groups were offered $NZ25 to participate. Any who were familiar with the conflict from an earlier session were excluded. Unaided-judgement participants were typically those who remained after roles were allocated to simulated-interaction participants. I gave unaided-judgement participants short briefings (less than 5 minutes) after which they left the lecture theatre where the simulations were to take place in order to complete their tasks. With one exception, the participants were told to work independently. Participants returned completed questionnaires to me in the lecture theatre.

I collected Personal Grievance responses from students using unaided judgement during two independent sessions. Participants at one session returned questionnaires after approximately 40 minutes and those at the other after approximately 45 minutes.

Telco Takeover responses were collected from students using unaided judgement during a single session. Six of the participants collaborated in pairs and returned three questionnaires. All participants returned questionnaires after approximately 50 minutes.

Water Dispute responses were collected from students using unaided judgement during a single session. All participants returned questionnaires after approximately 50 minutes.

3.4.3   Unaided judgement and structured analogies – experts

Method

The implementation of unaided judgement by experts was the same as for unaided judgement by novices, with three principal departures. First, the participants were selected because they were experts in forecasting, conflicts, judgement, decision making, human relations, or marketing. I appealed for help to individuals on organisations' contact lists, to email listservers (Table 12), and to a convenience sample of experts.

Table 12
Organisation contact lists and email lists that were sent appeals

Name	Abbreviation	Owner
International Association of Conflict Management contact list	IACM	International Association of Conflict Management
Judgment and Decision Making Listserver	JDM	Society for Judgment and Decision-Making
Behavioural decision making	DECISION	Risk Decision & Policy, and London Judgement and Decision Making Group
Conflict Management Division Listserver	CMDNET-L	Academy of Management Conflict Management Division
Human Resource Management Listserver	HRNET	Subscribers to former Cornell HRNET ListServ organised this
Pacific Region Industrial Relations Listserver	PRIR-L	Assoc. of Industrial Relations Academics of Australia & NZ
International Employment Relations Network Listserver	IERN-L	IIRA International Industrial Relations Study Group
An electronic mail network for marketing academics	ELMAR	Sponsored by American Marketing Association
International Institute of Forecasters' Research Associates contact list [a]	IIF R.A	International Institute of Forecasters

a From 1283 mostly current and former members of the IIF who responded to an appeal to become Research Associates. The 147 Research Associates I approached (8, who were familiar with my research or had not provided email addresses, were not approached) were 11 percent of the total IIF list.

Second, I communicated with participants via email, rather than verbally. As a consequence, it was not possible to enforce a time restriction for completing the forecasting task, nor was it possible to enforce collaboration (or its absence) nor abstaining from recourse to material that was not supplied. In the appeals I sent, I asked the participants to collaborate (or not), and *not* to seek more information about the conflict than was included in the material they were given (Appendix 5). I sent up to three follow-up appeals and included the conflict material with each.

Finally, I sent prospective participants all of the conflicts for which their expertise was likely to be relevant. This was so that individual experts who were willing to complete more than one conflict forecasting task could conveniently do so. The material was sent in the form of MS-Word™ documents – one for each conflict. The order of the conflicts in the email messages was varied so as to reduce the risk of some conflicts being under-represented in the responses and so that there was no bias across conflicts in the quality of forecasts as a result of tiredness or learning or some other change in the experts. Despite these measures, participants still had the opportunity to select a subset of conflicts according to their own criteria.

The implementation of the structured-analogies method was the same as for unaided judgement by experts, except that the task itself was as described in 3.2.3 (Structured analogies). The appeal I sent to potential structured-analogies participants was different from the one I sent to potential unaided-judgement participants as it included a description of the structured-analogies procedure (Appendix 6).

Response

I anticipated that I would need a large initial sample, due to the demanding nature of the task and the lack of extrinsic rewards for participants, in order to recruit sufficient participants to allow useful analysis. My first appeal to experts to participate using either unaided judgement or structured analogies was to members of the IACM list (Table 13). I allocated equal numbers of people on the list to each of four combinations of method and collaboration (unaided judgement or structured analogies, by solo or joint). The response from among people who had been allocated to use unaided judgement was markedly higher (14 respondents) than from among those who had been allocated to use structured analogies (three). Further, few (six) of the respondents were people who had been asked to collaborate, and none of them did so.

Table 13
IACM responses by allocated treatment

	Unaided judgement		Structured analogies		Total
	Solo	Joint	Solo	Joint	
Initial sample	84	84	84	84	**336**
Invalid address	14	11	7	9	**41**
Auto-reject / on-leave	3	2	1	5	**11**
Not an expert	1	3	4	3	**11**
Too busy	13	9	20	7	**49**
Further info. request [a]	7	6	5	6	**24**
Promise of help [a]	1	9	8	6	**24**
No response	38	40	38	46	**162**
At least one completed response	10 [b]	4	1	2	**17** [b]

a People who replied to the appeal for participants but did not provide completed responses.
b One professor did not participate himself but arranged for four conflict-management graduate students to do so

As a consequence of the poor response for the structured-analogies and collaboration treatments from those on the IACM list, in appeals to members of other lists I asked all potential participants to use the structured-analogies method and to collaborate. I assumed that the proportion of those respondents who would ignore my request to collaborate and the proportion who would be either unable to think of analogies or who would ignore their own analogies in making a forecast would be sufficient for the purposes of my analysis. These assumptions were based on the IACM list responses.

Between zero and six percent of the members of the nine formal lists participated (Table 14). The highest response rate from these lists was from IIF Research Associates (5.8 percent). The people on that list had, days before, volunteered to become Research Associates in response to an appeal to participate in research projects supported by the IIF. The eight participants from this source amount to 0.6 percent of the full IIF list. When the IIF response rate is calculated on this basis, the highest response rate was from the IACM list. Personalised appeals sent to a convenience sample of 14 achieved a 50 percent response rate.

Table 14
Sources of expert (non-game theorist) participants

Source	Appeal	Conflicts	Participants	Response rate(%)[a]
IACM	Personalised	All	17[b]	4.1[c]
JDM	Impersonal	All	2	0.3
DECISION	Impersonal	All	4	2.0
CMDNET-L	Impersonal	All	1	0.2
HRNET	Impersonal	Artists/55%/Grievance/Nurses	3	0.2
PRIR-L	Impersonal	Artists/55%/Grievance/Nurses	-	0.0
IERN-L	Impersonal	Artists/55%/Grievance/Nurses	1	0.3
ELMAR	Impersonal	Distribution / Telco Takeover	5	0.2
IIF R.A	Personalised	All	8	5.8[d]
Conv. samp.[e]	Personalised	All, or Grievance/Telco/Water	7	50
Total (unweighted)			**48**	**1.5**[f]

a Number of respondents divided by number of email addresses in original sample
b Includes four graduate students who participated in the place of their professor (Table 13)
c Numerator counts the four graduate students as a single participant
d Roughly 0.6 percent of all those on the full IIF list participated
e Convenience sample
f Excludes convenience sample.

I received between one and nine forecasts from the 48 expert participants. Eighteen forecasts were clearly inconsistent with forecasters' own analogies. In these cases, participants followed the structured-analogies procedure (subsection 3.2.3) but chose decision options other than the decision options implied by their own analogies. I recorded these inconsistent forecasts as an unaided-judgement forecast and, when the forecaster provided a single analogy or a single decision option was implied by the set of analogies provided, I adopted the analogy-implied forecast as a second forecast. For example, one participant was effectively responsible for nine forecasts for eight conflicts. I examine the effect on forecast accuracy of participants ignoring the implied decision of their own analogies in subsection 4.1.1.

There were 114 unaided-judgement forecasts. Fifteen of these were the product of collaboration between two or more people. There were 84 structured-analogies forecasts – 23 were collaborative forecasts. (Table 15). Despite repeated appeals to more than six thousand potential respondents, I was unable to achieve my target of 10 forecasts for each treatment for each conflict. Had the response rates from impersonal appeals to lists approached the response rate achieved with personalised appeals to IACM members, the target would almost certainly have been met.

Table 15
Unaided-judgement and structured-analogies forecasts by experts
Number of forecasts

	Unaided judgement			Structured analogies		
	Solo	**Joint**	**Total**	**Solo**	**Joint**	**Total**
Artists	20	4	**24**	5	4	**9**
Distribution	19	3	**22**	9	3	**12**
55% Pay	12	4	**16**	8	5	**13**
Grievance	4		**4**	12	2	**14**
Nurses	15	1	**16**	8	5	**13**
Telco	9		**9**	8	2	**10**
Water	6	1	**7**	4	1	**5**
Zenith	14	2	**16**	7	1	**8**
Total	**99**	**15**	**114**	**61**	**23**	**84**

Participants reported taking between 10 minutes and 12 hours to derive forecasts for single conflicts, including time for reading instructions and conflict description. Mostly they spent about 30 minutes (Table 16). Participants who worked on their own and also used structured analogies to forecast tended to spend longer (about 10 minutes more) on the task than did those who used their unaided judgement (permutation test for paired replicates, one-tailed, $P = 0.016$; Siegel and Castellan, 1988). Participants who used structured analogies and also collaborated, tended to spend longer again (about 10 minutes more again) than those who worked on their own (permutation test for paired replicates, one-tailed, $P = 0.06$; Siegel and Castellan, 1988).

Table 16
Unaided-judgement and structured-analogies forecasts by experts
Median time taken to forecast [a]
(number of forecasts)

	Unaided judgement [b]			Structured analogies [c]		
	Solo	**Joint**	**Total**	**Solo**	**Joint**	**Total**
Artists	30 (20)	120 (1)	**30**	30 (5)	45 (4)	**30**
Distribution	20 (17)	15 (1)	**20**	30 (9)	30 (3)	**30**
55% Pay	25 (11)		**25**	30 (8)	18 (5)	**30**
Grievance	30 (3)		**30**	30 (12)	60 (2)	**30**
Nurses	15 (14)	180 (1)	**15**	30 (8)	30 (5)	**30**
Telco	30 (7)		**30**	45 (8)	45 (2)	**45**
Water	30 (5)	30 (1)	**30**	39 (4)	60 (1)	**48**
Zenith	20 (13)	120 (1)	**25**	30 (7)	60 (1)	**45**
Total (unweighted)	**25 (90)**	**93 (5)**	**26**	**33 (61)**	**44 (23)**	**36**

a Minutes
b Forecasts from forecasters who did not also use structured analogies for the conflict
c Includes forecasts derived from analogies which were ignored by forecasters

In their instructions (Appendices 5 and 6) participants were told to "either pick an outcome or assign probabilities" and, in the questionnaires, they were told to "check one ✓, or %" in order to indicate their forecast. Participants provided probabilities for 21 unaided-judgement forecasts and eight structured-analogies forecasts (Table 17). Eight of the 21 probabilistic unaided-judgement forecasts were from analogy treatment participants. The eight responses were coded as unaided judgement because the probabilities were inconsistent with the participants' own analogies. I coded all but three of the probabilistic unaided-judgement forecasts (one each for Distribution Channel, Nurses Dispute, Water Dispute) and all but one of the probabilistic structured-analogies forecasts (for Telco Takeover) as single-decision forecasts for calculations of the proportion of correct predictions. The Nurses Dispute forecast was provided by collaborating participants. My rule for coding probabilistic forecasts to single-decision forecasts was that if there was a single probability allocated to a decision option that was larger than those allocated to all other decision options, that decision option was the single-decision forecast. In the case of the four exceptions, the allocation of probabilities was too even to allow this and they were excluded from analysis of proportion correct. I examine probabilistic forecasts in section 4.1.

Table 17
Probabilistic unaided-judgement and structured-analogies forecasts by experts
Number of forecasts

	Unaided judgement			Structured analogies		
	Solo	**Joint**	**Total**	**Solo**	**Joint**	**Total**
Artists	3	1	**4**			
Distribution	3	1	**4**	1	1	**2**
55% Pay	1		**1**	1		**1**
Grievance	1		**1**	2		**2**
Nurses	2	1	**3**		1	**1**
Telco	4		**4**	1		**1**
Water	3		**3**			
Zenith	1		**1**	1		**1**
Total	**18**	**3**	**21**	**6**	**2**	**8**

3.4.4   Game theory – experts

Method

With two exceptions, I implemented the game-theory method in the same way as unaided judgement by experts. The exceptions were, first, the participants were all game-theory experts. Second, the participants received feedback on their collective performance, relative to unaided judgement by novices and simulated interaction, for six situations (Artists Protest, Distribution Channel, 55% Pay Plan, Nurses Dispute, Panalba Drug Policy, Zenith Investment) in the form of a draft of Green (2002a), before they sent their responses for the final three conflicts (Personal Grievance, Telco Takeover, Water Dispute). One of the six conflicts provided to participants before they received feedback, Panalba Drug Policy, is not included in this research as the situation as described did not involve direct interaction between parties with divergent interests (Armstrong, 2002). That is, the conflict was not of the type that is the subject of my research.

I recruited participants by sending an email appeal for help to an initial sample of 558 game theorists composed of the members of the Game Theory Society, recipients of the International Society of Dynamic Games "E-Letter", and several prominent game-theory experts who were not otherwise included. My email message drew the attention of potential participants to the purpose of the research (Appendix 7). The message had the subject line "Using Game Theory to predict the outcomes of conflicts" and the first paragraph included the text "I am engaged on a research project which investigates the accuracy of different methods for predicting the outcomes of conflicts". The emails included five conflicts in the form of attached MS-Word™ documents. They were: Artists Protest, Distribution Channel, 55% Pay Plan, Panalba Drug Policy, and Zenith Investment.

Two weeks after my original appeal, I sent individualised email reminder messages to 413 addresses from which I had received no response. I included the letter sent in the first email and, in the original order, the attached files.

Approximately a year after my original appeal, I sent another appeal to those who had participated earlier. The appeal was for forecasts for Nurses Dispute. I sent up to two reminders, a week apart, to those who had not responded.

A year after the first Nurses Dispute appeal, I sent an email appeal for forecasts for Personal Grievance, Telco Takeover, and Water Dispute to those experts who had participated in the first round. I sent up to three reminders, between one and four weeks apart, to non-respondents.

Questionnaires that I used early in my research programme (for Artists Protest, Distribution Channel, 55% Pay Plan, Zenith Investment) did not specifically instruct respondents to include the time they took to read the material in the question on how long they took to derive a prediction (see Table 10). To rectify this omission, and to obtain information on the game theorists' confidence in their forecasts, I sent an email to game-theorist respondents asking three questions:

1. Did the times you provided for the task of deriving predictions include, or exclude, the time you took to read the material on the situation and the parties?
2. Roughly, how long did you spend reading the material on each of the situations? (an average figure is OK)
3. Do you think spending more time on the problems would have been likely to have changed your predictions?

Where respondents indicated they had not included reading time in their earlier responses, I added to their times the reading times they had provided in answer to the second question. In all cases, the respondents provided a single time in answer to this question rather than a time for each of the conflicts for which they had provided forecasts. One respondent did not answer any of these questions. Five others, who had indicated that they had not included reading time, failed to provide estimates of how long they had spent reading. For these five, I added the average of all the reading time responses to their times.

Response

I received forecasts from 21 game theorists – 3.8 percent of the initial sample. Participants each provided between one and eight forecasts. In total, 98 forecasts were received for the eight conflicts examined in this research – between five and 18 forecasts per conflict. Participants reported taking between 10 and 139 minutes to derive forecasts for individual conflicts, including time for reading instructions and conflict description.

Mostly, they spent about 30 minutes (Table 18) – the same as non-collaborating experts using structured analogies.

Table 18
Forecasts by game theory experts:
Median time taken to forecast [a] [b]
(number of forecasts)

Artists Protest [c]	30	(17)
Distribution Channel [c]	30	(12)
55% Pay Plan [c]	30	(16)
Personal Grievance	30	(5)
Nurses Dispute	30	(14)
Telco Takeover	30	(7)
Water Dispute	30	(6)
Zenith Investment [c]	30	(17)
Total (unweighted average)	**30**	**(94)**

a In minutes
b Times not available for four forecasts
c For these conflicts, the times that respondents
    spent reading the material were based on
    responses to an *ex post* survey.

I had received responses from 269 email addresses in total. This was 48 percent of the initial sample. Aside from the 21 who provided forecasts, in response to my original appeal and reminders, I received 78 invalid-address messages, 18 automatic messages informing me that my message had been rejected (typically because the addressee was on leave), and six messages stating that the addressee was not an expert in game theory. The balance of 146 messages consisted of 95 messages from game theorists who did not wish to participate, and 51 messages from game theorists who either promised to respond or who wanted more information.

Ninety of the 95 game theorists who refused to participate provided reasons (Green, 2002a: Appendix B; reproduced here as Appendix 8). Most of them (72) stated that they were too busy to help with this research. Eight others stated that their game-theory speciality was not applicable to the tasks that I provided. Six stated that it was not appropriate to apply game theory to the tasks, and four stated that the information I provided was not sufficient for them to derive forecasts.

Two game theorists provided forecasts for all eight conflicts, while three provided forecasts for a single conflict. The 16 other game theorists provided forecasts for intermediate numbers of conflicts. One of the conflicts (Artists Protest) was recognised by one respondent. His forecast for that conflict is neither included in the previous discussion nor is it included in any subsequent analysis. Questionnaires for Artists Protest, Distribution Channel, 55% Pay Plan, and Zenith Investment asked respondents to "check one – ✓" and the questionnaire for Nurses Dispute asked respondents to "tick one box only". Despite these instructions, one game theorist provided probabilities for each of the six possible Artists Protest decisions. He assessed the probability of the actual outcome as zero, and hence his forecast is coded as inaccurate for calculations of the proportion of correct predictions. In 4.1 (Relative accuracy) I examine probabilistic forecasts.

3.4.5   Simulated interaction – novices

Method – overview

The procedures I adopted for this method were substantially the same as those adopted by Armstrong (1987), and Armstrong and Hutcherson (1989). These procedures were described in Armstrong (2001a). As shown in Table 6, I took simulated-interaction findings for Artists Protest (14 forecasts), Distribution Channel (12), and 55% Pay Plan (10) from that source. I conducted simulated-interaction sessions for Nurses Dispute, Personal Grievance, Telco Takeover, Water Dispute, Zenith Investment.

I gave simulated-interaction participants a single role description to read and told them to adopt that role for the duration of their simulation. I then asked the participants (role players) to read a description of one of the five conflicts for which I had prepared material. I then divided role players into groups; each group comprising one role player for each of the roles. For example, in Nurses Dispute there were five roles: two management roles, two union roles, and one mediator role. Once the role players were in their groups, they were told to simulate the conflict from the point in time specified in the description until a decision was arrived at, or they ran out of time. On completion, each role player recorded the decision made by the group or, if the group had run out of

time, the role players each recorded the decision they thought would have been made had they been free to continue.

With the exception of four groups of Nurses Dispute role players, the role players were recruited on the basis of convenience rather than because they resembled the characters they were to play, or were familiar with the role they were to play, or were knowledgeable about the situation. In practice, the role players were mostly undergraduate university students. For most of the simulations, I asked participants who identified with general role types (for example, employer or employee) to step forward. I then allocated the participants to the roles that coincided with their preferences. I describe this process and the exceptions to it, for each of the five conflicts, later in this subsection.

Participants assembled in lecture theatres or similar locations for their briefings. I gave each an information sheet, which provided basic information on the research project and on simulated interaction, as well as a "consent" form (Appendix 4). I emphasised to participants that it was very important they take the simulations seriously. Additional space such as a second lecture theatre, meeting rooms, or lobby areas was available for the simulations. The role players were encouraged to make good use of the available space for holding both formal meetings and private discussions. At the end of their briefing I told role players that they were free to improvise, provided that they remained in-character and true to the situation description. They were allowed to retain their printed role and situation descriptions for the duration of their session. This material included the questionnaire that the role players were to fill in at the end of their simulated interactions. I drew the attention of role players to the decision options presented in the questionnaires, and told them that they would be expected to match one of these with their own group's simulated-interaction decision. The role players were also given printed self-adhesive name badges showing their character's name and position. They were told to meet with the rest of their group, wearing their name badges, and to introduce themselves to each other while in-character. Role players were instructed to arrange a time and a place for the first formal meeting between the parties and then to take time (typically 10 minutes) preparing with confederates (members of the same party) or potential allies. I told role players that they were free to hold several meetings or to interrupt a meeting for private discussions with confederates as they saw fit.

Method – Nurses Dispute

There were three independent Nurses Dispute sessions. I asked participants in the first and second sessions whether they would tend to identify more with union or management in a pay dispute. I then allocated participants according to their preferences, with the more equivocal participants given the mediator role. Participants in the third session were allocated to roles on the basis of my assessment of the compatibility of their experience with the demands of the roles. During the briefing, I told role players playing the mediator role to join the other mediators, one from each group at the session, at a designated place after they had greeted the other members of their group. They were told to discuss mediation and its application to the dispute among themselves until called upon by the parties in their own group, or 30 minutes had elapsed. If 30 minutes elapsed without agreement by the parties, the parties were obliged to accept the services of their government appointed mediator. This measure was intended to simulate the effect of employment relations legislation that had just taken effect at the time of the dispute.

Method – Personal Grievance

There were two independent Personal Grievance sessions. Briefing, allocation of roles, and reading of the material took 15 minutes in the first session and 20 minutes in the second. I asked participants whether they would tend to identify more with the employee or with management in a personal grievance. I then allocated participants according to their preferences, with the more equivocal participants given the mediator role. I told role players that, after introductions, the mediator should set a time and a place for the three parties to meet. During the simulations at the first session, I was asked by one role player to explain "salary bands". I interrupted the simulations to give all participants a brief description of the distinction between an employer's assessment of the value of a job (the salary band for the position) and the value of an individual employee (the actual salary of the employee). I drew a diagram on the lecture theatre white-board to show the relationship between the salary of the employee in this conflict and the salary band for her position that had been determined by her employer. At the end of their simulation, one of the groups had invented another decision option that allowed for special treatment

of the employee. I pointed out that this was inconsistent with the information they had been given and sent them away to choose from among the options that had been provided to them.

As a consequence of these experiences, in the second session briefing I emphasised to participants more strongly than before that they should read their role description and the decision options carefully and that their simulations should be consistent with the information they had been given. I told participants that they should think about what implications each of the decision options might have for them in their roles. I told them that the people who were involved in the actual grievance considered the list of decision options to be a realistic summary of the alternatives. Finally, I gave participants a talk on salary bands similar to the one I had given during the first session.

Method – Telco Takeover

There was one Telco Takeover session. Briefing, allocation of roles, and reading of the material took 17 minutes. Both parties to the dispute were represented by a chairman and a CEO in the simulations. I explained this, briefly described the typical distinctions between these roles, and asked for volunteers for the chairman roles. I continued to solicit volunteers until I had sufficient to fill these roles. The CEO roles were filled from among the remaining participants. I told role-players to report back to me with their completed questionnaires up to 50 minutes after the start of their simulations.

Method – Water Dispute

There was one Water Dispute session. Briefing, allocation of roles, and reading of the material took 17 minutes. Two of the three parties were represented by a foreign minister and a military advisor. I explained this, briefly described what these roles might typically entail, and asked participants to choose between the roles. When, after some encouragement, I had sufficient volunteers, I allocated people to the third party (mediator) role from among the remaining participants. I told role-players to report back to me with their completed questionnaires up to 38 minutes after the start of their simulations, but said that I was happy to wait if they needed longer.

Method – Zenith Investment

There were three independent Zenith Investment sessions. I allocated roles randomly to participants in the first two sessions. In the third session I first asked "natural leaders" and then "quantitative analysis experts" to come to the front of the theatre. Those who responded to the call for natural leaders I allocated the Chairman role, while the quantitative experts I allocated to the roles of either Finance Director or Chief Planner. The remaining participants were randomly assigned to the seven other roles in this conflict. I told the role players that the Chairman of Zenith would set a time and a place for a meeting and that, after this had been done, they should then prepare for the meeting in a manner consistent with their printed briefing material. During the verbal briefings, I emphasised that it was appropriate to hold informal discussions with other members of their group (the Zenith Policy Committee) prior to the meeting.

I made no efforts to increase the realism of any of the simulations beyond the measures I have described. There were no theatrical or technological devices, nor did I use any role players who were secretly in league with me.

Response – all conflicts

Nurses Dispute was simulated by 110 participants in 22 groups. In the first of three sessions, ten students with work experience enrolled in courses on dispute resolution simulated in two groups. In the second session, 90 students recruited with an offer of $NZ25 cash simulated in 18 groups. In the third session, 10 participants selected for experience relevant to the situation (union negotiators, managers, management negotiators, and professional mediators) simulated in two groups.

Personal Grievance was simulated by 50 participants in 10 groups. In the first session there were four groups of five role players and in the second there were six groups. The participants were all university students. They were each paid $NZ25 cash.

Telco Takeover was simulated by 40 participants in 10 groups. The participants were all university students. They were each paid $NZ25 cash.

Water Dispute was simulated by 50 participants in 10 groups. The participants were all university students. They were each paid $NZ25 cash.

Zenith Investment was simulated by 170 participants in 17 groups. The participants were all university students. Five groups simulated the situation in an organisational behaviour class and four in a conflict-of-laws class. The eight groups of role players in the third of the three sessions were recruited from among students attending mathematics and computer science lectures. Students who attended the third session had been promised $NZ25 to take part. All participants were told that taking part would help with research on decision-making in conflict situations and that taking part was likely to be both enjoyable and (in the case of the first two sessions) relevant to their studies. Participants were told that the conflict they were to simulate had occurred in the past and that it involved a group of senior managers making an important investment decision.

Ten minutes of each session, more-or-less, was allocated to reading role and conflict descriptions. The simulated interactions I conducted for this research took between 40 and 100 minutes, including reading time. Most simulations took between 45 and 60 minutes and 50 minutes was typical (Table 19).

Table 19
Forecasts from simulated interaction
Time taken to forecast, in minutes
(number of forecasts)

Artists[a]	48 (14)
Distribution [a]	48 (12)
55% Pay[a]	48 (10)
Grievance [b]	49 (10)
Nurses[b]	63 (22)
Telco[b]	46 (10)
Water [b]	45 (10)
Zenith [b]	60 (17)
Total (unweighted average)	**51 (105)**

a Average time across conflicts (Armstrong, 1987)
b Ten minutes reading plus median of simulation time
    including time to complete questionnaires.

3.4.6 Summary and implications

Data collection for this research is summarised in Table 20.

Table 20
Summary of data collection

	Unaided judgement		Game theory	Structured analogies	Simulated interaction
	Novice	**Expert**	**Expert**	**Expert**	**Novice**
Participants	undergraduate university students	miscellaneous experts (see Table 14) ✚	game-theory experts ✚	miscellaneous experts (see Table 14) ✚	undergraduate university students
Recruitment	promotion at lectures; Student Job Search; own email list	via email appeals (see Table 12) ✚	Game Theory Society & other email list ✚	via email appeals (see Table 12) ✚	promotion at lectures; Student Job Search; own email list
Motivation	$NZ25	collegiality ✚	collegiality ✚	collegiality ✚	$NZ25
Briefing	in person	via email	via email	via email	in person
Material	complete ✚	complete ✚	complete ✚	complete ✚	single role
Setting	lecture theatres; discretion	discretion ✚	discretion ✚	discretion ✚	lecture theatres & environs
Supervision	researcher; none	none ✚	none ✚	none ✚	researcher
Timing	<= 1 hour; discretion	discretion ✚	discretion ✚	discretion ✚	typically <= 1 hour
Forecast	judgement	judgement	game-theoretic reasoning; judgement	formal analysis of analogies; judgement	simulation outcome; judgement
Responding	researcher collected; post	email, fax, or post ✚	email, fax, or post ✚	email, fax, or post ✚	researcher collected

✚ A variation in the treatment of the method that, for the variable considered, seems likely to increase forecast accuracy for the method relative to the other methods.

Of the ten data collection variables listed in Table 20, there are eight for which it seems plausible to suppose that variations in the treatment of the methods might lead to biases that favour the forecast accuracy of some methods over others. One pattern is obvious: for these eight variables, the treatment of the methods seem likely to favour forecasts by experts using unaided judgement, game theory, or structured analogies.

The reasons why forecasts from these treatments seem likely to be favoured are, *first*, the participants were experts rather than novices. *Second*, because the participants were recruited via email appeals that included all the material required for the tasks, they were able to make well-informed decisions as to whether to participate or not – even to the extent that they could complete the tasks and then decide not to participate.

Third, the participants were likely to be more motivated than others because they knew that, once they had sent me their responses, I knew who they were and would publish acknowledgement of their help. Further, the participants knew that their expertise and their fields (conflict management, game theory, and so on) were being evaluated. By contrast, for the other participants, who were mostly undergraduate students, participating was an easy way to earn a useful sum of money. They had no attachment to the method they were using and their participation was completely anonymous. Many, if not most, seemed keen to finish as early as they could.

Fourth, the participants received explicit information on all the roles. Novices using unaided judgement also received all the information, but simulated-interaction participants were given only information on their own role.

Fifth, the participants were free to complete their tasks in whatever environment they wished, whereas the other participants were obliged to complete their tasks in environments that seemed inimical to considered thought – sometimes crowded lecture theatres and nearby public spaces often with people moving about and conversing loudly.

Sixth, the participants were not supervised and therefore had access to resources (for example libraries, the internet, other experts) that the supervised participants did not. The participants were asked not to refer to any material other than what they had been given and I am not aware that any did so, but rather pointing out that they had the opportunity whereas the other participants did not. On the other hand, might the supervised participants have been led to act in ways that were consistent with the researcher's expectations? Armstrong (2002) discounted such "demand effects", citing evidence that participants, rather than seeking to co-operate with researchers, tend to be concerned to present themselves in a favourable light (Sigall, Aronson, and van Hoose,

1970). As Armstrong (2002) pointed out, it is likely that experts who were all identified would have more concerned about looking good than anonymous novices.

Seventh, the participants had discretion over how long they spent on their tasks whereas the other participants had restricted time available to them.

Finally, because I was in communication with the participants by email, I was able to seek clarification on their responses and obtain responses for items that had been overlooked. This was not the case with the other participants: although I tried to check all the questionnaires as they were returned, the process was far from perfect due to the pressure of large numbers of students in a hurry to get paid and leave.

In the case of the variable, "briefing", it is possible that having the researcher brief participants in person might have tended to increase the accuracy of unaided-judgement and simulated-interaction forecasts from novices, but it is also arguable that written instructions and individual attention via email messages resulted in superior briefings.

The "forecast" variable represents the methods being evaluated.

3.5 Data collection – opinions

Method

I obtained assessments of the appeal of the forecasting methods to managers using the Delphi technique. Delphi has been shown to be superior to aggregating individual experts' opinions, and to traditional group approaches (Rowe and Wright, 2001). The Delphi panel's informed assessments of the forecasting methods were intended to provide a check on the extent to which recommendations based on the research are likely to be accepted by managers, and not rejected for reason or reasons unanticipated. The information collected from the panel thus supports the research purpose of providing *useful* recommendations to managers.

I implemented the Delphi process in accordance with Rowe and Wright's (2001) recommendations, as described in this paragraph and the next one. The Delphi procedure

was conducted in two parts. First, I sent panellists a description of the type of forecasting problem to be considered – forecasting decisions in conflicts – and examples of such problems. I asked the panellists to rate the importance of the 16 criteria for assessing the appeal of forecasting methods (listed in subsection 1.3.4) for such problems. I used the same seven-point scale as was used in the study described by Yokum and Armstrong (1995) (personal communication with the authors on 25 August, 2002). When they had done this I sent feedback to the panellists and requested that they review their ratings in response to the feedback. Feedback for each criterion was composed of the median, minimum, and maximum of the panellists' ratings together with reasons for a low rating from two panellists and reasons for a high rating from two other panellists. This material is included as Appendix 9. When the second set of responses were in, I informed panellists of the panels' average ratings of the importance of each of the criteria.

Second, I sent panellists descriptions of the four conflict forecasting methods I had investigated, together with findings on the relative accuracy of forecasts from the methods. I asked them to rate each of the methods against the 16 criteria using an 11-point (zero-to-ten) scale. Preston and Colman (2000) found that an 11-point scale, labelled at both ends, rated well relatively and absolutely for validity, reliability, and discriminating power. They reported that their respondents considered the 11-point scale to be not so quick and easy to use as shorter scales, but the spread of ratings was not great and all scales were rated highly (in absolute terms) against these criteria. The respondents gave the 11-point scale much higher ratings than shorter scales for allowing them to express themselves adequately. Unless there was a compelling reason to do otherwise, I used 11-point scales throughout this research for the purpose of obtaining ratings from respondents.

When panellists had rated the methods, I sent them feedback and requested that they review their ratings in response to this. I provided feedback on each criterion for each method. As with the first part, feedback included the median, minimum, and maximum of the panellists' ratings. In this part, I also provided reasons for ratings from *all* of the panellists. It was convenient to do this, as they typically provided a single comment covering the reasons for their ratings of the four methods against each of the criteria (Appendix 10). The Personal Grievance and Distribution Channel conflicts are treated differently in the table of findings included in the Appendix 10 material than they are later in this document. In the case of Personal Grievance this is because I did not collect

forecast usefulness ratings (discussed in 4.1.2) until after the Delphi process was complete. (Prior to the Delphi process, I had counted both options A and B as accurate). In the case of Distribution Channel, either/or responses from game theorists were coded as one correct forecast in the message to panellists whereas in the rest of this document, they are coded as half of a correct forecast (see subsection 3.3.1).

After receiving all the responses that were forthcoming, I sent panellists a summary of the results from the Delphi process. In particular, I provided the aggregate rating for each of the four methods (R_m). These figures were, for each method, the weighted sum of the average of the panellists' ratings for each criterion, where the weights were the averages of panellists' importance ratings for the criteria as a proportion of the total of the average importance ratings, or

<div align="center">

Formula 1

Aggregate rating for method m

</div>

$$R_m = \sum_{c=1}^{C} \left\{ \frac{1}{P} \sum_{p=1}^{P} R_{mcp} \cdot \frac{\frac{1}{P} \sum_{p=1}^{P} I_{cp}}{\sum_{c=1}^{C} \frac{1}{P} \sum_{p=1}^{P} I_{cp}} \right\}$$

Where:
R_m = aggregate rating for method m
C = number of criteria
P = number of panellists who provided a rating
R_{mcp} = rating of method m against criterion c by panellist p
I_{cp} = importance of criterion c for panellist p.

In practice, P varied between calculations, as not all panellists provided a full set of ratings.

In the same email message, I asked panellists to, for each of the methods, assess the likelihood that they would use or recommend the method the next time they were faced with an important conflict forecasting problem (Appendix 11). Panellists were given a simplified Juster Scale (Morwitz, 2001) for this purpose.

I communicated with panellists via email. The panellists did not know the identity of the other participants during the Delphi process, but were told that I would publish acknowledgement of their help after the event.

Response

I had hoped to recruit diverse senior managers with an interest in forecasting decisions in conflicts for the Delphi panel. After some initial enquiries, I decided that this was impractical. Instead, I recruited one senior manager using a personal appeal by email message and six members of the International Association of Conflict Management using a general appeal to the Association's list. Two of the latter had earlier provided unaided-judgement or structured-analogies forecasts.

Of the seven panellists, three competed the whole process described at the beginning of this section. Two panellists completed only the first part and two withdrew after the first round of the second part. One of the latter two also filled-in the post-Delphi questionnaire. There were, therefore, four panellists who made predictions of the likelihood that they would use each of the methods.

4.   Findings

In this chapter, I examine the findings from my research in three sections. In section 4.1, I examine the relative accuracy of forecasts from the four forecasting methods. In section 4.2, I assess the generalisability of the findings. In particular, I examine the effects of collaboration and expertise on forecast accuracy. Finally, in section 4.3, I examine the appeal of the methods to managers using the findings from a Delphi panel process.

4.1   Relative performance of methods

I use several methods to compare the accuracy of forecasts. The proportion of correct predictions is the principal measure I use, and the principal unit of analysis is that measure for one method applied to one conflict. Recognising that forecasts that are not entirely accurate may nevertheless still be useful, I also compare the performance of methods using independent ratings of how useful forecasts of each of the options provided to participants are likely to have been.

4.1.1   Effect of method on accuracy

Percent correct

Table 21 presents findings on the accuracy of forecasts made using four methods. The forecasts from the first three methods shown in the table are those of experts who reported that they did not collaborate with others: solo experts. The methods are unaided-judgement, game-theoretic, and structured-analogies forecasts. The fourth set of forecasts are from student role-players' simulated-interactions. On average, forecasts from simulated interaction were more accurate than structured-analogies forecasts, which were in turn more accurate than game theorists' forecasts and experts' unaided-judgement forecasts, and these in turn were somewhat more accurate than chance.

Table 21
**Accuracy of solo-experts' forecasts,
and forecasts from simulated-interaction by novices** [a]
Percent correct forecasts (number of forecasts)

	Chance	Unaided judgement		Game theorist		Structured analogies		Simulated interaction	
Telco Takeover	25	0	(9)	0	(7)	14	(7)	**40**	(10)
Artists Protest	17	10	(20)	6	(18)	20	(5)	**29**	(14)[b]
55% Pay Plan	25	17	(12)	29	(17)	38	(8)	**60**	(10)[b]
Zenith Investment	33	29	(14)	22	(18)	43	(7)	**59**	(17)
Distribution Channel [c]	33	33	(18)	23	(13)	50	(9)	**75**	(12)[b]
Personal Grievance	25	50	(4)	**60**	(5)	42	(12)	**60**	(10)
Water Dispute	33	60	(5)	67	(6)	75	(4)	**90**	(10)
Nurses Dispute	33	71	(14)	50	(14)	75	(8)	**82**	(22)
Totals (unweighted) [d]	28	34	(96)	32	(98)	45	(60)	**62**	(105**)**

a All forecasts are by individual experts except those from simulated interaction which, apart from four Nurses forecasts from groups of experts, are from groups of novices. Four probabilistic forecasts that could not be coded as single-decision forecasts are not included

b Forecast accuracy data reported in Armstrong (2001a)

c The game-theorist percent correct figure for this conflict differs from that reported in Green (2002a) as an either/or forecast was there coded as one correct forecast whereas here it is coded as half of a correct forecast (see subsection 3.3.1)

d Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

The average accuracy figures reported in Table 21 are *un*weighted averages of the accuracy figures for individual conflicts. The number of forecasts varies widely across methods and conflicts. Consequently, weighted averages would inappropriately place more weight on the accuracy figures from method / conflict combinations for which there happen to be more forecasts.

A breakdown of all forecast responses across all decision options is attached as Appendix 12. The breakdown does not include forecasts from others' research.

Using the Page test for ordered alternatives (Siegel and Castellan, 1988), I tested the hypothesis that the more realistically a forecasting method can represent a conflict, the more accurate will be the forecasts from that method (subsection 1.3.2). Table 21 displays the findings for the methods in the hypothesised order. The test result (N = 8, k = 4, L = 231.5, $P < .001$) supports the hypothesis.

The Page test result implies that the difference between at least one pair of treatments (forecasting methods) in the ordered sequence of treatments is a strict inequality, while the rest may be equalities. With the four forecasting methods in the proposed order, there

are three comparisons that can be made between adjacent pairs of methods. These are: simulated interaction > structured analogies; structured analogies > game-theorists forecasts; and game-theorist forecasts > unaided judgement. It is clear from Table 21 that the largest difference in accuracy between the forecasts of any pair of methods is between simulated-interaction and structured-analogies. As is implied by the Page test result, the difference is statistically significant (permutation test for paired replicates, one-tailed, $P = 0.004$; Siegel and Castellan, 1988). The difference between the accuracy of structured-analogies and game-theorist forecasts is smaller, but is also statistically significant (permutation test for paired replicates, one-tailed, $P = 0.027$). Finally, game-theorist forecasts were no more accurate (strictly, they were *less* accurate) than experts' unaided-judgement forecasts.

The finding that simulated-interaction forecasts were more accurate than structured-analogies forecasts is robust. Simulated-interaction forecasts were more accurate for each of the conflicts: the difference in accuracy between forecasts from the methods is statistically significant ($P < 0.05$) for any combination of five or more of the conflicts. Further, as the discussion in subsection 3.4.6 suggests, any bias in the treatment of the different methods is likely to favour experts' unaided-judgement, game-theorist, and structured-analogies forecasts over simulated-interaction forecasts.

The difference in accuracy between structured-analogies forecasts and game-theorist forecasts was not statistically significant for the set of five conflicts for which the differences were smallest (permutation test for paired replicates, one-tailed, $P = 0.22$).

The conflicts appear to be more-or-less difficult to forecast, regardless of the forecasting method used. That is, forecasting methods each rank the conflicts in similar orders of percent correct forecasts. The Kendall coefficient of concordance for the ranking of the eight conflicts by the four methods is, after adjusting for ties, 0.888 (equation 9.18b; Siegel and Castellan, 1988). The coefficient is close to the level of complete agreement ($W = 1.0$) and is highly significant ($X^2 = 24.9$, df = 7, $P < 0.001$; from equation 9.19). An alternative measure of agreement between the rankings given by the methods, the average of the Spearman rank-order correlation coefficients between the six possible pairs of forecasting methods, has a value of 0.850 (equation 9.16; Siegel and Castellan, 1988). For four sets of rankings, such as are considered here, the Spearman measure can take on values between negative one-third and positive one.

The question of whether it is possible to determine, *a priori*, if a conflict will be relatively easy or difficult to forecast is examined later in this subsection.

Analogy forecast rule

The structured-analogies participants were, after a number of formal steps, asked to "make your prediction" (Appendix 6), but were not told how they should do this. It is unsurprising then, that some of the participants provided forecasts that were clearly inconsistent with their own analogies. For these cases, I used my judgement to derive 19 single-decision structured-analogies forecasts from participants' own analogies. That is, where possible, I selected as a structured-analogies forecast the decision that was suggested by the highest-rated analogy or that was suggested by the weight of analogical evidence.

When both solo and joint forecasts were included, there were five conflicts (Artists Protest, Distribution Channel, 55% Pay Plan, Telco Takeover, Zenith Investment) for which there were at least two inconsistent forecasts – 16 forecasts in all. On the basis of unweighted averages of the percentage of correct forecasts for the five conflicts, the participants' inconsistent (classified as unaided-judgement) forecasts were correct for 13 percent of forecasts whereas the structured-analogies forecasts that I derived from the participants' analogies were correct for 56 percent of forecasts. Given the finding that structured-analogies forecasts were more accurate than unaided-judgement forecasts, this is not surprising. While the difference in accuracy was large, it was not statistically significant at conventional levels (permutation test for paired replicates, one-tailed, $P = 0.06$).

Should judgement be used for deriving a prediction from analogies data? In order to answer this question, I set rules for choosing a single-decision forecast from the analogies data that were assembled using the structured-analogies method (Figure 1). I used the rules to derive forecasts from the analogies data and compared the accuracy of these forecasts with the participants own structured-analogies forecasts together with those I had derived from participants' data using my judgement.

Figure 1

**Rules for choosing a single-decision forecast
from a set of up to five analogies that have been rated for similarity to a target conflict**
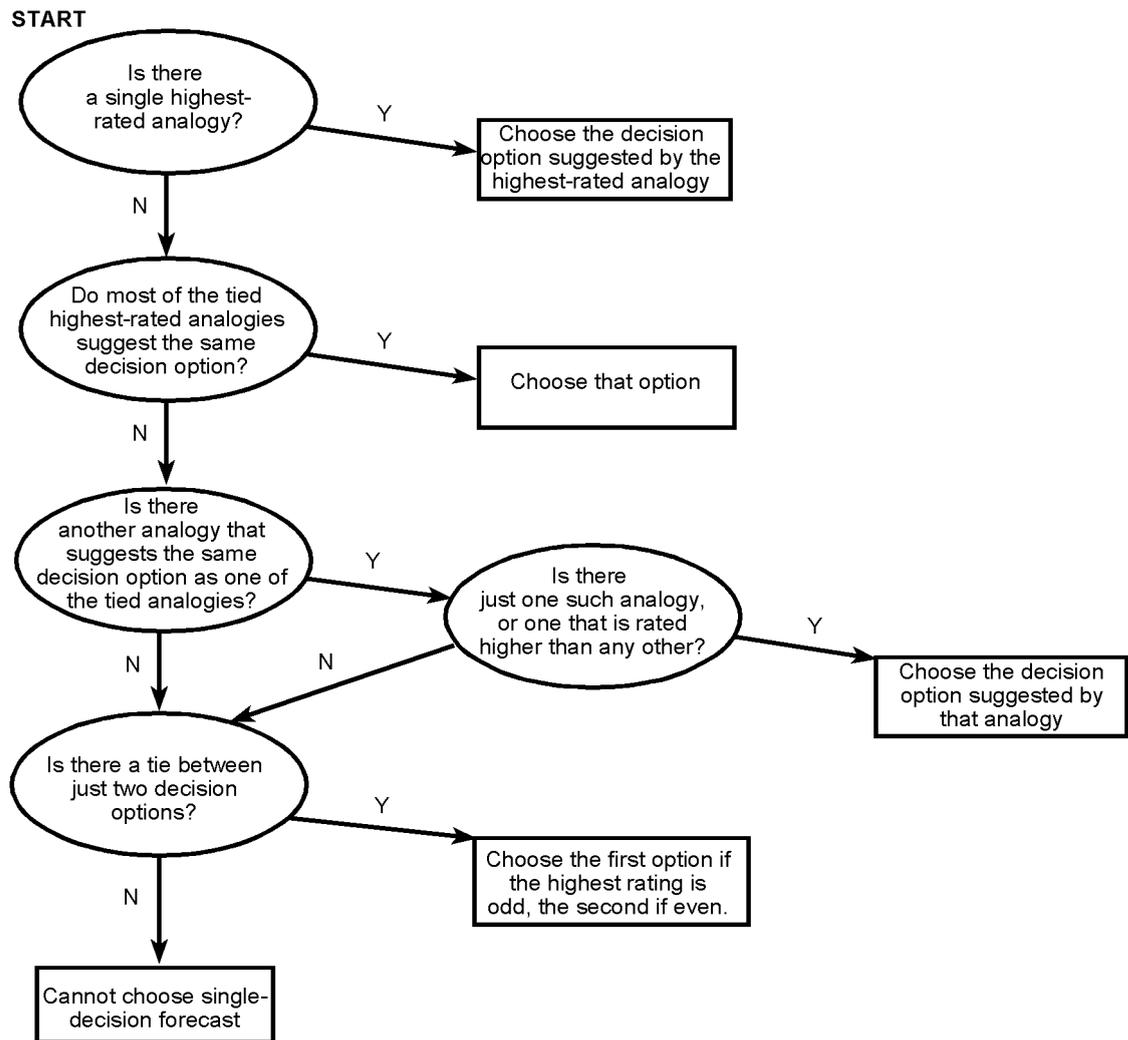
There were no meaningful differences in accuracy between the rule-based structured-analogies forecasts and the judgement-based structured-analogies forecasts. Rule-based forecasts were 44 percent correct (unweighted average across conflicts, n = 61) and, from Table 21, the structured-analogies forecasts from solo-experts were 45 percent correct (n = 60). If collaborative forecasts were included, accuracy would be 46 percent (n = 83) in both cases. Had I accepted forecasts that were at odds with participants' own analogies as structured-analogies forecasts, however, overall accuracy for the method would have been lower at 40 percent correct (n = 82).

Probabilistic forecasts

When, as is the case with this research, forecasters have to choose between several outcome options, some might prefer to assign probabilities to the options rather than to choose one of them. Comparisons of accuracy are straightforward when forecasters each choose a single option, but not when probabilistic forecasts are provided. Brier scores are used to assess the relative accuracy of probabilistic forecasts and are recommended for this purpose by authors including Armstrong (2001g), Brier (1950), Doggett (1998), and Lichtenstein, Fischhoff, and Phillips (1982). These authors define the Brier score as an average of the sums of the squared errors of probabilistic forecasts. For one set of probabilities for the possible outcomes of a single event, the formula for the Brier score (*BS*) is more succinct (the average calculation is avoided) and, arguably, more easily comprehended. On that basis the measure is…

Formula 2
Brier score (*BS*)

$$BS = (\mathbf{f} - \mathbf{a}) \cdot (\mathbf{f} - \mathbf{a})$$

Where:
$\mathbf{f}$ = probabilistic-forecast vector for the k outcome options of the forecast objective, $(f_1, \ldots, f_k)$
$\mathbf{a}$ = actual-outcome vector for the k options of the forecast objective, $(a_1, \ldots, a_k)$, where the element representing the actual outcome is coded as one and all other elements are coded as zero.

The Brier score can take on any value between zero and two, where a completely accurate forecast would have a Brier score of zero and one that was completely inaccurate would have a Brier score of two[12]. That is, a forecast that allocates 1.00 to the outcome option that actually occurs will have a *BS* of 0.00, irrespective of the number of options. In a case with four options, the Brier score for such a forecast could be represented as *BS*(1*,0,0,0) = 0.00, where the asterisk marks the actual outcome. A

---

12 A potential source of confusion over the Brier score, is that it has also been formulated as the squared error of the probability assigned to a single event that may or may not occur (for example, Fuller, 2000). Using this alternative formulation, a forecast of 0.90 chance of rain tomorrow would have a *BS* of 0.01 if it did rain $((0.90-1.00)^2)$ and 0.81 if it did not $((0.90-0.00)^2)$. With this formulation, the Brier score can take on any value between zero and one.

forecast that allocates 1.00 to another option will have a *BS* of 2.00: for example, *BS*(0*,1,0,0) = 2.00. The actual-outcome vector would be (1,0,0,0) in both examples.

Armstrong (2001e) examined the evidence on measures for evaluating forecasting methods and found that measures based on squaring forecast errors were not reliable and were difficult to interpret (also Mean Squared Error entry in Armstrong, 2001g). On the other hand, Armstrong (2001e) found that the relative absolute error (*RAE)* measure performed well against other error measures.

Armstrong's discussion and formulation of the *RAE* is based on forecasting of time series, rather than on probabilistic forecasts of the outcomes of situations. I have defined an analogous measure for the latter purpose, and have called it the probabilistic forecast accuracy rating (*PFAR*) in this document. It is the sum of the absolute errors of a set of probabilities assigned to the outcome options of a situation, divided by the sum of the absolute errors from a naïve forecast, or…

Formula 3
Probabilistic forecasting accuracy rating (*PFAR*)

$$PFAR = \frac{|\mathbf{f} - \mathbf{a}| \cdot \mathbf{u}}{2(k - 1)/k}$$

Where:
$\mathbf{f}$ = probabilistic-forecast vector for the k outcome options of the forecast objective, $(f_1, \ldots, f_k)$
$\mathbf{a}$ = actual-outcome vector for the k options of the forecast objective, $(a_1, \ldots, a_k)$, where the element representing the actual outcome is coded as one and all other elements are coded as zero
$\mathbf{u}$ = unity vector for the k options of the forecast objective, $(u_1, \ldots, u_k)$, where each element is equal to one.

The naïve probabilistic forecast is the even allocation of probabilities across outcome options. The formula for the sum of the absolute errors from the naïve forecast reduces to the term $2(k - 1)/k$ in the *PFAR* formula. Like the Brier score, *PFAR* can take on any value between zero and two but, unlike the Brier score, the maximum value (for a completely inaccurate forecast) declines towards an asymptote at one as the number of outcome options increases from two towards infinity.

In Appendix 13, I compare the Brier score and *PFAR* measures. In sum, Brier scores are biased against forecasts with polarised probability allocations (one outcome option allocated a high probability) and against forecasting for situations with more rather than fewer outcome options. Consequent on the second bias, Brier scores cannot legitimately be compared across situations with different numbers of outcome options. These problems may lead to poor decisions when evaluating forecasting methods. As I show in Appendix 13, the *PFAR* measure does not have these problems, and so I have used *PFAR* scores to assess the relative accuracy of probabilistic forecasts.

All expert participants in the unaided-judgement and structured-analogies treatments were given the option of providing probabilistic forecasts. Although few provided explicitly probabilistic forecasts, because they had the option of doing so it seems reasonable to assume that in cases where participants selected a single option they were implicitly allocating a probability of one to their choice and of zero to all other options. In other words, all forecasts that were provided when the forecaster had an option of providing a probabilistic forecast could be regarded as probabilistic forecasts. It is on that basis that I conduct the following analysis of the data presented in Table 22. To maintain consistency with the percent correct analysis, forecasts are by solo experts.

I set out to answer three questions. First, does including participants' probabilities improve accuracy compared to simply re-coding their first choices (for each forecast, the option allocated the highest probability) as one? Second, does the inclusion of participants' probabilities lead to a different conclusion about the relative accuracy of experts' forecasts from unaided judgement compared to structured analogies than was concluded from percent correct data? Third, in the case of structured-analogies forecasts, do probabilities derived from participants' analogy decisions and ratings using a rule tend to be more accurate than first-choice-set-to-one and participants' probabilities forecasts?

<div align="center">

Table 22

**Probability forecast accuracy ratings of solo-experts' forecasts
by forecasting method and derivation of probabilities** [a]

Average *PFAR* (number of forecasts)

</div>

	Unaided judgement		Structured analogies		
	One option set to 1.0 [b]	Participants' probabilities	One option set to 1.0 [b]	Participants' probabilities	Rule-based probabilities
Telco Takeover	1.19 (9)	**1.09** (10)	1.14 (7)	1.11 (6)	1.26 (8)
55% Pay Plan	1.11 (12)	1.11 (12)	.83 (8)	.83 (8)	**.82** (8)
Artists Protest	1.08 (20)	1.08 (20)	.96 (5)	.96 (5)	**.87** (6)
Zenith Investment	1.07 (14)	1.04 (14)	**.86** (7)	1.05 (6)	.97 (7)
Distribution Channel	1.00 (18)	.94 (19)	**.75** (9)	.98 (7)	.90 (9)
Personal Grievance	**.67** (4)	.80 (4)	.78 (12)	.80 (12)	.80 (12)
Water Dispute	.60 (5)	.70 (6)	.38 (4)	.38 (4)	**.20** (4)
Nurses Dispute	.43 (14)	.49 (15)	**.38** (8)	.43 (7)	.56 (8)
Totals (median) [c]	1.04 (96)	.99 (100)	**.81** (60)	.90 (55)	.85 (62)

a The probabilities from which the *PFAR*s were calculated were derived in three different ways: 1/ see note b; 2/ the participants' own probabilities were used unchanged; 3/ probabilities were derived from participants' analogy decisions and ratings using the rule described in Appendix 13 and illustrated in Table 48. In all cases, forecasts of C for Distribution Channel were re-coded with half allocated to A and half to B

b When a participant allocated the highest probability for a conflict to a single option, that option was re-coded as one and the rest as zero. When participants' probabilities were inconsistent with their own analogies, any forecasts from their probabilities were coded to unaided judgement and, where this was reasonable, single forecasts were derived from the analogies

c Figures in this row are medians of the *PFAR*s reported for each conflict.

In answer to the first question, forecasts incorporating participants' probabilities were, on average, no more accurate than participants' first choices. For unaided-judgement forecasts, the median *PFAR* score was only slightly lower for participants' probabilities (0.99) than for first choices (1.04). Both medians were close to one, which indicates that the forecasts were on average no better than chance. The differences between the scores for the conflicts were not statistically significant (permutation test for paired replicates, one-tailed, $P = 0.72$; Siegel and Castellan, 1988). For structured-analogies forecasts, the median *PFAR* score was higher (that is, worse) for participants' probabilities. The *PFAR* score was 0.90 for participants probabilities compared to 0.81 for first choices.

The answer to the second question is no. Inclusion of participants' probabilities does not lead to a different conclusion about the relative accuracy of solo experts' structured-analogies forecasts compared to solo expert's unaided-judgement forecasts than does percent correct data. Generally, first-choice forecasts by experts using structured analogies were more accurate than experts' first-choice unaided-judgement forecasts

(permutation test for paired replicates, one-tailed, $P = 0.02$). As with the comparison by percent correct, structured-analogies forecasts were more accurate than unaided-judgement forecasts for conflicts other than Personal Grievance. Further, forecasts from structured analogies that incorporated participants' probabilities were more accurate, generally, than forecasts from unaided-judgement with participants' probabilities (permutation test for paired replicates, one-tailed, $P = 0.078$).

Finally, the answer to the third question is also no. While the median *PFAR* score for rule-based probabilities was lower (better) than that for participants' probabilities from structured analogies, there was considerable variability across conflicts and hence the difference was not statistically significant (permutation test for paired replicates, one-tailed, $P = 0.32$). Further, rule-based probability forecasts from structured analogies were less accurate than first-choices from experts' structured-analogies forecasts.

Percentage error reduction versus chance

Conflicts vary in the number of plausible decisions that are of interest to managers. This number is partly a function of the conflict and partly of client or forecaster judgement. For example, instead of the three options I used for Water Dispute, the United Nations might have simply wanted to know whether war would be declared or not (two options). On the other hand, it might have been important for the United Nations to know not only whether war would be declared but, should it be declared, whether the commitment by the parties would be major or minor and whether Deltaland would seek to capture and hold territory or would simply attempt to force concessions on Midistan (five options). In the two-option case there is a 50 percent chance (1/2) that a forecast will be wrong while in the five-option case there is an 80 percent chance (4/5). Note that calculating the chance of an inaccurate forecast from the number of decision options is a form of naïve predictability assessment. A measure of forecast accuracy that discounts such *a priori* forecasting difficulty is the percentage error reduction versus chance (Formula 4). As well as insensitivity to *a priori* difficulty, *PERVC* is independent of scale as it is calculated as a percentage rather than a level.

<div align="center">

Formula 4
Percentage error reduction vs chance (*PERVC*)

$$\mathbf{PERVC}_{SM} = (\mathbf{COE}_S - \mathbf{Error}_{SM}) / \mathbf{COE}_S * 100$$

</div>

Where:
  COE = percentage chance of error: (1 - (1 / number of options)) * 100
  Error = inaccurate forecasts as a percentage of all forecasts
  S = conflict situation
  M = method

The *PERVC* transformation does not alter the rankings of methods by accuracy for each of the conflicts (Table 23). Consequently, the results of statistical tests are identical to those conducted for the Table 21 data.

<div align="center">

Table 23
Accuracy of forecasts: Percent error reduction versus chance (*PERVC*) [a]

</div>

	Chance of error	Unaided judgement	Game theorist	Structured analogies	Simulated interaction
Nurses Dispute	*67*	57	25	63	73
Water Dispute	*67*	40	51	63	85
Distribution Channel	*67*	0	-15	25	63 [b]
Zenith Investment	*67*	-6	-16	15	39
Median		**20**	**5**	**44**	**68**
Personal Grievance	*75*	33	47	23	47
55% Pay Plan	*75*	-11	5	17	47 [b]
Telco Takeover	*75*	-33	-33	-15	20
Artists Protest	*83*	-8	-13	4	14 [b]
Median		**-10**	**-4**	**10**	**33**
Totals [c]		**-3**	**-4**	**20**	**47**

a All forecasts are by individual experts except those from simulated interaction which, apart
   from four Nurses forecasts from groups of experts, are from groups of novices. Four
   probabilistic forecasts that could not be coded as single-decision forecasts are not included
b Forecast accuracy data reported in Armstrong (2001a)
c Figures in this row are medians of the *PERVC*s reported for each conflict.

The data in Table 23 show an interesting pattern: *PERVC* was substantially lower when participants were presented with four or more decision options than it was when they were presented with three. For example, consider the average error rate (percent *in*correct – see Table 21 for percent correct) for simulated-interaction forecasts. For conflicts with four or more decision options the average error rate (53 percent) was more than twice that for forecasts of conflicts with fewer than four options (24 percent). Artists Protest was the only conflict that did not have either three or four decision

options. When it is excluded from the comparison, the average error rate figures are 47 percent and 24 percent. This is surprising as, on the basis of the number of options, the *a priori* chance of forecasting inaccurately is only 11 percent (one-ninth) lower when there are three options rather than four[13].

Judgements of a priori *predictability*

While an expectation of forecast accuracy might reasonably be derived from a knowledge of the forecasting method used and the number of decision options provided, it is clear from Table 23 that such an expectation would be an unreliable predictor. Greater precision for estimates of likely forecast accuracy should help managers to make better decisions. Kahneman and Tversky (1982) suggested, in the absence of records of past predictions and outcomes for an appropriate reference class, that experts' subjective assessments of predictability should be used.

In order find out whether assessments of the predictability of decisions in conflicts might be useful, I obtained 84 ratings from 28 university students attending a third-year strategy course lecture. I describe this process, and why it was reasonable to obtain ratings from students rather than from experts, in Appendix 14.

In the event, there was little variation between the averages of the students' ratings for the various conflicts. When the responses are scored from one for "very unlikely" to five for "very likely", the eight averages have a range of little more than one (from 2.8 to 3.9). Further, there was a great deal of variation in the ratings given for the conflicts other than Artists Protest.

I tested the reliability of the ratings using the *Kappa coefficient of agreement K* (Siegel and Castellan, 1988). The coefficient *K* is zero if there is only chance agreement between ratings and one if there is complete agreement. I calculated *K* using 10 ratings per conflict by excluding the first rating for cases where I had obtained 11. On this basis, the value of *K* was 0.05. With little agreement between the raters (*K* near zero) and little variation in average ratings, *a priori* subjective ratings of predictability do not provide useful forecasts of forecast accuracy.

---

13 Calculated as $((2/3) / (3/4) - 1) * 100 = -1/9 * 100$.

The accuracy of unaided-judgement forecasts is another **C** *ex post* in this case **C** benchmark of forecasting difficulty. This seems a sensible benchmark, as unaided judgement is likely to be used in most conflict forecasting (Armstrong, et al., 1987). A measure of forecast accuracy that discounts unaided-judgement forecast accuracy is the percentage error reduction versus unaided judgement or *PERVUJ* (Formula 5).

Formula 5
Percentage error reduction vs unaided judgement (*PERVUJ*)

$$\textbf{PERVUJ}_{SM} = (\textbf{Error}_{S;M=UJ} - \textbf{Error}_{SM}) / \textbf{Error}_{S;M=UJ} * 100$$

Where:
      Error = inaccurate forecasts as a percentage of all forecasts
      S = conflict situation
      M = method
      UJ = unaided-judgement method

As with *PERVC*, the *PERVUJ* transformation does not alter the rankings of methods by accuracy for each of the conflicts (Table 24). The details of the Page test for ordered alternatives (Siegel and Castellan, 1988) are different to those of the test conducted on the Table 21 data – there are three and not four alternatives being tested. Nevertheless, the result is the same ($N = 8$, $k = 3$, $L = 110$, $P < .001$). The results of statistical tests on paired replicates are identical to those conducted for the Table 21 data.

Table 24
Accuracy of forecasts:
Percent error reduction versus unaided judgement (*PERVUJ*) [a]

	Error from unaided judgement	Game theorist	Structured analogies	Simulated interaction
Telco Takeover	*100*	0	14	40
Artists Protest	*90*	-4	11	21 [b]
55% Pay Plan	*83*	14	25	52 [b]
Zenith Investment	*71*	-10	20	42
Median	***87***	**-2**	**17**	**41**
Distribution Channel	*67*	-15	25	63 [b]
Personal Grievance	*50*	20	-16	20
Water Dispute	*40*	18	38	75
Nurses Dispute	*29*	-72	14	38
Median	***45***	**2**	**20**	**50**
Totals [c]		**-2**	**17**	**41**

a All forecasts are by individual experts except those from simulated interaction which, apart from four Nurses forecasts from groups of experts, are from groups of novices. Four probabilistic forecasts that could not be coded as single-decision forecasts are not included
b Forecast accuracy data reported in Armstrong (2001a)
c Figures in this row are medians of the PERVUJs reported for each conflict.

The conflicts in Table 24 are ordered from most to least difficult to forecast using unaided judgement. In the case of the four conflicts that experts using unaided-judgement found hardest to predict, they were almost always wrong (median 87 percent). For the other four conflicts, experts using unaided judgement were right more often than they were wrong – just. The distinction between conflicts that were hardest to predict and those that were easier to predict using unaided judgement does not hold for the other three forecasting methods. Median *PERVUJ* figures for the four conflicts that were hardest to predict using unaided judgement were *lower* than for the four easiest to predict conflicts.

The Table 24 data are surprising, as it might reasonably be supposed that: (a) the conflicts that were chosen for use in this research were biased towards those that were interesting; (b) such conflicts were found to be interesting, at least in part, because the decisions that were made were surprising to the researchers; and (c) decisions that were surprising to the researchers would, *ipso facto*, be relatively more difficult to forecast using unaided judgement than they would using a formal forecasting method. As the preceding discussion and the data shown in Table 24 demonstrate, this hypothesis on relative difficulty is not supported.

4.1.2   Effect of method on forecast usefulness

Method

To test the possibility that forecasts that are not entirely accurate may nevertheless still be useful, I asked independent raters to read descriptions of the actual decisions that were made in the conflicts. After reading the actual decisions, they rated for usefulness the decision options that had been provided to forecasters and role-players on a zero-to-ten scale. It was explained to raters that a decision option that matched the actual decision should be given a rating of 10, a decision that was the opposite of the actual decision should be rated *zero*, and all other decisions should be given some intermediate value. A copy of the questionnaire I used for this purpose is attached as Appendix 15.

Ten students from a third-year strategic management class completed the questionnaire during 30 minutes of class time. These student raters did not consistently interpret the task as I had intended. For example, several students rated all decision options for a conflict as 10, giving reasons such as "they are all important things to know". Others rated decision options that matched the actual decision lower than options that did not. One gave a reason for doing this: "a prediction of a long strike [the actual outcome for 55% Pay Plan] may have encouraged it". It is possible that the strategic management lecture that preceded the students' participation led them to interpret the task in ways that I had not intended.

Rather than use my judgement to select responses that appeared to result from interpreting the instructions in the way that I had intended, I rejected the students'

responses and recruited a convenience sample of five mature professional people to complete the task. I sent them the questionnaire by email. The questionnaire was identical to the one used for the students. Although it included the text "you have about three minutes for each conflict" I did not repeat this instruction in my communications to the participants, nor did I ask them how long they spent on the task. As I had hoped, these participants appeared to interpret the task in ways that were consistent with my intentions.

Response

The five participants each provided ratings for all of the 31 decision options provided for the conflicts. I tested the inter-rater reliability of the ratings using the Kappa coefficient of agreement K (Siegel and Castellan, 1988). The coefficient K can take on any value between zero and one: zero where there is no more than chance agreement and one where there is complete agreement. While modest, the level of agreement between the raters ($K = 0.249$) is statistically highly significant ($z = 7.89$, $P < 0.001$).

I used the median of the raters' ratings for each decision option as the usefulness rating for that option. I used medians in order to avoid the influence of outlier ratings and so that any decision option that was rated as 10 by a majority of raters would receive a usefulness rating of 10. The usefulness ratings for the decision options are shown in the copy of the rating questionnaire attached as Appendix 15.

In the cases of Personal Grievance, Nurses Dispute, and Water Dispute, none of the decision options received a median rating of 10. I had anticipated that this would be the case with Personal Grievance as I considered that at least two of the four decision options, and possibly a third, could be regarded as partly accurate and none as entirely accurate. Although I had anticipated imposing ratings of 10 on decision options that I considered matched the actual decisions, in the case of Nurses Dispute I was sufficiently persuaded by the reasons given by respondents who had not rated these options as 10 to stick with the median ratings. Raters giving ratings of less than 10 argued that the option I considered matched the actual decision was: "incomplete in that it does not give information about where the parties met"; "literally correct, but 'wishy-washy' (vague) and really no better than the other two"; and "a compromise is a subtle thing". In the

case of Water Dispute, I did not find the arguments of the raters who gave ratings of less than 10 persuasive. They argued that the option I considered matched the actual decision: "…implies that the amount will be sufficient to meet the needs – which is not stated in the actual decision"; "explains that Midistan will release water, but doesn't say how"; and "reflects part of what happened". Nevertheless, rather than impose a rating of 10 for the decision I considered matched the actual decision solely on the basis of my judgement, I adopted the raters' median rating.

Table 25 shows the average usefulness ratings of forecasts from the four forecasting methods for each of the conflicts. The average usefulness ratings are the averages of the usefulness ratings associated with participants' predicted decision options. Instead of assigning one to correct predictions and zero otherwise (as with the Table 21 data) each prediction was assigned the usefulness rating for the option that was chosen.

Table 25
Accuracy of forecasts: Average usefulness rating out of 10 [a]

	Expected usefulness	**Unaided judgement**	**Game theorist**	**Structured analogies**	**Simulated interaction**
Telco Takeover	2.5	0.0	0.0	1.4	**4.0**
Artists Protest	3.7	4.0	3.8	4.6	**5.4** [b]
Zenith Investment	5.0	4.3	3.3	6.4	**7.4**
55% Pay Plan	4.5	4.8	5.3	6.3	**6.8** [b]
Water Dispute	2.7	4.8	5.3	6.0	**7.2**
Distribution Channel	5.5	5.6	5.3	6.8	**8.5** [b]
Nurses Dispute	4.0	6.3	5.0	6.5	**6.9**
Personal Grievance	4.8	6.5	**6.6**	5.4	**6.6**
Totals (unweighted) [c]	4.1	4.5	4.3	5.4	**6.6**

a All forecasts are by individual experts except those from simulated interaction which, apart from four Nurses forecasts from groups of experts, are from groups of novices. Four probabilistic forecasts that could not be coded as single-decision forecasts are not included

b Figures based on forecast accuracy data reported in Armstrong (2001a), a personal communication with J. Scott Armstrong on 11 January 2002 regarding distribution of inaccurate responses for Artists Protest, and an assumption that in the absence of contradictory information inaccurate forecasts were evenly distributed: Artists Protest 4, 5, 4, 0, 0, 1; Distribution Channel 9, 1, 1, 1; 55% Pay Plan 6, 1, 1, 2

c Percentage figures in this row are unweighted averages of the rating reported for each conflict.

In Table 25, the methods are presented in the same order as they were in Table 21. A comparison with Table 21 shows that the ranking of the methods by usefulness ratings (Table 25) is the same as the ranking by percent correct (Table 21). As a consequence of this, the Page test for ordered alternatives gives the same result (N = 8, $k = 4$, $L = 231.5$,

$P < .001$). Further, the difference in usefulness between the simulated-interaction and the structured-analogies forecasts is statistically significant (permutation test for paired replicates, one-tailed, $P = 0.004$; Siegel and Castellan, 1988). Also statistically significant is the difference between the usefulness of structured-analogies and game-theorist forecasts (permutation test for paired replicates, one-tailed, $P = 0.02$).

4.2   Generalisability

4.2.1   Effect of collaboration on accuracy

I obtained responses from 12 participants who collaborated on 15 unaided-judgement forecasts and 23 structured-analogies forecasts. These participants collaborated with between one and seven other people – a median of one other person.

Collaborating experts using their unaided judgement were more accurate than individual experts for one of four conflicts, but were less accurate for two. There was one draw (Table 26). Collaborating experts using the structured-analogies method were more accurate than individual experts for four of six conflicts, and were less accurate for two.

Table 26
Effect of collaboration on experts' forecast accuracy [a]
Percent correct forecasts (number of forecasts)

	Unaided judgement				Structured analogies			
	Solo		**Joint**		**Solo**		**Joint**	
Telco Takeover	-	-	-	-	**14**	(7)	0	(2)
Artists Protest	**10**	(20)	0	(4)	20	(5)	**50**	(4)
55% Pay Plan	**17**	(12)	0	(4)	38	(8)	**40**	(5)
Zenith Investment	29	(14)	**100**	(2)	-	-	-	-
Distribution Channel	33	(18)	33	(3)	50	(9)	**67**	(3)
Personal Grievance	-	-	-	-	42	(12)	**50**	(2)
Nurses Dispute	-	-	-	-	**75**	(8)	60	(5)
Totals (unweighted) [b]	22	(64)	33	(13)	40	(49)	45	(21)

a Conflicts are included in the table if there were two or more forecasts for both levels
    of the variable
b Percentage figures in this row are unweighted averages of the percent correct
    forecasts reported for each conflict.

While joint-forecasts from both unaided judgement and structured analogies were, on average, more accurate than solo-forecasts from the same methods, there is insufficient

evidence to conclude that the differences are other than the result of chance. Had the joint unaided-judgement forecasts for the four conflicts each been more accurate than the corresponding solo forecasts, there would still have been insufficient data to achieve a conventional level of statistical significance using the permutation test. The difference between joint and solo forecasts from structured analogies is quite small and not statistically significant (permutation test for paired replicates, one-tailed, $P = 0.27$).

There were some marked differences between the characteristics of joint forecasters and solo forecasters. Joint forecasters tended to rate their experience with similar conflicts more highly than did solo forecasters. Joint forecasters had a median of 14 years conflict management experience while solo forecasters had a median of four years. Finally, on balance, joint forecasters reported spending more time deriving their forecasts than did solo forecasters (Table 27).

Table 27
**Characteristics of structured-analogies forecasts and forecasters
by collaboration**
Median statistics [a]

	Time to complete task (minutes)		Confidence in forecast (percent) [b]		Conflict management experience (years)		Experience with similar conflicts (0-10)	
	Solo	Joint	Solo	Joint	Solo	Joint	Solo	Joint
Telco Takeover	45	45	**15**	5	2	**13**	**3.5**	3.0
Artists Protest	30	**45**	**30**	23	10	**15**	2.0	**5.0**
55% Pay Plan	**30**	18	20	20	5	**20**	2.0	2.0
Distribution Channel	30	30	40	**50**	5	5	2.0	**5.0**
Personal Grievance	30	**60**	**15**	0	2	**13**	**5.5**	3.5
Nurses Dispute	30	30	20	20	0	**20**	4.0	**5.0**
Totals (unweighted) [c]	30	**38**	20	20	4	**14**	2.8	**4.3**

a Numbers of forecasts are shown in Table 26.
b Self-assessed likelihood of changing forecast if more time were available for the task
c Percentage figures in this row are unweighted medians of the statistics reported for
   each conflict.

As there is insufficient data to disentangle differences in participant characteristics and behaviour from any genuine improvement in accuracy that may arise from collaboration, I have confined the remainder of my analysis to the forecasts of solo-forecasters.

4.2.2   Effect of expertise on accuracy

Novices and experts

The data permit only one comparison between the accuracy of experts and that of novices (students). That is for unaided-judgement forecasts. Experts were more accurate than novices for five of eight conflicts, while novices were more accurate for two. There was one draw. Overall, there was little difference between the accuracy of experts' unaided-judgement forecasts and that of novices (Table 28).

Table 28
Accuracy of experts' and novices'
unaided-judgement forecasts
Percent correct forecasts (number of forecasts)

	Novices		Experts	
Artists Protest	5	(39)[a]	**10**	(20)
Distribution Channel	5	(42)[a]	**33**	(18)
Telco Takeover	**10**	(10)	0	(9)
55% Pay Plan	**27**	(15)[a]	17	(12)
Zenith Investment	29	(21)	29	(14)
Personal Grievance	44	(9)	**50**	(4)
Water Dispute	45	(11)	**60**	(5)
Nurses Dispute	68	(22)	**71**	(14)
Totals (unweighted) [b]	29	(169)	34	(96)

a Forecast accuracy data reported in Armstrong (2001a)
    except for eight Artists Protest forecasts (one correct) and
    five Distribution Channel forecasts (one correct)
b Percentage figures in this row are unweighted averages of
    the percent correct forecasts reported for each conflict.

Approach to analysis of expertise

In the remainder of my analysis of expertise, I examine the effect of the relative expertise, and length of experience, of expert participants. Because there were no forecasts for many combinations of conflict and variable level, it was necessary to re-code variables into two levels in order to analyse the available data.

The following variables were re-coded as dichotomies for this purpose: (a) participants' self-assessed level of experience with similar conflicts to the one they were forecasting; (b) participants' years-of-experience as conflict management experts; (c) participants'

111

confidence (self-assessed likelihood of changing a forecast given more time); (d) analogy similarity to a target conflict, and (e) number of analogies.

I chose the cut-points of each variable for this re-coding so that roughly half of all experts' forecasts, across all methods, fell either side of the cut-points. Doing this maximises the chances that, given the data available, there would be forecasts for both levels of the variables for each conflict (subject to the constraint that the same cut-points were used for each of the three methods used by expert participants). The more conflicts for which comparisons are possible, the greater are the chances of obtaining statistically significant findings.

It is logically possible that the choice of cut-points for re-coding could lead to erroneous conclusions about the relationship between a variable that was re-coded and forecast accuracy. This possibility does decrease with the number of forecasts and the number of conflicts but, in order to reduce the chances of mistaken conclusions further, I also compare the average levels of the variables (by conflict) for accurate and inaccurate forecasts. Further, I compare the accuracy of the most experienced participants with average accuracy for each of the methods used by experts.

Experts' experience – unaided judgement

I re-coded participants' self-assessed level of experience with similar conflicts to the one they were forecasting into low (0-4) and high (5-10).

Overall, experts using unaided judgement who identified themselves as having high levels of experience with conflicts that were similar to the target conflict were less accurate than those with low levels of experience with similar conflicts. Those experts using unaided judgement who had five or more years experience as conflict management experts were no more accurate than those who had fewer than five years experience (Table 29).

Table 29
**Effect of experience
on the accuracy of experts' unaided-judgement forecasts** [a]
Percent correct forecasts (number of forecasts)

	Experience with similar conflicts (0-10)		Conflict management experience (years)	
	Low (0-4)	High (5-10)	Under 5 years	5+ years
Artists Protest	6 (16)	**25** (4)	10 (10)	10 (10)
55% Pay Plan	**22** (9)	0 (3)	14 (7)	**20** (5)
Zenith Investment	- -	- -	17 (6)	**38** (8)
Distribution Channel	- -	- -	29 (7)	**36** (11)
Nurses Dispute	64 (11)	**100** (3)	**86** (7)	57 (7)
Personal Grievance	**100** (2)	0 (2)	- -	- -
Totals (unweighted) [b]	48 (38)	31 (12)	31 (37)	32 (41)

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable
b Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

Accurate forecasts were provided by participants who rated their experience with conflicts similar to the one they were forecasting somewhat lower, on average, than did those who provided *in*accurate forecasts (Table 30). The difference is not great, however, and accurate forecasts for four out of six conflicts were provided by participants who rated their experience with similar conflicts higher, on average, than those who provided inaccurate forecasts.

Participants who provided accurate forecasts had, on average, less experience as conflict management experts (3.3 years) than did participants who provided inaccurate forecasts (6.8 years).

Table 30
Forecaster characteristics associated with accurate and inaccurate unaided-judgement forecasts by experts [a]

	n	Experience with similar conflicts (0-10) [b]		Conflict management experience (years) [c]	
		Inaccurate	Accurate	Inaccurate	Accurate
Artists Protest	20	2.0	**3.0**	3.5	**4.5**
55% Pay Plan	12	**1.8**	0.5	2.0	**13.3**
Nurses Dispute	14	0.0	**2.9**	**18.5**	2.0
Personal Grievance	4	**6.5**	0.0	**10.0**	0.0
Water Dispute	5	0.5	**1.0**	5.0	0.0
Zenith Investment	14	0.8	**1.5**	8.5	8.5
Totals (unweighted) [d]	69	1.9	1.5	6.8	3.3

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable. Distribution Channel is excluded because some forecasters chose a combination decision that could not be conveniently coded for this analysis
b Averages
c Medians
d Forecaster characteristic figures in this row for "similar conflicts" are averages of the figures reported for each conflict and for "conflict management" are medians.

Some participants were very experienced indeed. Ten of the experts who provided unaided-judgement forecasts claimed twenty or more years of conflict management experience. As with all expert participants, they were able to chose which conflicts to forecast (none forecast Water Dispute) and, game theorists excepted, whether or not to collaborate (five of their forecasts were collaborative). Taking their expertise at face value, it is reasonable to assume that they used the choices they were given wisely. It follows from this that a calculation of average forecast accuracy that included all their forecasts would bias any comparison in favour of these veterans' forecasts. In spite of this, the veteran experts' forecasts were substantially less accurate than the overall

*un*weighted-average for solo-expert unaided-judgement forecasts of 34 percent (Table 21). Of their 27 forecasts, only 17 percent were accurate on the basis of a weighted average.

The term "conflict management expert" is one that members of the International Association of Conflict Management will readily have identified with. In my pre-testing of materials, one respondent who did not have this background reacted strongly when I questioned his response that he had many years of conflict management experience. He had interpreted the term broadly and his reply was that he had spent all of his working life as a conflict management specialist. On the strength of that response, I included the question about conflict management experience in the questionnaires that I sent to all expert unaided-judgement and structured-analogies participants.

It is possible, however, that some non-IACM members may not have interpreted the term in the same broad way that my pre-test respondent had done. The IACM respondents had an average of 13 years conflict management experience compared to five years for other respondents. Yet other respondents had slightly more experience with similar conflicts (average rating of 2.5) than had IACM respondents (2.1). There was no significant difference in the accuracy of forecasts from the two groups. (All figures are unweighted averages across conflicts). Given these statistics, a cautious interpretation of the findings that relate to conflict management experience would be sensible.

Experts' experience – game theorists

I asked game theorists about their experience with similar conflicts only in relation to Personal Grievance, Telco Takeover, and Water Dispute. Among the 12 responses to this question there were none that indicated the participant had a "high" level of experience with similar conflicts. This precludes useful analysis of that data.

Game-theorist participants with five or more years experience as game-theory experts were more accurate than those with fewer than five years experience for two conflicts, but were less accurate for three. Overall, there was little difference between the accuracy of those with more experience and those with less experience (Table 31).

Table 31
**Effect of experience as a game theorist
on the accuracy of game-theorist forecasts** [a]
Percent correct forecasts (number of forecasts)

	Under 5 years		5+ years	
Nurses Dispute	0	(2)	**58**	(12)
Artists Protest	**17**	(6)	0	(12)
Zenith Investment	17	(6)	**25**	(12)
55% Pay Plan	**33**	(6)	27	(11)
Distribution Channel	**38**	(4)	17	(9)
Totals (unweighted) [b]	21	(24)	25	(56)

a Conflicts are included in the table if there were two or
    more forecasts for both levels of the variable
b Percentage figures in this row are unweighted averages
    of the percent correct forecasts reported for each
    conflict.

Similarly, there was little difference in the median experience of those who provided accurate forecasts (seven years) and those who provided inaccurate forecasts (six years). Accurate forecasts were provided by participants who were on average more experienced than those who provided inaccurate forecasts for three conflicts and the reverse was the case for the other two for which sufficient data were available for making a comparison (Table 32).

Table 32
**Game-theory experience of game-theorist forecasters
by accuracy of forecasts** [a]
Median years of experience (number)

	Inaccurate		Accurate	
55% Pay Plan	5.5	(12)	**6.0**	(5)
Nurses Dispute	5.0	(7)	**8.0**	(7)
Personal Grievance	**20.0**	(2)	7.0	(3)
Water Dispute	**10.0**	(2)	7.0	(4)
Zenith Investment	5.5	(14)	**8.0**	(4)
Totals (unweighted) [b]	5.5	(46)	7.0	(27)

a Conflicts are included in the table if there were two or more
    forecasts for both levels of the variable. Distribution Channel is
    excluded because two forecasters chose a combination
    decision that could not be appropriately coded for this analysis
b Experience figures in this row are medians of the medians
    reported for each conflict.

As with unaided-judgement participants, some game-theorist participants were very experienced. Five of the participants who provided forecasts claimed twenty or more years experience as game-theory experts. As with all expert participants, they were able to chose which conflicts to forecast (between them, they provided forecasts for all conflicts), but the game theorists were asked not to collaborate and none reported doing so. Taking their expertise at face value, it is reasonable to assume that they used the choices they were given wisely. It follows from this that a calculation of average forecast accuracy that included all their forecasts would bias any comparison in favour of these veterans' forecasts. In spite of this, the veteran experts' forecasts were less accurate than the overall *un*weighted-average for solo-expert game-theorist forecasts of 32 percent (Table 21). Of their 20 forecasts, only 30 percent were accurate on the basis of a weighted average.

Experts' experience – structured analogies

The forecasts of experts who used structured analogies and who claimed a high level of experience with similar conflicts were more accurate than those of experts claiming a low level of such experience for two out of six conflicts. They were less accurate for three of the conflicts, and there was no difference in accuracy for one. Overall, the forecasts from the two groups were similar in accuracy (Table 33).

The forecasts of experts with five or more years of experience as conflict management specialists were more accurate than those claiming a low level of such experience for five out of eight conflicts. They were less accurate for the other three conflicts. Overall, the forecasts of those with more experience were more accurate than those with less, but the difference is not significant at conventional levels (permutation test for paired replicates, one-tailed, $P = 0.16$)

Table 33
Accuracy of structured-analogies forecasts by experience [a]
Percent correct forecasts (number of forecasts)

	Experience with similar conflicts				Conflict management experience			
	Low (0-4)		High (5-10)		Under 5 years		5+ years	
Personal Grievance	20	(5)	**57**	(7)	38	(8)	**50**	(4)
Telco Takeover	**20**	(5)	0	(2)	**25**	(4)	0	(3)
Artists Protest	0	(3)	**50**	(2)	0	(2)	**33**	(3)
55% Pay Plan	**40**	(5)	33	(3)	0	(4)	**75**	(4)
Water Dispute	-	-	-	-	50	(2)	**100**	(2)
Zenith Investment	-	-	-	-	**50**	(4)	33	(3)
Distribution Channel	**58**	(6)	33	(3)	38	(4)	**60**	(5)
Nurses Dispute	75	(4)	75	(4)	**83**	(6)	50	(2)
Totals (unweighted) [b]	36	(28)	41	(21)	36	(35)	50	(25)

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable
b Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

There was no significant difference in the average levels of experience with similar conflicts or experience as a conflict management expert between those who provided accurate and those who provided inaccurate forecasts (Table 34).

Table 34
Forecaster characteristics associated with accurate and inaccurate structured-analogies forecasts [a]

	n	Experience with similar conflicts (0-10) [b]		Conflict management experience (years) [c]	
		Inaccurate	Accurate	Inaccurate	Accurate
55% Pay Plan	8	2.6	**3.7**	0.0	**15.0**
Nurses Dispute	12	3.0	**4.2**	**3.5**	0.0
Personal Grievance	8	4.7	**6.2**	0.0	**4.0**
Zenith Investment	7	**3.5**	1.7	**7.5**	0.0
Totals (unweighted) [d]	35	3.5	4.0	1.8	2.0

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable. Distribution Channel is excluded because some forecasters chose a combination decision that could not be conveniently coded for this analysis
b Averages
c Medians
d Forecaster characteristic figures in this row are averages of the figures reported for each conflict, except for "experience as c-m specialist" which are medians.

As with unaided-judgement and game-theorist participants, some structured-analogies participants were very experienced indeed. Seven of the experts who provided structured-analogies forecasts claimed twenty or more years of conflict management experience. As with all expert participants, they were able to chose which conflicts to forecast (between them, they provided forecasts for all conflicts) and, game theorists excepted, whether or not to collaborate (11 of their forecasts were collaborative). Taking their expertise at face value, it is reasonable to assume that they used the choices they were given wisely. It follows from this that a calculation of average forecast accuracy that included all their forecasts would bias any comparison in favour of these veterans' forecasts. In spite of this, the veteran experts' forecasts were no more accurate than the overall *un*weighted-average for solo-expert structured-analogies forecasts of 45 percent (Table 21). Of their 20 forecasts, 45 percent were accurate on the basis of a weighted average.

Experts' confidence

I asked experts using the methods of unaided judgement and structured analogies to rate, using an 11-point scale, the chances that they would change their forecasts given more time. I asked game theorists the confidence question only in relation to Personal Grievance, Telco Takeover, and Water Dispute. There were too few responses to conduct a useful analysis.

The experts' average responses for each conflict are shown in Table 35 – the raw responses were multiplied by 10 to give the percentage figures. Most of the figures are quite low. This indicates that, overall, the experts believed spending more time on the tasks would be unlikely to increase the accuracy of their forecasts. Unless the experts were choosing decisions at random and therefore extra time would not have made any difference, the low figures suggest that the experts were generally quite confident that their forecasts were accurate. I have assumed that the responses do represent the experts' confidence in their forecasts.

Table 35
Solo-experts' confidence in their forecasts
Percent chance of changing forecast given more time
(number of forecasts)

	Unaided judgement		Structured analogies	
Personal Grievance	30.0	(4)	15.8	(12)
55% Pay Plan	25.0	(12)	26.3	(8)
Zenith Investment	24.3	(14)	40.0	(7)
Water Dispute	24.0	(5)	22.5	(4)
Distribution Channel	21.7	(18)	33.3	(9)
Artists Protest	21.1	(19)	28.0	(5)
Nurses Dispute	20.8	(13)	28.8	(8)
Telco Takeover	18.9	(9)	14.3	(7)
Totals (unweighted) [a]	23.2	(94)	26.1	(60)

a   Percentage figures in this row are unweighted averages of the
    percent chance of changing forecast averages reported for
    each conflict.

Although the overall average levels of confidence for unaided-judgement and structured-analogies forecasts were similar, there is no obvious relationship if confidence levels for each of the conflicts are compared between the two methods.

The average of experts' confidence turned out to be a poor indicator of the forecast accuracy achieved for conflicts. There were negative correlations between confidence and forecast accuracy (Table 21) for the conflicts, for both unaided-judgement forecasts (Spearman rank-order correlation coefficient $r_s = -0.21$; equation 9.5, Siegel and Castellan, 1988) and structured-analogies forecasts ($r_s = -0.44$; equation 9.7, which includes adjustment for ties).

Nevertheless, this finding does not preclude the possibility that those experts who had high confidence in their forecasts for a particular conflict were more accurate than those who were less confident in their forecasts for the same conflict. In order to investigate this possibility, I re-coded the confidence data as "low confidence" for scores greater than two – probability that the forecast would be changed greater than 20 percent. Lower scores were coded as "high confidence" – probability that the forecast would be changed 20 percent or less (Table 36).

Table 36
Accuracy of experts' forecasts by forecaster confidence [a]
Percent correct forecasts (number of forecasts)

| | Unaided judgement | | Structured analogies | |
	Low [b]	High [c]	Low [b]	High [c]
Personal Grievance	- -	- -	0 (2)	**50** (10)
Telco Takeover	0 (3)	0 (6)	- -	- -
Artists Protest	**17** (6)	8 (13)	0 (3)	**50** (2)
55% Pay Plan	**20** (5)	14 (7)	33 (3)	**40** (5)
Distribution Channel	33 (6)	33 (12)	**60** (5)	38 (4)
Water Dispute	50 (2)	**67** (3)	50 (2)	**100** (2)
Zenith Investment	**43** (7)	14 (7)	40 (5)	**50** (2)
Nurses Dispute	**100** (4)	67 (9)	33 (3)	**100** (5)
Totals (unweighted) [d]	38 (33)	29 (57)	31 (23)	**61** (30)

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable
b Self-assessed likelihood of changing forecast if more time were available for the task is greater than 20% (between 3 and 10 on an 11 point scale)
c Self-assessed likelihood of changing forecast if more time were available for the task is 20% or lower (between 0 and 2 on an 11 point scale)
d Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

The forecasts of experts who used unaided judgement and who reported high confidence were more accurate than those of experts who reported low confidence for a single

conflict. Confident experts were less accurate for four and no different for two conflicts. Overall, the forecasts of experts using unaided judgement who were highly confident about their forecasts were markedly less accurate (at 29 percent) than were those from forecasters who were less confident (38 percent).

The forecasts of experts who used structured analogies and who reported high confidence were more accurate than those of experts who reported low confidence for six conflicts. They were less accurate for a single conflict. Overall, the forecasts of experts using structured analogies who were highly confident about their forecasts were substantially more accurate (at 61 percent) than were those from forecasters who were less confident (31 percent). The difference is statistically significant (permutation test for paired replicates, one-tailed, $P = 0.04$).

Those participants using unaided judgement who provided accurate forecasts were less confident, on average, than those who provided inaccurate forecasts. On the other hand, those using structured analogies who provided accurate forecasts were more confident than those who provided inaccurate forecasts (Table 37). I discuss why this might be in Chapter 5.

Table 37
Forecaster confidence associated with accurate and inaccurate forecasts [a]
Percent likelihood of changing forecast given more time [b] (number of forecasts)

	n	Unaided judgement		n	Structured analogies	
		Inaccurate	Accurate		Inaccurate	Accurate
Artists Protest	19	20	**30**	-	-	**-**
55% Pay Plan	12	25	25	8	26	**27**
Nurses Dispute	13	7	**25**	12	**65**	17
Personal Grievance	4	15	**45**	8	**20**	10
Water Dispute	5	**25**	23	**-**	**-**	-
Zenith Investment	14	23	**28**	7	**45**	33
Totals (unweighted) [c]	67	19	29	35	39	22

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable. Distribution Channel is excluded because some forecasters chose a combination decision that could not be appropriately coded for this analysis

b Average of self-assessed likelihood of changing forecast if more time were available for the task, from an 11-point scale with a score of zero indicating no chance of change (complete confidence), expressed as percentages (10=100%)

c Confidence figures in this row are averages of the figures reported for each conflict.

Analogies

Structured-analogies participants described the sources of their analogies and I coded them into one of five categories for each forecast. These source-of-knowledge categories were: (1) literature (fiction); (2) history (accounts and analysis of past events); (3) media (current events covered in newspapers, television, radio, and so on); (4) the experience of other people known personally to the participant; and (5) the participant's own experience. My intention was that the categories should reflect an increasing level of realism in the participants' knowledge of the analogies they used to derive their forecasts. Where participants provided more than one analogy for a conflict, I coded the source of their highest rated analogy to one of the five categories. Where participants provided more than one source for a single analogy (for example, history *and* the experience of people known personally to the participant) I coded the one that was closest to being the participant's own direct experience (the latter source in the example).

In the event, none were coded to literature, three were coded to history, and four were coded to the experience of others. As a consequence of the paucity of forecasts in these categories, I reduced the number of categories to two: "direct experience" (categories 4 and 5) and "media, history, fiction" (categories 1,2 and 3 – in effect, *in*direct experience).

The hypothesis that forecast accuracy increases with increasing realism suggests that forecasts based on analogous conflicts that were directly experienced by the forecaster will be more accurate than those that were not. This was the case for five out of six conflicts (Table 38). Overall, forecasts based on directly experienced analogous conflicts were accurate for 47 percent of forecasts, while those that were not were accurate for 36 percent of forecasts. The differences, although large, are not statistically significant at conventional levels (permutation test for paired replicates, one-tailed, $P = 0.16$).

Table 38
Accuracy of forecasts by source of analogy [a]
Percent correct forecasts (number of forecasts)

	Media, history, fiction [b]		Direct experience [c]	
Artists Protest	0	(2)	**33**	(3)
Telco Takeover	0	(3)	**25**	(4)
55% Pay Plan	33	(6)	**50**	(2)
Zenith Investment	33	(3)	**50**	(4)
Distribution Channel	50	(2)	**58**	(6)
Nurses Dispute	**100**	(2)	67	(6)
Totals (unweighted) [d]	36	(18)	47	(25)

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable
b Knowledge of analogous conflict obtained from media, history, or fiction
c Knowledge of analogous conflict obtained from participants' own experience or from that of people close to them
d Experience figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

Structured-analogies forecasts based on at least one highly rated analogy were, on average, no more accurate those based on lower rated analogies. Forecasts based on two or more analogies were more accurate overall than those based on a single analogy (Table 39). The difference was not, however, significant at conventional levels (permutation test for paired replicates, one-tailed, $P = 0.27$)

Table 39
Accuracy of forecasts by quality and by quantity of analogies [a]
Percent correct forecasts (number of forecasts)

| | Similarity to target [b] | | Number of analogies [c] | |
	Low (0-7)	High (8-10)	One only	Two or more
Artists Protest	0 (2)	**33** (3)	- -	- -
Telco Takeover	**20** (5)	0 (2)	0 (2)	**20** (5)
55% Pay Plan	33 (6)	**50** (2)	**40** (5)	33 (3)
Personal Grievance	29 (7)	**60** (5)	33 (6)	**50** (6)
Water Dispute	**100** (2)	50 (2)	**100** (2)	50 (2)
Zenith Investment	- -	- -	33 (3)	**50** (4)
Distribution Channel	- -	- -	40 (5)	**63** (4)
Nurses Dispute	**80** (5)	67 (3)	67 (6)	**100** (2)
Totals (unweighted) [d]	44 (27)	43 (17)	45 (29)	52 (26)

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable

b Participants' ratings of their analogies for similarity to the conflict being forecast on an 11-point scale. In cases where participants provided more than one analogy for a conflict, the top rating is used

c Number of analogies provided by participants

d Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

I tested the proposition that, if forecasts based on direct experience are more accurate than those that are not and if forecasts based on two or more analogies are more accurate than those based on a single analogy, forecasts based on two or more analogies from a forecasters' direct experience might be more accurate again (Table 40). The accuracy of forecasts based on two or more analogies from direct experience was high (69 percent) and the difference in accuracy between those forecasts and forecasts based on a single analogy from direct experience is large, but not statistically significant (permutation test for paired replicates, one-tailed, $P = 0.125$). There are too few forecasts based on analogies from media, history, or fiction to make useful comparisons.

Table 40
Forecast accuracy by source and quantity of analogies [a]
Percent correct forecasts (number of forecasts)

	Media, history, fiction [b]		Direct experience [c]	
	One analogy	2+ analogies	One analogy	2+ analogies
Personal Grievance	- -	- -	33 (6)	**60** (5)
Distribution Channel	- -	- -	50 (3)	**67** (3)
Nurses Dispute	- -	- -	50 (5)	**100** (2)
Zenith Investment	- -	- -	50 (2)	50 (2)
55% Pay Plan	**50** (4)	0 (2)	- -	- -
Water Dispute	**100** (2)	50 (2)	- -	- -
Totals (unweighted) [d]	75 (6)	25 (4)	46 (16)	69 (12)

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable
b Knowledge of analogous conflict obtained from media, history, or fiction
c Knowledge of analogous conflict obtained from participants' own experience or from that of people close to them
d Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

The data in the table highlight a potential weakness in the structured-analogies method. For some conflicts, for example 55% Pay Plan and Water Dispute, it may be difficult for forecasters to think of realistic analogies because similar conflicts have not occurred before or because there is a prevalent lack of knowledge of similar conflicts that have occurred.

Time taken

Although some expert participants spent as little as 10 minutes deriving a forecast for a single conflict and others spent two or more hours, the great majority spent more-or-less 30 minutes on the task. There was no difference in the median time spent on deriving accurate forecasts using unaided judgement or structured analogies than was spent on deriving inaccurate forecasts (Table 41). Neither was their any difference between the median time spent by game theorists deriving accurate forecasts and the median time they spent deriving inaccurate forecasts. Thirty minutes in both cases.

Table 41
Accuracy of experts' forecasts by time taken[a]
Median time in minutes

		Unaided judgement			**Structured analogies**	
	n	Inaccurate	Accurate	n	Inaccurate	Accurate
Artists Protest	20	**30**	21	-	-	-
55% Pay Plan	12	25	**43**	8	**30**	25
Nurses Dispute	14	15	**20**	8	30	30
Personal Grievance	4	23	**45**	12	30	**30**
Water Dispute	5	24	**30**	-	-	-
Zenith Investment	14	**25**	23	7	23	**60**
Totals (unweighted) [b]	79	25	27	35	30	30

a Conflicts are included in the table if there were two or more forecasts for both levels of the variable. Distribution Channel is excluded because some forecasters chose a combination decision that could not be appropriately coded for this analysis
b Times in this row are medians of the figures reported for each conflict.

Top forecasters

Two of the 18 game theorists who provided more than one forecast were accurate for half of their forecasts. The two provided forecasts for four and eight conflicts each. There were no game theorists who were more than 50 percent accurate. The two top game-theorist forecasters had between three and seven years game-theory experience when they made their forecasts. Their median experience of four years was less than the median of six years over all game-theorist participants' forecasts. These top game-theorist forecasters spent more time, on average, on their six accurate forecasts than they did on their six inaccurate forecasts.

Of the 17 solo-experts who provided more than one structured-analogies forecast, eight were 50 percent accurate or better – two were 67 percent accurate and one was 70 percent accurate. These top eight provided forecasts for between two and five conflicts each – 24 forecasts in all. Their forecasts were 57 percent accurate in an unweighted average across eight conflicts. Excluding Artists Protest and Water Dispute, for which there was only one forecast each, they were 59 percent correct on average.

The top eight had between zero and 25 years conflict management experience, a median of 2.8 years and an average of 5.7 years. This compares with a median of 2.7 years (average of 6.2 years) conflict management experience across all 17 who provided more than one forecast. The top eight's average self-rated experience with similar conflicts was 3.3 out of ten for accurate forecasts and 3.4 for inaccurate forecasts. These figures are the unweighted averages across the three conflicts for which at least two accurate and two inaccurate forecasts were provided by the top eight forecasters. These conflicts were Distribution Channel, 55% Pay Plan, and Personal Grievance. Table 34 shows the overall structured-analogies figures.

Three of the top eight reported spending less time on their accurate forecasts (on average) than on their inaccurate forecasts. Three spent the same time, and two reported spending more time on their accurate forecasts. The top eight spent between 10 and 60 minutes on deriving their accurate forecasts. Most spent 30 minutes. This was the median time spent on accurate forecasts by the top eight, and across all their forecasts. In contrast to the overall figures for experts using structured-analogies, the top eight forecasters were *less* confident in their accurate forecasts than in their inaccurate forecasts. The average assessment of the likelihood that they would change their forecasts given more time was 22 percent for accurate forecasts and 16 percent for inaccurate forecasts. The difference between the figures and the sizes of the samples are small, however, and the figures might best be interpreted as suggesting that top forecasters are generally more confident than others. As with the self-rated experience figures, these figures are unweighted averages across three matched conflicts with adequate data: Distribution Channel, 55% Pay Plan, and Personal Grievance. Table 37 shows the overall structured-analogies figures.

4.3   Appeal to managers

4.3.1   Selection criteria weights

Interpretation of the criteria

The Delphi panel I assembled rated the same 13 criteria for selecting a forecasting method as had been rated by individual forecasting experts in a study reported by Yokum and Armstrong (1995). The authors wrote of the descriptions of the criteria that they had used and I had adopted: "Although we believed that some of the phrases might be considered to be a bit ambiguous, respondents did not voice concerns about the definitions in either our pretest or our study" (p. 592). I did not find this to be the case with my panellists.

Some Delphi panellists had difficulty in interpreting six of the 13 criteria. One panellist wrote of *cost savings resulting from improved decisions*, "Not clear on this question: are we talking about improved methodology selection decisions or decisions consequent upon the outbreak of conflict?". The panellist did not provide a rating for that criterion. One panellist wrote of *flexibility*, "I don't know what this might mean" in the first round and "I think people interpreted this question very differently" in the second. The panellist did not provide a rating for that criterion. One panellist wrote of *ease of use*, "I am unclear what this refers to" in the first round. The panellist provided a rating for the criterion in the second round. Two panellists had difficulty with *ease of implementation*. One wrote "not sure if this is outcome or process" in the first round, but provided a rating in the second. The other wrote "Not sure how implementation of a method differs from use" in the first round, but provided a rating in the second. Two panellists had difficulty with *reliability of confidence intervals*. One wrote "guessing only" and provided a rating in the first round, and wrote "Maybe I missed the point first time" and provided a revised rating in the second. The other provided neither comment nor rating in either round. Finally, one panellist wrote of *development cost (computer, human resources)*, "If resources are available, not important; if not, then important" in the first round. The panellist provided a rating for the criterion in the second round.

In addition to the 13 criteria used by Yokum and Armstrong (1995), the Delphi panel rated three criteria that were suggested by Armstrong (2001c). One panellist had

difficulty in interpreting *ability to examine different environments*. The panellist wrote "I'm not sure enough that I know what this means" in the first round, and provided neither a comment nor a rating in the second.

All panellists' responses are shown in Appendix 16.

There are reasons to suppose that there was a genuine difference in the understanding of the criteria between Yokum and Armstrong's (1995) respondents and Delphi panellists. First, the Yokum and Armstrong respondents were professional or academic forecasters, whereas the panellists were experts in conflict management. Second, the forecasters were asked to rate the criteria for selecting a forecasting method for an unspecified purpose, whereas the Delphi panellists rated the criteria for the specific purpose of selecting a conflict forecasting method. Criteria that make sense in one context, may make less sense in another.

An alternative explanation is that the apparent differences in understanding between the two groups might have been the manifestation of differences in motivation. The Delphi panellists were asked to provide reasons for their ratings and to participate in a second round, and so were encouraged to test their understanding in ways that the respondents in the Yokum and Armstrong (1995) study were not. Further, as the panellists' names were known to me and were likely to be published, it seems reasonable to assume that they would have taken greater care to understand the material than the anonymous Yokum and Armstrong (1995) respondents (Sigall, Aronson, and van Hoose, 1970).

Criteria ratings

Rowe and Wright (2001) suggest using medians to aggregate Delphi panellists' responses in order to avoid the influence of extreme values. In the case of this research, values are limited to the range of the seven-point scale used to rate the criteria. Yokum and Armstrong (1995) use averages (means) of their respondents' ratings and so, for the sake of comparison, do I.

130

In common with Yokum and Armstrong's (1995) forecasters, the Delphi panel rated accuracy as the most important criterion for choosing a forecasting method (Table 42). The average of the forecasters' ratings was 6.2, and the panel's figure was 6.4. The two groups gave timeliness the same high average rating of 5.9. The average ratings for development costs and for theoretical relevance from the two groups were similarly low. Both forecasters and panellists rated all criteria as being relatively important: the average ratings of both groups are all greater than 4.0, the mid-point of the scale. Beyond these similarities the two sets of ratings have little in common.

Table 42
Importance ratings [a] of criteria for selecting a forecasting method[b]:
Yokum and Armstrong (1995) vs Delphi panel
Averages (number)

	Y&A [c]	Delphi panel	
1. Accuracy	6.2	**6.4**	(7)
2. Timeliness in providing forecasts	**5.9**	**5.9**	(7)
3. Cost savings resulting from improved decisions	**5.8**	4.5	(6)
4. Ease of interpretation	**5.7**	4.7	(7)
5. Flexibility	**5.6**	5.0	(6)
6. Ease in using available data	**5.5**	4.9	(7)
7. Ease of use	**5.5**	4.4	(7)
8. Ease of implementation	**5.4**	4.8	(6)
9. Ability to incorporate judgemental input	5.1	**6.4**	(7)
10. Reliability of confidence intervals	4.9	**5.8**	(6)
11. Development cost (computer, human resources)	4.9	**5.1**	(7)
12. Maintenance cost (data storage, modifications)	**4.7**	4.3	(7)
13. Theoretical relevance	4.4	**4.6**	(7)
14. Ability to compare alternative policies		5.7	(7)
15. Ability to examine alternative environments		5.5	(6)
16. Ability to learn (experience leads forecasters to improve procedures)		6.1	(7)

a On a scale of 1 = "unimportant" to 7 = "important"
b The Delphi panel rated the criteria for the purpose of selecting a conflict forecasting method
c 322 respondents.

Yokum and Armstrong (1995) expressed surprise that their forecasters, who rated the cost savings from improved decisions criterion the third-most-important at 5.8, did not rate it more highly. For the Delphi panel, it was the third-to-lowest rating criterion at 4.5. The forecasters also rated flexibility and ease (of interpretation, in using available data, of use, of implementation) relatively highly, whereas the panel did not. On the other hand, the panel rated the ability to incorporate judgemental input as highly as accuracy.

Overall, there is no significant correlation between the Yokum and Armstrong (1995) forecasters' ratings of forecasting method selection criteria and the Delphi panel's ratings of criteria for selecting a *conflict* forecasting method (Spearman rank-order correlation coefficient, $r_s = 0.28$, $P = 0.35$, two-tailed; Siegel and Castellan, 1988). Only the 13 criteria rated by both forecasters and panellists are used in this comparison. As well as the overall average of the forecasters' ratings, Yokum and Armstrong (1995) also provide the average ratings of four sub-groups. These are: decision makers, practitioners, educators, and researchers. The correlations between the ratings of these sub-groups are all high and statistically significant. There is no significant correlation between the Delphi panel's ratings and any one of these.

This is an interesting result. One reasonable interpretation, is that choosing a method for forecasting decisions in conflicts is markedly different from the task of choosing a forecasting method for some or all other purposes. In the absence of any evidence to the contrary, I will assume that the Delphi panel's ratings provide an accurate reflection of the importance of the individual criteria to managers who are faced with the task of choosing a conflict forecasting method.

4.3.2   Method ratings

Between three and five of the Delphi panellists rated each of the four conflict forecasting methods against the 16 forecasting method selection criteria. They used an 11-point scale for this purpose, where zero was defined as "inadequate" and 10 as "excellent".

Overall, the panel rated simulated interaction as the best forecasting method for conflicts ($R_m = 6.6$). Interestingly, the overall rating for unaided judgement ranks the method a close second ($R_m = 6.4$). The panel rated structured analogies third ($R_m = 6.0$), and game theory fourth ($R_m = 5.4$). The Delphi panel's average ratings are shown in Table 43. The selection criteria are presented in order of decreasing importance for selecting a conflict forecasting method. All ratings and comments are shown in Appendix 17.

Table 43
**Delphi panel's ratings [a] of conflict forecasting methods
by forecasting method selection criteria**
Averages

	Criteria Importance [b]	Method [c]				
		SI	UJ	SA	GT	n
A. Accuracy	6.4	**7.0**	2.8	3.4	2.8	5
B. Ability to incorporate judgemental input	6.4	7.4	**9.8**	8.0	4.0	5
C. Ability to learn [d]	6.1	7.2	**8.2**	7.2	7.2	5
D. Timeliness in providing forecasts	5.9	5.0	**7.8**	7.4	5.8	5
E. Reliability of confidence intervals	5.8	**7.0**	1.7	3.3	3.7	3
F. Ability to compare alternative policies	5.7	6.8	6.2	6.2	**8.2**	5
G. Ability to examine alternative environments	5.5	**7.0**	5.0	6.0	6.7	3
H. Development cost [e]	5.1	6.0	**7.6**	6.8	4.2	5
I. Flexibility	5.0	**8.2**	7.4	6.8	3.2	5
J. Ease in using available data	4.9	**7.5**	**7.5**	5.8	6.0	4
K. Ease of implementation	4.8	6.0	7.5	6.0	**7.8**	4
L. Ease of interpretation	4.7	6.6	**7.2**	5.8	5.4	5
M. Theoretical relevance	4.6	6.0	5.3	6.5	**8.0**	4
N. Cost savings from improved decisions	4.5	**5.4**	2.6	3.2	2.0	5
O. Ease of use	4.4	5.2	**8.2**	6.6	5.0	5
P. Maintenance cost [f]	4.3	6.0	**8.0**	7.0	7.0	3
Aggregate rating out of 10 (weighted by criterion importance), R_m		**6.6**	6.4	6.0	5.4	

a On a scale of 0 = "inadequate" to 10 = "excellent"
b On a scale of 1 = "unimportant" to 7 = "important"
c SI: simulated interaction; UJ: unaided judgement; SA: structured analogies; GT: game theory
d Experience leads forecasters to improve procedures
e Computer, human resources
f Data storage, modifications.

Simulated interaction received the highest average ratings for accuracy, reliability of confidence intervals, ability to examine alternative environments, flexibility, ease of using available data (first equal), and cost savings from improved decisions. It was rated relatively poorly for timeliness, ability to compare alternative policies (relative to game theory), ease of implementation, ease of use, and theoretical relevance.

The panel, appropriately, gave very low ratings to unaided judgement for accuracy and cost savings from improved decisions. Nevertheless, the method received highest

average rankings for more criteria than any other method: eight in total. The panel gave very high rankings to unaided judgement for some important criteria. In particular, ability to incorporate judgemental input, ability to learn, and timeliness.

The panel did not give the method of structured analogies highest average rankings for any of the 16 criteria. The method did, however, receive relatively high ratings for ability to incorporate judgemental input, timeliness, and development cost.

While the panel gave game theory the lowest overall ranking of the four methods, it was ranked highest for three criteria: ability to compare alternative policies, ease of implementation, and theoretical relevance. On the other hand, game theory ranked lowest, or lowest equal, for eight of the 16 criteria – including the three most important criteria.

4.3.3 Likely use of methods

The Delphi panellists were almost certain (94 percent likely) to use or recommend unaided judgement the next time they were faced with an important conflict forecasting problem. Perhaps this should not be surprising – it would be almost impossible to stop a decision maker from using unaided judgement when faced with a new conflict. Simulated interaction, the most accurate of the methods and the method with the panellists' highest aggregate rating, is likely to be used or recommended three times in four. Structured analogies is likely to be used or recommended more often than not, and game theory less than one time in three. Table 44 shows the average likelihood of the panellists using or recommending each of the four methods in future.

Table 44
**Likelihood[a] that Delphi panellists [b]
would use or recommend methods
for their next important conflict
forecasting problem**
Percent

Unaided judgement	**94**
Simulated interaction	75
Structured analogies	58
Game theory	28

a Average of Juster scale responses
   (Morwitz, 2001)
b Four panellists responded.

The form of the question that was put to the panellists is one that is useful for making predictions of behaviour (Morwitz, 2001). But is only useful for this purpose if the question is answered by respondents who are representative of the population whose behaviour one wishes to predict. The four Delphi panellists who responded to these questions were not such a group – they were chosen because of their expertise on the subject of conflicts, rather than because they were representative of managers making decisions on conflict forecasting methods. Further, after having participated on the panel, they were more informed on the subject of forecasting methods for decisions in conflicts than any manager is likely to be.

It seems reasonable to suppose that they would have been influenced by the research findings that I presented to them. Managers making decisions on conflict forecasting methods are unlikely to be exposed to the type of intensive and involving education programme the panellists were subject to. What is interesting, therefore, is that the panellists expect to use or recommend unaided judgement and game theory as much as they do, and simulated interaction as little.

5.   Discussion, conclusions, and implications

In this chapter, I present my conclusions (section 5.1), discuss the implications for researchers of limitations in the research (section 5.2), and describe the implications of my findings for managers (section 5.3).

5.1   Discussion and conclusions

In this thesis, I set out to evaluate forecasting methods in order to make useful recommendations to managers who face problems of forecasting decisions in conflicts. In order to do this, I sought published evidence on the relative accuracy of four methods, and made public requests and approached prominent researchers to the same end. The methods I considered were unaided judgement, game-theorist forecasts, structured analogies (a formal analysis of similar situations), and simulated interaction (a type of role playing).

The only pertinent evidence I was able to find was summarised in Armstrong (2001a). Armstrong presented evidence on the relative accuracy of forecasts from the unaided judgement of students and from the decisions made by role-playing students. I obtained new evidence on the relative accuracy of forecasts from experts who used unaided judgement, game theory, or structured analogies. I also obtained evidence from five new conflicts for these methods and from simulated interactions using student role-players.

In conducting my research I was guided by the methodology employed in the research reported in Armstrong (2001a) and by his principles for evaluating forecasting methods (Armstrong, 2001e).

In this section, I present my conclusions on the relative accuracy of the four forecasting methods, the generalisability of the findings, and the appeal of the methods to managers.

5.1.1   Relative accuracy

As Yokum and Armstrong (1995) showed, the expected accuracy of forecasts is the most important criterion for managers selecting a forecasting method. It is the principal criterion I used in comparing the four methods. I used the percentage of correct forecasts for that purpose. I tested the sensitivity of my findings to type of conflict, number of decision options, and use of rules versus judgement. I also tested the findings using alternative measures of forecast accuracy including probabilistic forecast accuracy scores, and forecast "usefulness". The findings that support my discussion and conclusions were presented in section 4.1.

Forecasting methods

For the purpose of forecasting decisions in conflicts, the method of simulated interaction is clearly the best available. Overall 62 percent of forecasts from simulated interactions by student role-players were accurate, a performance substantially better than that of other methods. (These results are summarised in Table 21 – reproduced below). Simulated-interaction forecasts were more accurate than those from other methods for all eight conflicts that I used in this research (there was one draw) and than chance. The finding is highly statistically significant. Moreover, for each of the conflicts, the most popular decision option from among the simulated-interaction forecasts was the same as (similar to, in the case of Artists Protest) the decision that was made in the actual conflict (Appendix 12, and Table 25 note b). Despite this impressive record, when simulated interaction is used in practice achieved accuracy could be depressed as a result of inherent variability in conflict situations, ineluctable simplification in describing them, variability in implementing the forecasting method, and difficulties in forecasting conflicts as they are unfolding.

The usual method for forecasting decisions in conflicts, unaided judgement by experts, was no more accurate than chance. Overall, unaided judgement was correct for 34 percent of forecasts – one would expect to be correct for 28 percent of forecasts by choosing decisions at random (chance). Whereas simulated interaction, as its name implies, involves simulating the interactions of parties to a conflict in a realistic way, unaided judgement involves thinking about a conflict. Armstrong (2001a) suggests that

simulated interaction offers superior forecast accuracy because it takes account of the influence of role on the way people behave and of the complexity that arises from the series of actions and reactions that typically occurs in conflicts.

Game theory has been proposed for predicting decisions in conflicts (subsection 1.2.2), but the findings of this research give no reason to believe that the method, as implemented in this research, would be useful for that purpose. Forecasts by game-theory experts, at 32 percent correct overall, were no more accurate than forecasts by experts using unaided judgement, nor were they significantly more accurate than chance.

Table 21 – reproduced
**Accuracy of solo-experts' forecasts,
and forecasts from simulated-interaction by novices** [a]
Percent correct forecasts (number of forecasts)

	Chance	Unaided judgement		Game theorist		Structured analogies		Simulated interaction	
Telco Takeover	25	0	(9)	0	(7)	14	(7)	**40**	(10)
Artists Protest	17	10	(20)	6	(18)	20	(5)	**29**	(14)[b]
55% Pay Plan	25	17	(12)	29	(17)	38	(8)	**60**	(10)[b]
Zenith Investment	33	29	(14)	22	(18)	43	(7)	**59**	(17)
Distribution Channel	33	33	(18)	23	(13)	50	(9)	**75**	(12)[b]
Personal Grievance	25	50	(4)	**60**	(5)	42	(12)	**60**	(10)
Water Dispute	33	60	(5)	67	(6)	75	(4)	**90**	(10)
Nurses Dispute	33	71	(14)	50	(14)	75	(8)	**82**	(22)
Totals (unweighted) [c]	28	34	(96)	32	(98)	45	(60)	**62**	(105**)**

a All forecasts are by individual experts except those from simulated interaction which, apart from four Nurses forecasts from groups of experts, are from groups of novices. Four probabilistic forecasts that could not be coded as single-decision forecasts are not included
b Forecast accuracy data reported in Armstrong (2001a)
c Percentage figures in this row are unweighted averages of the percent correct forecasts reported for each conflict.

While game theorists seek to transform conflicts into mathematical relationships for the purpose of analysis, experts who used the method of structured analogies based their forecasts on the outcomes of analogous real conflicts. Forecasts from structured analogies were, at 45 percent correct overall, significantly more accurate than those of game theorists.

The findings support the hypothesis that the more realistically a forecasting method models a conflict the more accurate will be that method's forecasts of decisions made in the conflict (subsection 1.3.2). As Raiffa's (1982) conflict taxonomy suggests (section 1

preamble), real conflicts tend to be complex. The assumptions and simplifications of game-theoretic modelling are at odds with that complexity (p. 14-15, p.31). Further, to the extent that game theory is inherently unrealistic in its approach, experts who base their thinking on game theory are not advantaged when forecasting real conflicts. Raiffa (1982) himself found this to be so: "Practically every case I looked at included an interactive, competitive decision component, but I was at a loss to know how to use my expertise as a game theorist" (p. 2). It is clear from Table 45 (from Table 1) that while the hypothesis on realism presents a plausible explanation for the findings from this research, its implications are not widely accepted: the findings are quite at odds with the expectations of the diverse experts who were surveyed on forecasting accuracy (subsection 1.2.5).

Table 45
Experts' expectations of forecasting methods' accuracy [a]
Percent correct (number of responses)

Method	Actual	Expectation [b]	Difference
Unaided judgement (by novices) [c]	27 (139)	30 (60)	3
Simulated interaction (using novices) [d]	61 (75)	40 (60)	-21
Unaided judgement (by experts)	32 (78)	50 (62)	18
Game theory (by experts)	26 (80)	50 (60)	24
Structured analogies (by experts)	45 (37)	50 (61)	5
Simulated interaction (using experts)		50 (57)	

a Forecasts for conflicts: Artists Protest, Distribution Channel, 55% Pay Plan, Nurses Dispute, and Zenith Investment. Experts forecasts are by solo experts.
b Median expectation for the five conflicts listed in note "a".
c Findings from Armstrong (2001a) for Artists Protest, Distribution Channel, and 55% Pay Plan except for 13 forecasts from Green (2002a): Artists Protest (1 correct / n=8); Distribution Channel (1/5). Findings for Nurses Dispute and Zenith Investment from Green (2002a).
d Findings from Armstrong (2001a) for Artists Protest, Distribution Channel, and 55% Pay Plan. Findings for Nurses Dispute and Zenith Investment from Green (2002a).

The experts dramatically *under*estimated the accuracy of forecasts from simulated interaction with novice role players, and dramatically *over*estimated the accuracy of experts' unaided-judgement forecasts and game-theorist forecasts. The accuracy of forecasts from simulated interaction using expert role players was not examined in this research.

The experts' expectations are consistent with the finding by Armstrong et al. (1987) that unaided judgement by experts was the most commonly used method for forecasting

decisions in conflicts. If managers and their advisors believe, as the expectations survey suggests, that there is no method that will consistently provide forecasts more accurate than can be obtained from the unaided judgement of experts, then they are unlikely to use alternative methods that are less familiar and can be more costly to implement.

There is good reason to believe the expectations of the experts in the survey are a fair representation of general expectations on forecast accuracy. The four groups of experts included in the survey were geographically dispersed and diverse in experience and interests. One group was in the UK, one was in the US, and two groups were in New Zealand. They were, respectively, academics, business executives, conflict management specialists, and police college educators. Remarkably, the mean and median expectations of the four groups were quite similar, despite the different nationalities and occupations of the experts.

Types of conflict

The data presented in Table 21 show that the conflicts used in this research varied considerably in their predictability and that this predictability was independent of the forecasting method used. That is, the ranking of the conflicts by the accuracy with which they were forecast hardly varies between the forecasting methods. Explaining this specific ranking is beyond the scope of this research as the sample of eight conflicts is too small to allow meaningful conclusions about which conflict characteristics affect predictability and hence forecast accuracy. Nevertheless, the consistency of the rankings is statistically highly significant and this suggests that conflicts in general are likely to have qualities that make them more, or less, predictable than other conflicts.

One salient characteristic associated with the conflicts, but not intrinsic to them, is the number of decision options provided. That number is largely discretionary, and should be strongly influenced by the interests of the client or manager who has commissioned the forecast. For example, a client for a forecast of the 55% Pay Plan decision may have been interested to know only whether there would or would not be a strike (two decision options). On the other hand, the client may have been interested to know how long a strike to expect, should one occur (the research is based on four options for this conflict). An inspection of Table 21 reveals that this characteristic cannot be the only explanation

of the ranking of the conflicts by predictability – predictability does not increase monotonically with chance and there is considerable variation in accuracy for any combination of chance and forecasting method. The latter point is shown most clearly in Table 23.

I tested the possibility that, in the absence of a formal model, it might be possible to judge the predictability of conflicts in advance of forecasting them. This appears not to be possible. The ratings I obtained for the *a priori* predictability of the eight conflicts varied to such an extent between the raters as to be indistinguishable from a chance allocation (subsection 4.1.1 and Appendix 14). On reflection, this is not surprising as the finding is consistent with that on the accuracy of unaided-judgement forecasts: the experts were no better than chance (Table 21). In sum, judgement appears inadequate for the task of assessing predictability, and a formal model of predictability would require a great deal more data than is available from this research.

Number of decision options

Logically, one would expect an inverse relationship between the number of decision options from which forecasters must choose and the accuracy of their forecasts – the more options the lower, *ceteris paribus*, the chance of selecting the correct one. This suggests that conflict forecasting problems should be framed with as few decision options as are consistent with the requirements of the clients of the forecasts.

Seven of the eight conflicts used in this research were each presented with either three or four decision options. On average, one decision option from three can be taken to represent one-third of the decision space (of all plausible decisions) for a conflict and one from four to represent one-quarter of that space. Consequently, a decision from among three options is likely to include an extra one-twelfth of the decision space compared to a decision from among four options ($\frac{1}{3} - \frac{1}{4}$ equals one-twelfth). On this basis, it would be reasonable to expect the error rate for forecasts of a conflict which was presented with four decision options to have been reduced by one-ninth had three options been presented instead (one-twelfth / three-quarters equals one-ninth).

The analysis presented in subsection 4.1.1, Table 23 in particular, showed that the error rates for three-option conflicts were much lower than expected on this basis, relative to four-option conflicts. Or, to put it another way, increasing the number of decision options from three to four reduced accuracy more than proportionately for all methods. For example, the average accuracy of simulated-interaction forecasts of conflicts with three decision options was 76 percent correct. By contrast, for conflicts with four decision options the average accuracy was 53 percent correct. Table 46 shows the unexplained component of the error reduction between four- and three-option conflicts.

Table 46
Unexplained relationship between number of decision options and error rates [a]
Average percent *in*correct forecasts [b]

	Four-option conflicts [c] (A)	Three-option conflicts [d] (B)	Actual error reduction (C=A-B)	Expected error reduction [e] (D)	Unexplained error reduction (C-D)
Chance	*75*	*67*	*8*	*8*	*0*
Unaided judgement	78	52	26	9	**17**
Game theorist	70	58	13	8	**5**
Structured analogies	69	39	29	8	**22**
Simulated interaction	47	24	23	5	**18**
Unweighted averages [f]	66	43	23	7	16

a Based on Table 21 data
b Unweighted averages of percent *in*correct forecasts for conflicts (complements of Table 21 figures)
c 55% Pay Plan, Personal Grievance, and Telco Takeover (n = 3)
d Distribution Channel, Nurses Dispute, Water Dispute, and Zenith Investment (n = 4)
e Equal to one-ninth of the four-option error rate
f Excluding chance.

The unexplained error reductions are positive for all of the four forecasting methods, and large for three of them. Given the finding that the relative predictability of the conflicts is independent of the forecasting method employed (subsection 4.1.1), however, it is to be expected that what is true for one of the methods will be true for all. One possible reason for the surprisingly large differences in the error rates is that, by chance, the four-option conflicts used in this research were intrinsically more difficult to forecast than the three-option conflicts.

A second is that the order in which the decision options were presented biased forecasters against choosing the actual decisions for the four-option conflicts relative to

those with three. Such a bias might be general or specific to the number of options, or some combination of these. A general bias might be a tendency to avoid first and last options. The effect of such a bias would be greater in cases where fewer options were presented. Taken to the extreme, if first and last options were avoided altogether, forecasts for a conflict with three decision options and the actual decision as the middle option would always be correct. On the other hand, a tendency to avoid first and last options might be specific to sets of four or more options – there being no such bias in choices among three options – for example. There is insufficient data to draw conclusions on the presence or absence of any such biases.

Decision options for any given conflict can, and in the case of this research did, differ in their specificity. For example, "Two ACMA plants [were purchased]" (Zenith Investment) is more specific than "A compromise was reached" (Nurses Dispute). This suggests a third possible reason for the large differences in error rates: forecasters' selections were biased away from the actual decision in the four-option conflicts relative to those with three due to variations in specificity. Such a bias might occur because options that are more specific appear to forecasters to be more representative of actual decisions. It seems reasonable to assume that the extent to which forecasters perceive an option to be representative or otherwise will be at least partly a function of the way the option is described. On the other hand, a specific option could be seen by forecasters as being less likely than a more inclusive option. The latter is an argument that an option that is more specific than others in a set of options represents a disproportionately small share of the decision space and is, therefore, less likely to occur.

With so many possibilities, resolving the question of whether the unexplained error differences in this research arose as a consequence of inherent differences in the conflicts' predictability or as a consequence of biases in the formulation or presentation of the decision options would require much more research. A very large research programme indeed would be needed to identify grounded principles for formulating and presenting decision options in such ways that forecaster biases would be avoided. Without the findings from such a programme, the possibility exists that there are biases that strongly favour the accuracy of forecasts from among three, as opposed to four, options. As the available evidence is consistent with this supposition, it seems sensible to present no more than three decision options to forecasters if this is at all possible.

Rules

With the method of structured analogies, using rules to transform analogies information into a forecast is likely to increase accuracy relative to using forecaster judgement to do so. An example of such a rule is, if an expert rates one analogy as more similar to a target conflict than any other, choose the decision option suggested by that analogy as the forecast (Figure 1). This conclusion on rules is consistent with evidence in the literature that accuracy is improved by adopting formal processes (subsection 1.2.3).

I had assumed that participants assigned to use structured analogies would choose the decision option that was suggested by the weight of their own analogical evidence. In the case of some forecasts, however, participants preferred their judgement ahead of the evidence of their own analogies. The findings presented in Table 21 suggest that this preference was likely to have been wrong-headed – the method of structured analogies was more accurate than unaided judgement – and so it was. With collaborating experts' forecasts included, eliminating judgement from the process of choosing a forecast (using the rules shown in Figure 1) increases the accuracy of structured-analogies forecasts from 40 percent to 46 percent.

Using a rule to transform analogies information into probabilistic forecasts (Appendix 13 and Table 48) results in greater accuracy than using forecasters' own probabilities. Using a simpler rule – adopt as forecasts the first-choices implied by forecasters' analogies – appears to offer the greatest accuracy (Table 22). The relative accuracy of probabilistic forecasts is discussed below.

In sum, rules are superior to unaided judgement for deriving forecasts from analogical information and should, therefore, be preferred.

Probabilistic forecasts

By presenting forecasts in the form of probabilities of possible events occurring and assigning non-zero probabilities to all options, forecasters can avoid the stigma of inaccurate forecasts. When an event occurs that had been assigned a low-probability by a

forecaster, sceptical forecast users might question whether such a forecast was truly accurate. Forecasters can respond that their probabilities accurately represented the *a priori* likelihood of the various events occurring. This assertion cannot be proved or disproved in regard to any single situation, as each situation that is forecast will in some way be unique. This research is concerned with forecasting methods, however, and it is possible to rate the overall accuracy of probabilistic forecasts from different methods by applying them to situations belonging to a class of forecasting problem – in this case, decisions in conflicts.

Brier scores (Formula 2) have been recommended for the purpose of rating the accuracy of probabilistic forecasts but I found that another measure, the probabilistic forecast accuracy rating or *PFAR* (Formula 3), had better characteristics (subsection 4.1.1 and Appendix 13). Regardless of which of these two measures is used, on balance probabilistic forecasts appear to offer no gain in accuracy compared to a policy of adopting the decision option with the highest probability (first-choice) as the forecast (Table 47). This is interesting as, while allocating a probability of 1.0 to an option that occurs will achieve a better score than any alternative, in cases where an option allocated a probability of 1.0 does *not* occur any alternative allocation would achieve a better score.

Research participants provided relatively few explicitly probabilistic forecasts. I have argued that, as they were offered the option of doing so, expert participants using unaided judgement or structured analogies effectively provided implicit probabilistic forecasts when they ticked one box. When both implicit and explicit probabilistic forecasts are included in analysis of structured analogies, first-choice forecasts resulted in greater accuracy overall relative to using probabilities (Table 22). In the case of unaided judgement, using probabilities was slightly more accurate than first-choices. In neither case, however, were the differences statistically significant. Finally, analysis using probabilistic forecasts makes no difference to conclusions about the relative accuracy of forecasts from the four methods.

It is possible that, had I insisted on participants providing probabilities for each option, the probabilities provided would have been different from the 0.0 or 1.0 figures I have assumed were implied by a tick in one box. Further research would be needed to test the assumption. Nevertheless, the analysis of the explicit probabilistic forecasts alone

suggests that if there is any accuracy advantage to probabilistic forecasts, it is likely to be small.

Usefulness

Forecasts from simulated interaction were judged by five independent experts to be more useful, on average, than forecasts from other methods (subsection 4.1.2; Table 25). The experts were asked to rate the usefulness of the decision options that might have been forecast in the light of the actual decisions that were made. This approach to comparing the methods has face validity, as the experts' ratings could reasonably be expected to represent the extent to which managers would have taken appropriate action in response to a particular forecast. Further, the level of agreement between the raters, while well-short of being complete, was statistically highly significant. In a sense, the absence of complete agreement is reassuring: it suggests that diverse perspectives or interpretations or both were incorporated into the overall usefulness ratings. That the raters interpreted the text in diverse ways is supported by the comments that they made in support of their ratings (subsection 4.1.2).

Comparison of the forecasting methods on the basis of the usefulness to managers of their forecasts led to the same conclusions about the relative merits of the methods as did a comparison of percent correct forecasts. This finding is remarkable when the apparent diversity of rater perspective and interpretation is considered. The consistency of the findings across the different measures of forecast accuracy, broadly defined, provides considerable support for the hypothesis on realism (subsection 1.3.2).

5.1.2   Generalisability

With this research, I set out to provide useful advice to managers. Advice that can be applied in a wide range of circumstances is likely to be more useful than advice that has limited applicability. For example, I included diverse conflicts in the research in order to maximise the chances that managers would find at least one conflict that was relevant to their concerns. In this subsection, I discuss the extent to which the principal conclusions on forecasting accuracy can be generalised in the face of variations in forecaster

collaboration and forecaster expertise. Discussion on generalisation across expertise is extended with an examination of the effect on forecasting accuracy of variations in two indicators of expertise: the quality and quantity of forecasters' analogies, and forecasters' confidence in their forecasts. The findings which support the discussion and conclusions were presented in section 4.2.

Collaboration

Collaboration by forecasters might be expected to improve forecast accuracy (subsection 1.3.3). However, while collaborative forecasts were, on average, somewhat more accurate than solo forecasts the average differences were small and, for some conflicts, differences were opposite to that expectation (Table 26). The differences in accuracy between joint and solo forecasts from unaided judgement and structured analogies were not statistically significant.

In addition, there were differences between the characteristics of forecasters who provided joint forecasts and those who provided solo forecasts. On average, joint forecasters spent more time deriving their forecasts, had more years of conflict management experience, and had more experience with similar conflicts than solo forecasters (Table 27). On the face of it, the small differences in average accuracy might be explained by the greater time and expertise that was applied to joint forecasts. This contention is not supported by the evidence of this research, however, and so the best explanation remains that implied by the lack of statistical significance: the differences arose by chance.

Forecasters were free to decide whether or not they would collaborate with others in making their forecasts, and relatively few did so. This discretion over collaboration may also explain the differences between the measured characteristics of the forecasters providing solo and joint forecasts. More collaborative forecasts for all conflicts would provide greater certainty over whether there is or is not any gain in accuracy to be had from collaboration. It would be sensible to systematically allocate participants to one or other of the treatments to the same end. Nevertheless, the existing data suggest that collaboration is unlikely to provide large increases in accuracy when forecasting decisions in conflicts, and so this matter should not be a priority for further research.

Expertise

Surprisingly, the expertise of forecasters appears to have no material influence on the accuracy of their forecasts of decisions in conflict. Experts' unaided-judgement forecasts were little more accurate than those of novices (Table 28). Among experts who used unaided judgement or structured analogies, neither more years of experience with conflicts nor more experience with conflicts similar to the target conflict translated into forecasts that were more accurate (Tables 29 and 33). More years of experience with game theory was also not associated with forecasts that were more accurate (Table 31).

One might expect that the forecasts by the most experienced experts would be substantially more accurate than those of others. After all, it is such people who are sought out by government, business, and the media for predictions on how conflicts (wars, civil disturbances, takeover battles, strikes, and so on) will turn out. Surprisingly, then, for the methods of unaided judgement, game theory, and structured analogies, the forecasts of experts with even 20 or more years of experience were no more accurate than the average.

The most accurate forecasters cannot be identified in advance; at least not on the basis of experience with similar conflicts, experience as a conflict management expert, time spent forecasting, or confidence. These characteristics varied widely across the more accurate of the game-theory and structured-analogies participants. Not only are the best forecasters not identifiable in advance, even the best were not outstandingly accurate. No game-theory expert's forecasts were more than 50 percent accurate and only one structured-analogies participant was 70 percent accurate.

Although it is tempting to speculate that forecasters' accuracy records may be good predictors of their accuracy in future, the evidence is against this. Armstrong (1980) sought evidence for the existence of consistently accurate judgemental forecasters and found none. In the light of this, and the absence of any distinguishing characteristics in the top forecasters in this study, it seems likely that the accuracy of the top forecasters' forecasts will tend to regress to the mean. That is, there is no reason, other than short records of no more than eight and typically fewer forecasts, to expect that future forecasts from top forecasters would be any more or less accurate than the average for all forecasters using the same method. The most likely explanation of the top

forecasters' exceptional (relative to other experts) accuracy is chance and, by chance, their average accuracy would be likely to decline were they to make further predictions.

The findings from this research on the value of expertise are consistent with those of Armstrong's (1991) study comparing the accuracy of forecasts by experts and novices. Armstrong found that high school students were *more* accurate than marketing academics at predicting the outcomes of published consumer behaviour experiments. The findings are also consistent with Tetlock (1992). In his study of experts' forecasts of real conflict outcomes, Tetlock did not compare their accuracy with that of novices, but he did find that the experts were no better than could be expected from chance.

Another indicator of the quantum of expertise applied to a forecasting problem is the time spent deriving a forecast. Game-theory experts and forecasters using the methods of unaided judgement or structured analogies can and do derive accurate forecasts of decisions in conflicts in 30 minutes. This is no more nor less than the time typically taken to derive *in*accurate forecasts, of which there are typically more from these methods (Table 41). In sum, there is no reason to believe that if forecasters who used these methods spent more time in deriving a forecast that this would result in forecasts that were more accurate.

It is clear from Table 41, however, that data are sparse or absent for some combinations of accuracy, conflict, and forecasting method. More data would, therefore, allow greater confidence in any conclusion on the effect, on accuracy, of the time experts spend deriving forecasts. Nevertheless, it is questionable whether further research on this matter would be useful. In practice, for important conflict forecasting problems, it would be just as well to give experts an hour to consider the situation and derive a prediction. One hour was the maximum median time taken to accurately forecast any conflict using structured analogies. Given the findings of this research, it would be reasonable to be sceptical of the value of paying experts for more time than this – better to pay for more experts for an hour each that the same expert for more hours.

Unlike the measures discussed above, two indicators of expertise do appear to be associated with meaningful variations in forecast accuracy. They are (1) the quality and quantity of analogies identified by structured-analogies forecasters and (2) their confidence in their forecasts. Although, as the preceding discussion suggests, accurate

forecasters cannot be identified in advance, these indicators can be measured at the time forecasters make their forecasts. They are discussed below.

Quality and quantity of analogies

It appears that structured-analogies forecasts based on two or more analogies from a forecaster's own experience will tend to be more accurate than forecasts based on fewer analogies or analogies that are not part of the forecaster's own experience (Tables 38 - 40). While this conclusion is not based on findings that are statistically significant at conventional levels (at best $P = 0.125$ for two-plus analogies from direct experience versus one analogy from direct experience, and $P = 0.16$ for analogies from direct experience versus others) the differences are large.

The differences are also consistent with the hypothesis on realism. That is, first, analogous conflicts from personal experience can and often will be remembered more vividly, accurately, and comprehensively by forecasters than will other conflicts. Second, the more analogous conflicts that forecasters can identify, the more likely it is that the target conflict will be represented accurately. Conversely, the fewer analogous conflicts that forecasters can identify, the more likely it is that the target conflict is atypical and therefore poorly represented by analogies.

The combination of large differences that are consistent with the central hypothesis of this research and *P*-values that indicate the differences are quite unlikely to have arisen by chance, suggests that further research on the effect of number and source of analogies on structured-analogies forecast accuracy is likely to be worthwhile.

Experts' confidence

Research participants were asked to assess how likely it was that they would have changed their forecasts had they more time available to derive them. On average, the participants thought it quite *un*likely that they would change their forecasts (Table 35). This finding supports the earlier contention that one hour of forecaster time per conflict should suffice. It also supports the principal findings of this research on the relative

accuracy of forecasts from the four methods in that the finding furnishes further evidence that the forecasts of expert forecasters were not disadvantaged as a consequence of the experts spending insufficient time deriving their forecasts.

In the case of structured analogies, the data show that forecasts from forecasters who were most confident were correct more often than forecasts from forecasters who were less confident. In the case of unaided judgement, the reverse was true (Table 36). In both cases, there was no positive relationship between *average* forecaster confidence for individual conflicts and the relative accuracy with which the experts were able to predict decisions in those conflicts, as a comparison of Table 35 with Table 21 shows. This implies that structured-analogies forecaster confidence is a valid indicator of the relative likelihood of forecasts being correct for a particular conflict, but that nothing useful can be said about the relative likelihood of forecasts being correct across conflicts.

The conclusions on unaided-judgement forecaster confidence are consistent with the finding that experts' unaided-judgement forecasts are not significantly better than chance. In other words if, on average, experts cannot accurately predict decisions in conflicts there is little reason to believe that their confidence in their own accuracy for specific conflicts would be accurate. The same mechanism is likely to be at work in both cases: the pervasiveness of delayed and ambiguous feedback in conflict forecasting situations. If this were not the case, it is arguable that experts would either avoid forecasting decisions in conflicts using their unaided judgement as a failed cause or they would become much better at it.

It seems plausible that the superior calibration of forecasters using structured analogies, relative to that of forecasters using their unaided judgement, is a result of formal consideration of alternatives to their own judgmental assessments. This accords with the recommendation by Arkes (2001) to consider alternatives, especially when forecasting in new situations. The conclusion is also consistent with the hypothesis on realism in that the alternatives being considered are typically real conflicts and outcomes.

5.1.3   Appeal to managers

On the basis of the evidence from this research, selecting a conflict forecasting method solely on the likely accuracy of forecasts is clear-cut. When other criteria that are important to managers are also considered, the decision is less obvious. Using the Delphi panel approach described in section 3.5, I elicited ratings for conflict forecasting method selection criteria (subsection 4.3.1) and for the four methods against those criteria (subsection 4.3.2) from seven conflict management specialists. The panel rated simulated interaction at 6.6 out of 10, little better than unaided judgement (6.4). The rating for structured analogies was 6.0 and for game theory 5.4.

Despite the ratings, unaided judgement is the method most likely to be used for an important conflict forecasting problem, even by practitioners who are familiar with the findings of this research (subsection 4.3.3). While such well-informed practitioners assess the chances of using simulated interaction for an important conflict as 75 percent on average, it may be doubted whether, in the event, usage would be as high as this.

Accuracy aside, the importance attributed to criteria for the purpose of selecting a conflict forecasting method is mostly quite different from the importance attributed to the same criteria for selecting a forecasting method for an unspecified purpose (Table 42). For example, it seems cost saving from improved decisions is not an important criterion for selecting a conflict forecasting method. Given that some of the conflict management specialists had difficulty interpreting the criteria, the question of how managers select forecasting methods for conflicts warrants more further research.

5.2   Implications for researchers

Forecasting decisions in conflicts has been subject to remarkably little research effort given the frequency and cost of inaccurate forecasts. I have drawn on the commentaries of Armstrong (2002), Bolton (2002), Erev et al. (2002), Goodwin (2002), Shefrin (2002), and Wright (2002), as well as on the findings of this research, in order to identify limitations in the research that, once overcome, appear most likely to lead to further useful findings on conflict forecasting methods. In this section, I note those limitations and propose directions for further research.

Limitations of this research

The principal limitation of this research is the lack of sufficient data to draw firm conclusions on the conditions that favour particular methods and the effect of the number of decision options on forecast accuracy. Resolving these questions would require many more forecasts. I expended considerable effort in obtaining the forecasts that I did. In order to make substantial progress in this field, a large budget to pay expert participants for their time may well be needed.

Further limitations are, first, the absence of comparable research by researchers with rival theories on the relative merits of different conflict forecasting methods. Publicity for this research may encourage replications. Second, because the game theorists themselves decided how to go about making their predictions, and because their brief descriptions of how they did so generally made no mention of model building, it is impossible to say with certainty that formal modelling would not have led to greater accuracy. Third, two only of the conflict descriptions used in this research were written without any knowledge of the final outcomes. This is because preparing material on conflicts that are in progress and obtaining sufficient forecasts before an outcome is known tends to be more expensive than preparing material on conflicts from the past. Finally, for the sake of comparability, I used established general-purpose forecasting method selection criteria to assess the appeal of conflict forecasting methods. Devising a set of criteria that are specific to forecasting for conflicts may be both feasible and useful.

These limitations in the research are addressed in the suggestions for further research which follow.

Suggestions for further research

▲ **Determine the conditions that favour different methods**

Forecasts for many more conflicts will be needed to determine which conditions, if any, favour particular forecasting methods.

Research to date has found that simulated interaction consistently provides the most accurate forecasts of decisions in conflicts (Table 21). Nevertheless, the accuracy of forecasts varies with the conflict being forecast and it would be useful for managers to know how likely it is that forecasts for a particular conflict will be accurate.

I have presented findings on structured analogies that suggest forecasts from the method may be similarly accurate to those from simulated interaction if certain conditions are met (Tables 38 to 40). More forecasts for the eight conflicts used in this research, and for new conflicts, are needed in order to increase confidence in this finding.

There is no evidence to suggest that either unaided-judgement or game-theorists' forecasts are likely to be more accurate than simulated-interaction or structured-analogies forecasts (Table 21). Consequently, it would not be a sensible expenditure of limited research resources to pursue research on unaided judgement or game theory for forecasting conflicts, other than to replicate the research presented here. Subject to the caveat that the influence of formal modelling on accuracy remains uncertain, this recommendation particularly applies to game theory, which has been the subject of enormous research effort to date.

▲ Pursue replications by researchers with contending theories

Researcher bias may affect research findings in ways that are unclear. For this reason, properly conducted replications by researchers with contending theories would provide important information for researchers and managers.

In order to facilitate replication and commentary, I have published a description of part of the research reported here (Green, 2002a) and have made supporting material available on the internet (kestencgreen.com). Also, I presented papers at the 2001, 2002, and 2003 International Symposia on Forecasting. I expect to publish more of the findings, and to make more material available on the internet (at forecastingprinciples.com) in the future.

▲ Obtain forecasts from game theory models

In the case of structured analogies, I found that using a rule to derive forecasts from experts' analogies resulted in greater accuracy than was achieved by accepting experts' forecasts, whether or not they were consistent with their own analogies (p. 97). Were the game-theorist participants in this research also inclined to ignore the implications of their own knowledge to the detriment of forecast accuracy? This question might be answered by asking game theorists to develop game-theoretic models of the conflicts used in this research, and to record the decisions predicted *by the models*. Earlier, however, I argued and presented evidence (p. 140) that game theory's representation of conflicts is inherently unrealistic. If that is so, forecasts that are certain to have come from game-theoretic models are unlikely to be more accurate than the game-theorist forecasts presented in this research.

▲ Forecast conflicts that are unresolved at the time of forecasting

Findings from research based on unresolved conflicts are likely to be more convincing to managers than the alternative. Forecasts of decisions in newsworthy unresolved conflicts may be newsworthy themselves and media coverage is likely to help persuade managers of the usefulness of the research findings.

It can be difficult and will often be impossible for researchers not to be influenced by knowledge of the actual outcome of a conflict when they prepare material for research. This problem may be exacerbated if the researcher's knowledge is based on accounts that were written after the conflict occurred. Material that is prepared and used while a conflict is unfolding will provide a better test of forecasting methods than material prepared afterwards, as the test will more closely match the conditions in which conflict forecasting methods must be used.

The material for two of the conflicts used in this research, 55% Pay Plan and Nurses Dispute, was prepared before the decisions that were the subject of the forecasts were made. More tests of this kind would provide useful information on how well conflict forecasting methods perform in conditions of actual use and how procedures can be improved in order to accommodate the requirements of live forecasting.

▲ Examine the effect of the number of decision options

I found that forecasts of conflicts with three decision options were markedly more accurate than those with four options (subsection 4.1.1, Table 23). The finding is important if it represents a regularity, rather than an artefact of the conflicts that I included in my research.

In subsection 5.1.1 I identified possible explanations for the finding. Namely: (1) intrinsic differences in the predictability of the conflicts; (2) forecaster bias for or against options based on the position of the option in a list that is (a) general (independent of the number of options presented), or is (b) specific to the number of options; (3) forecaster bias for or against options based on their relative specificity that is (a) related to the representativeness of options, or is (b) positively related to the option's share of the decision space.

Explanation 1 could be tested by varying the number of decision options for the conflicts used in this research, and obtaining new forecasts to compare with those already obtained. This approach would not provide a definitive result, however, as the findings might be confounded by the factors involved in explanations 2b, 3a, and 3b. In order to overcome this difficulty, it would be necessary to undertake a research programme that addressed all of the possible explanations.

Such a programme would involve obtaining forecasts for a set of conflicts for which it was reasonable to provide, as a minimum test, both three and four decision options. The order of the decision options would need to be varied, as would the specificity – both in terms of share of the decision space and the way in which the decisions were described.

While a programme to address this issue would need to be very large, the size of unexplained difference between the accuracy of forecasts of conflicts presented in this research with three- or four-options suggests that findings from further research may be very useful.

▲ **Investigate the effect of forecasting in stages**

If further research finds that providing forecasters with three decision options tends to result in forecasts that are much more accurate than if four or more are provided, it would be useful to conduct research on forecasting in stages. Whether or not forecasts are *much* more accurate when only three decision options are provided, providing fewer options will more likely result in accurate forecasts than providing more options. Forecasting in stages may allow high levels of forecast accuracy to be maintained while still addressing managers' desires to distinguish between more than three decision options (Table 2, principles C3 and C4). For example, my client for research on the Personal Grievance conflict wanted to distinguish between 11 decision options. Useful forecasts were not obtained in tests of the material and four decision options were used in the research as a compromise solution (subsection 3.3.1).

To illustrate further, Artists Protest was provided with more decision options than any other conflict, and forecasts for that conflict were among the least accurate (Table 21). It would be possible to obtain forecasts for three, rather than six, decision options at a time without "losing" any of the options, as follows. The current first four options (A-D) could be collapsed into one: "The government will make concessions to the protesters". Role players who made that decision could be given another set of three decisions (A, B *or* C, and D) and be asked to continue with their simulation until they had decided between them.

▲ **Investigate the effect of variations in descriptions**

It seems likely that the accuracy of forecasts will be influenced by characteristics of the conflict descriptions given to forecasters. For example, the conflict descriptions I used in my research were mostly one-page long. An editor who did not know the outcome of the conflict could be asked to prepare versions of the description which had elements, such as descriptions of individuals or historical context, excised. Shorter and simpler descriptions may lead to more forecasts that were accurate, and changes in accuracy may vary across methods. Identifying critical elements of conflict descriptions may allow descriptions to be prepared faster and at lower cost than otherwise.

Research on the effect of variations in conflict descriptions would "test assumptions for construct validity" (principle B2, Table 2 and Appendix 1). This principle of forecasting method evaluation was not addressed in this research.

▲ Devise a set of forecasting method selection criteria for conflicts

I asked a panel of conflict management practitioners to rate the importance of a set of established general-purpose forecasting method selection criteria (subsection 1.3.4, and section 4.3). While I expected their ratings to ensure that only criteria they considered relevant to forecasting for conflicts carried any weight, the difficulty some of the panel had in interpreting some of the criteria suggests that might not have occurred. Depth interviews with forecasting practitioners and experts could be used to compile descriptions of the criteria that were less ambiguous to conflict management practitioners. These descriptions could be used in interviews with conflict management practitioners to determine whether any of the criteria could be eliminated and whether any should be added to the set of criteria used to assess conflict forecasting methods. By repeating the Delphi process described in this research using the revised set of criteria, a better understanding of how practitioners select conflict forecasting methods could be obtained.

5.3   Implications for managers

How a conflict forecasting method should be chosen

There are some clear findings from my research that justify strong recommendations. Following these recommendations will result in cost savings (increased profits) from improved decisions. Other recommendations are based on more tentative findings. Nonetheless, they may result in savings and are unlikely to leave managers worse off.

Yokum and Armstrong (1995) seem to have expected the criterion of cost savings from improved decisions to have dominated other criteria for the selection of forecasting methods. Logically, for a profit maximising or cost minimising organisation, it *should* be the principle criterion for forecasting method selection. Other criteria do not need to be

considered unless there is a cost-savings tie or there is uncertainty over the relative cost savings from improved decisions that will arise from the use of contending forecasting methods. If costs are small relative to likely savings or they are similar across methods, then the criterion of cost savings from improved decisions is equivalent to the criterion of accuracy. At least this is so *if* a method does not fail against a criterion that is critical in the particular circumstances. For forecasting decisions in conflicts, criteria that are potentially critical are: timeliness in providing forecasts, ability to compare alternative policies, and ability to compare alternative environments.

The conflict forecasting methods that I have considered can all be conducted cheaply relative to the savings possible from improved decisions in important conflicts. The costs are also likely to be similar across methods (Green, 2001a) if they are implemented with equivalent care. The method that provides forecasts that are most often accurate (simulated interaction) is also the best of the methods for comparing alternative environments and alternative policies.

On the latter point I differ from the Delphi panel (subsection 4.3.2). The panellists rated game theory 8.2 out of 10 for ability to compare alternative policies, but simulated interaction only 6.8 (Table 43). As simulated interaction provides a framework for varying every aspect of a situation and testing the effects of those variations, there appears to be no good reason why simulated interaction should be ranked behind game theory on this criterion.

Timeliness in providing forecasts is thus the only criterion that could reasonably preclude the use of simulated interaction for forecasting decisions in conflicts. While unaided judgements can in practice be made with little preparation, the other methods do require preparation of material and recruitment and organisation of people. This need not take a great deal of time. For example, the New Zealand Armed Offenders Squad uses simulated interaction to test different approaches to dealing with a violent confrontation before they take action. If other managers consider there is insufficient time for such a process, their forecasts would be just as accurate if they rolled dice rather than went to the extra trouble of making an unaided-judgement forecast. Although unaided-judgement forecasts were, on average, more accurate than chance, the difference was small and not statistically significant.

The material presented here is advice to managers based on empirical findings on the performance of forecasting methods, rather than a description of how forecasting methods for conflicts are currently chosen. As it happens, the advice appears to be contrary to common practice and to evidence on what is likely to prove acceptable advice to managers. These are not good reasons, however, for failing to give advice that is likely to be useful to managers who do follow it.

Recommendations on choosing and implementing forecasting methods for conflicts

◆ **Use simulated interaction for forecasting decisions in conflicts whenever the decision that is made may have important consequences**

This advice holds good across diverse conflicts. It also holds good against alternative forecasting methods regardless of the level of expertise of the forecasters using these methods and regardless of whether they collaborate with other forecasters or not. Recommendations on how to organise simulated interactions for forecasting are provided in Armstrong (2001a).

◆ **Use simple measures to increase the representativeness of role-players**

Armstrong (2001a) recommends realistic casting of role-players where this is not too costly, but points out that the available evidence suggests that this is not critical. It may be that being human is all that is required. For example, Sugiyama, Tooby, and Cosmides (2002) found that non-literate Shiwiar hunter-horticulturists of Ecuadorian Amazonia were no different from Harvard undergraduates in the frequency of their choices in regard to cheater behaviour. High levels of accuracy are achievable using student role-players, and students can be hired at relatively low rates.

One measure to increase the representativeness of role-players is to determine which of the parties or roles role-players tend to identify more with, and to allocate roles accordingly. This is easy to do, and seems likely to increase the realism and hence accuracy of simulated-interaction forecasts. While there is no evidence in favour of this practice, it is consistent with the hypothesis on realism and it is unlikely to result in

161

reduced accuracy. Other simple measures may occur to forecasters who use this method. For example, it may be practical and useful in some circumstances to allocate roles on the basis of sex, age, or ethnicity.

◆ Provide only two or three decision options to forecasters

Doing this appears to reduce forecast error rates dramatically relative to providing four or more options. The evidence for this recommendation is limited to the findings from the eight conflicts used in this research. Nevertheless, there is little risk in following this advice as, reduced error rates are likely, even if the reduction is not dramatic.

◆ If more detail is required than can be provided by a forecast of one of three decision options, split the forecasting task into stages

In cases where greater detail is desired on a decision than can be obtained from a forecast of one of three, it may be possible to formulate a forecasting problem in two or more stages with no more than three options at each stage. This advice is speculative.

◆ Use structured analogies when several forecasters can each think of at least two analogies from their own experiences that are similar to the target conflict

If forecasters can think of several conflicts from their own experiences that are similar to a target conflict, their forecasts from structured analogies are quite likely to be accurate. Such forecasts should be combined with simulated-interaction forecasts. Armstrong (2001d) provides evidence that combining forecasts from different forecasting methods tends to increase forecasting accuracy.

◆ **Give simulated-interaction forecasts more weight than structured-analogies forecasts**

Simulated-interaction forecasts are more accurate in more circumstances than are structured-analogies forecasts. There is also less evidence on the accuracy of structured-analogies forecasts. These are good reasons to give more weight to simulated-interaction forecasts.

◆ **Where forecasters using structured analogies are highly confident in the accuracy of their forecasts, give those forecasts more weight**

Among structured-analogies forecasts, more weight should be given to those that are provided by forecasters who are highly confident about them.

◆ **Use several inexpensive experts for structured-analogies forecasts**

The structured-analogies forecasts of experts with many years experience with conflicts, even conflicts similar to a target, are no more accurate than the forecasts of those with lesser experience. Combining the forecasts of several inexpensive experts is likely to provide greater accuracy than the forecast of a single expert with a great deal of experience. This research shows that very experienced experts are no better than those with less experience, and Stewart (2001) provides evidence that combining forecasts is likely to improve accuracy in many circumstances.

◆ **Convert probabilistic forecasts into single-decision forecasts**

Some forecasters may prefer to assign probabilities to decision options rather than to select a single option. In such cases, use the option that was allocated the highest probability by the forecaster as the forecast. That is, if a forecaster provides the probabilities (0.10, 0.20, 0.70) for options A, B, and C, use option C as the forecast. This practice will not reduce accuracy and may increase it.

◆ Use rules to transform analogies information into a single-decision forecast

Forecasters using structured-analogies sometimes provide a forecast that is at odds with their own analogies. Such a forecast is less likely to be accurate than one that is consistent with the forecaster's analogies. Using the rules shown in Figure 1 avoids the problem, and forecasts derived using the rules are no less accurate than those derived from analogies using forecasters' own judgements.

Conclusion

There is compelling evidence that simulated interaction is the best method available for forecasting decisions in conflicts and should therefore be used for that purpose. Forecasts from simulated interaction are substantially more accurate than forecasts from other methods and need be no more expensive. The method can be used to test different strategies and this can be done in a timely manner should this be necessary. Structured-analogies forecasts are likely to be accurate if experts are able to identify two or more analogies from their own experience. The method is less likely to be useful for unusual conflicts or when managers wish to test alternative strategies. Finally, while obtaining unaided-judgement or game-theoretic forecasts from experts may seem convenient, their forecasts are unlikely to be more accurate than chance.

Appendix 1: Application of forecasting method evaluation principles

Application of Principles [a]

Principle	Rating [b] and description of application

A/ Using reasonable alternatives

1 Compare reasonable forecasting methods	U	Methods for comparisons: unaided judgement (usual method); game theory (expert recommendations); structured analogies (usage and expert recommendations); simulated interaction (evidence of superior accuracy)

B/ Testing assumptions

1 Use objective tests of assumptions	Y	Objective tests of verity of individual conflict descriptions & outcome options not possible. Actual partisans reviewed Nurses Dispute & Personal Grievance; others reviewed material in all cases. Tested.
2 Test assumptions for construct validity	Y	Ideally: obtain independent descriptions of conflicts & generate outcome options using brainstorming or similar, then obtain forecasts for all versions. In practice: resources not available; evidence that accurate forecasts can be obtained using one description that has been reviewed
3 Describe conditions for generalisation	U	Problems of forecasting decisions in conflicts with interaction between a small number of parties in conflict; effect of expertise & collaboration
4 Match tests to the problem	U	Some *ex ante* test results available for forecast validity; most tests use disguised or obscure historical conflicts; a variety of conflicts from different domains are used
5 Tailor analysis to the decision	U	Assessed appeal of the methods to managers; decision scores to compare forecasting accuracy C "wrong" options may have non-zero value

C/ Testing data and methods

1 Describe potential biases	U	Selection of interesting conflicts may bias sample towards those difficult to forecast using judgement => tested for larger gains in accuracy from role-play vs judgement for conflicts that are most difficult for judges
2 Assess reliability and validity of data	U	Endeavoured to collect 10 or more forecasts from each method for each conflict; implement the methods in the same (or a similar) way to the way they would be implemented in the field. Number of forecasts low in some cases
3 Provide easy access to data	U	Publish data on the internet (kestencgreen.com); maximise disclosure without breaching confidentiality; will make more data available
4 Disclose details of methods	U	Provide full descriptions of methods, including any deficiencies in implementation
5 Do clients understand [and accept] the methods?	U	Compare practical forecasting methods; use Delphi panel to assess appeal to managers of methods

D/ Replicating outputs

1 Use direct replication to identify mistakes	**U**	Examine prior research for errors in analysis; replicate some unaided judgement research using same conflict descriptions
2 Replicate studies to assess reliability	**U**	Apply unaided judgement and simulated interaction methods to more conflicts
3 Extend studies to assess generalisability	**U**	Increase the variety of conflicts used to test the methods; add methods **C** game theory & structured analogies; vary conditions
4 Conduct extensions in realistic situations	**U**	Implement forecasting methods in ways that are the same as or similar to the way they are likely to be implemented in the field
5 Compare with forecasts from different methods	**U**	The research programme is designed to compare methods

E/ Assessing outputs

1 Examine all important criteria	**U**	Accuracy is main criterion; other criteria will be rated for importance by Delphi panel who will then rate methods on basis of criteria
2 Specify criteria in advance	**U**	Criteria from Armstrong (2001c) rated, in this research, by Delphi panel for importance for methods for forecasting decisions in conflicts
3 Assess face validity of methods & forecasts	**U**	Client acceptance provides measure of method face validity [c]; unaided judgement forecasts by experts tests face validity of other methods; but risk that these will be poor measures of forecast validity; Delphi results
4 Adjust error measures for scale	**F**	Scale is not an issue in this research (see also E11)
5 Ensure error measures are valid	**U**	Seek peer review on measures (face validity); compare with other reasonable measures e.g. would forecast lead to wrong decision (construct validity)
6 Ensure error measures *in*sensitive to difficulty	**U**	Discount method error by a priori difficulty: (1 - (1 / number of options)); and by unaided error rate
7 Ensure error measures are *un*biased	**U**	Test sensitivity of rating of methods by excluding each conflict in turn from the calculations (jackknife procedures)
8 Ensure error measures are *in*sensitive to outliers	**U**	All forecasts will be used in comparisons *unless* there is compelling evidence for exclusion, such as: conflict does not match research criteria; participant recognised situation; etc.; compare medians also (Page test for ordered alternatives does this)
9 Do *not* use R^2 to compare models	**U**	
10 Do *not* use RMSE	**U**	
11 Use multiple error measures	**U**	Measure error based on percentage accuracy, *PERVC*, *PERUJ*, *BS*, *PFAR*, & decision scores
12 Use ex ante tests for accuracy	**U**	Participants do not know the actual outcomes of the conflicts they are forecasting as the conflicts are disguised or unresolved

13 Use statistical significance to test only reasonable models	U	With a small sample of conflicts, tests of the statistical significance of differences in accuracy between methods may be needed to confirm any apparent practical significance in accuracy differences
14 Use *ex post* tests for policy effects	U	This is not an issue for conflicts for which descriptions are written *ex post*; may need to drop conflicts with descriptions written *ex ante if* it becomes clear that relevant information or options, which could and should have been included, were not
15 Obtain large samples of independent forecast errors	U	Use sample of heterogeneous conflicts and jackknife procedures to test errors, but a risk remains that the sample conflicts are not representative
16 Conduct an explicit cost/benefit analysis	U	Costs are easy to estimate and not high; benefits depend on how much is at stake; assume a lot at stake => cost becomes insignificant => net benefit ranking is reverse of average error ranking of methods

[a] Based on Exhibit 10 in Armstrong J. S. (2001e).

[b] U: Principle was applied; Y: Principle was not applied or was poorly applied; F: Principle was irrelevant to this study

[c] Client for role-play research on Nurses Dispute & Personal Grievance; Zenith Investment used in teaching; Armstrong J. S., Brodie R. J. & McIntyre, S. H. (1987).

Appendix 2: Conflict descriptions and questionnaires provided to game theorist participants

<u>ARTISTS' REPRIEVE</u>

The Country

Histavia is a relatively small country which is also one of the most densely populated nations in the world. It has some energy resources (mostly gas) but very few other natural resources. Farming, fishing and trading have always been important parts of this country's economy. More recently the country has become quite involved in manufacturing with over 40% of the people employed in various industries. The government is renowned for its bureaucracy and for the extravagant systems it puts together. For example, in order to help with the population problem and overcrowding, the government subsidised emigration.

The government is a constitutional monarchy. Executive power lies in a cabinet selected by the majority in Parliament. Histavia's parliament has two houses. The 1st House contains appointed representatives from each province's legislative body. The 2nd House is elected by direct vote and develops all legislation. A bill passed by the 2nd House is sent to the 1st House for a vote.

Artistic Problems

Visual art has always been a very important part of Histavian culture. Three hundred years ago, Histavia's seaports and rivers made it a major trading hub for the surrounding nations. Histavians became relatively wealthy and wanted to chronicle the times. They wanted pictures of themselves, their families, homes, and the countryside. Demand was so great that an artist could specialise in whatever subject she wanted to paint. Out of this environment came some of the world's greatest artists. The Histavian people thus developed a special regard for art and artists.

This prosperity didn't last. The private market for art has been very weak for a long time. There are currently sufficient jobs available for artists but there isn't enough demand for many artists to make a living selling their works. The government feels a compelling need to do something about this problem. Since the private market isn't there, the government feels obliged to protect the artists' financial freedom and thereby to preserve art for the sake of art. Partially to reward artists for their contribution to Histavia's culture and partially because the main artists' union has made a big fuss over the continuing conditions for art, the government has developed a simple programme to keep artists from having to drive taxi cabs. An artist will sign up on the programme for up to a year. During that time the government will buy the artist's wares. Once the artist's time period is up, the artist must sell his works on his own or find another occupation. During the government purchase period, if the artist begins to sell her works in the private market, the government will end its assistance early.

Counterculture

The programme was accepted and seemed to work well for about twenty years. In the late 1960's the counterculture boomed worldwide. Artists in Histavia were particularly upset with the continuing conditions for artists. They felt anyone who said they were an artist was one, and that the programme's length wasn't sufficient time for an artist to get established given the continuing weak demand for art. To press their point, members of the main artists' union occupied the most valuable room of Histavia's primary art museum. They refused to leave until entrance requirements were relaxed and artists were allowed to remain in the programme indefinitely.

Additional Information

- The conflict occurs in the late-1960's
- This is a developed, industrialised country (i.e.; not Third World)
- Not all artists are members of the main union, but union leaders feel they speak for and represent all of Histavia's artists. They also believe they can gain public support for their cause fairly easily.

Current entrance requirements to the programme involve acceptance by the government on the basis of some criteria.

ARTISTS' REPRIEVE

Roles: Histavian Bureaucrats

You are two Histavian bureaucrats responsible for administering the government's programme for purchasing art. You report directly to the executive cabinet, which represents the majority in both Houses and the view of the Queen. The cabinet has given you negotiating power to resolve the artists' sit-in and feels confident it can win support for whatever agreement you reach with the artists. The government and the people are very concerned for the well being of art and artists. The majority party, as always, is concerned with negative public opinion affecting the outcome of the next election. Additionally, the room the artists are holding contains some of the country's most treasured art works. It's your job to negotiate and resolve the issue.

All facts and impressions that are available are given. Extrapolate as necessary.

ARTISTS' REPRIEVE

Roles: Artists' Union Leaders

You are two leaders of the main artists' union and have taken it upon yourselves to represent not only your fellow union members in the room, but all Histavian artists. You know one of the most important things to Histavian people is their art. Still the private demand for art isn't sufficient to support all the artists Histavia's rich culture has inspired. You're meeting with two government agents responsible for administering the current programme. They are influential people who have been sent to resolve the sit-in and have the backing of the majority party in power. Right now you occupy the room in the art museum which contains the country's most treasured art works. You want exit and entrance requirements to the programme relaxed before you'll leave. You feel strongly you represent the viewpoint of all artists and it will be easy to gain public support for your cause.

All facts and impressions that are available are given. Extrapolate as necessary.

ARTISTS' REPRIEVE

1) **What will be the final resolution of the artists' sit-in?**   *(check one - ✓)*

    (A)   The government will relax entrance rules and allow an artist to remain
            in the programme for an indefinite period.   [ ]

    (B)   The government will extend an artist's time in the programme to
            2 or 3 years.   [ ]

    (C)   The government will extend the programme 2 or 3 years and relax entrance rules.
                                                                               [ ]

    (D)   The government will relax the entrance requirements only   [ ]

    (E)   The government will make no change in the programme.   [ ]

    (F)   The government will end the programme completely.   [ ]

2) **Broadly, what approach did you use to derive your prediction?**

3) **Roughly, how long did you spend on the task of deriving a prediction for this situation?**
                                                                 [____] hours

4) **If you have not provided a prediction, please state your reasons:**

5) **Roughly, how many years have you spent as a game theory practitioner or researcher?**
                                                                 [____] years

When you have completed this questionnaire, please return either
this document as an email attachment to
kesten.green@vuw.ac.nz
or this page (with your initials printed below) by fax to
(64 4) 499 2080.

Your initials: [_____]

<div align="center">**DISTRIBUTION PLAN**</div>

Background

The year is 1961. The Ace Company has been in business over seventy years and has become a major producer of home appliances. The home appliance industry has had a terrible start this decade. Sales have been weak, inventories are high, dealers are demoralised, and mass merchandisers and foreign competitors have entered the market slashing margins. While the recession seems to have bottomed out by mid-1961, Ace's operating deficit is approaching $6 million for the year. The company, however, feels the fall introduction of the colour TV set might recover some the loss.

Ace's problems seem to be short term. Some existing new products in the development cycle are draining funds and the year's poor sales has created a cash flow crisis. Now that consumer purchases are picking up, funds aren't available for heavy promotion and prices are still soft due to foreign competition and excess inventories. One component of the marketing mix can be attacked – distribution.

Appliance Discount Plan

Competition between supermarkets for customers has always been heavy. Discount houses are opening up and every store seems to be giving trading stamps or running some kind of promotion. Ace feels it has come up with an excellent new distribution plan that will make everyone involved happy and return much needed appliance sales. It's called the Cash Register Tape Plan (CRTP). An Ace dealer will link up with an area supermarket. The supermarket will be given an exclusive contract in return for floor space to display some major appliances. The dealer will supply a sales person to explain the appliances and show pictures of items not on display. A shopper accepted by Ace Financing receives a 12-36 month, no down payment instalment plan and has the item immediately delivered by the dealer. Each month the purchasers bring in their cash register tapes from that supermarket. Five and a half percent of the tape total is taken off the monthly instalment (up to 75% of the payment). The payment of the monthly discount is split between Ace and the supermarket based on a sliding scale. If a shopper purchases less than $50, Ace pays the full 5½%. If she purchases more than $120, then Ace pays 2½% and the supermarket 3%.

The CRTP is designed to benefit all those involved. Naturally Ace and their dealers expect to enjoy increased sales. The shopper will get a reduced price on a major appliance by altering purchasing habits. The supermarket should be able to reduce split market shopping, increase purchases by regular customers of items often purchased elsewhere, obtain new customers, and generally build traffic. There are, however, costs to the plan. Ace receives a lower price, the dealer has a salesperson at a remote location, the customer won't be able to "shop around" for the lowest prices and the supermarket has to give up floor space and pay part of the discount.

Selection Process

Ace wants to deal with regional chains. This strategy will require getting agreement from only one source (per regional market) before they're able to start up the CRTP in the metropolitan areas they've selected. Expansion to additional cities will also be easily accommodated. Ace will select the dealers who will participate based on their proximity to the individual stores selected to participate from the supermarket chains. Both the dealer and the supermarket must be approved by Ace Financing before they can join the CRTP. Dealers not near a selected supermarket or in a region not selected for the plan won't participate. If dealerships overlap a supermarket's territory they will sell in the stores on alternate days.

Additional Information

- Ace carries a full line of major household appliances and is a well-known national manufacturer.
- Small appliances (like toasters) have been sold through supermarkets as promotion items
- Assume Ace can present some attractive return figures for the supermarket based on the given sliding scale discount payment procedure
- No actual test of the CRTP has been conducted

DISTRIBUTION PLAN

Roles: ACE Executives

You are the ACE marketing vice-president and the marketing director of the appliance division. You are preparing to meet with the management of BIG VALUE supermarkets (a mid-West grocery store chain). They have been approached about the CRTP, received an outline of the plan, and have agreed to meet with you. In addition, you've sent a market analysis your division pulled together indicating how much a chain like BIG VALUE should profit by participating in the CRTP.

It's your job to sell the CRTP! You see it as a potential saviour for ACE's current problems and a unique distribution tool. You are particularly worried about other stores, dealers, and/or manufacturers copying the CRTP. You want to move as quickly as possible once the programme starts. Obviously the continued sales slump and cash flow problems will negatively reflect on your careers.

All facts and impressions that are available are given. Extrapolate as necessary.

DISTRIBUTION PLAN

Roles: BIG VALUE Vice-Presidents

You are the two top vice-presidents of BIG VALUE supermarkets, a mid-West, regional grocery store chain. You have agreed to meet with the marketing v-p and marketing director of ACE's appliance division after having received information on ACE's CRTP and a market analysis (from ACE) which projects favourable returns for a store chain like yours. You have also heard about ACE's current problems.

It's your job to discuss the pros and cons of the CRTP for your chain and be prepared to react to the ACE personnel. Ultimately you must decide whether or not to accept the CRTP and in what form. BIG VALUE's top management is aware of the plan and will accept your decision. Positive or negative results of your decision will obviously impact on your career.

All facts and impressions that are available are given. Extrapolate as necessary.

DISTRIBUTION PLAN

1) **Will the management of a supermarket chain accept the CRTP in their stores?**

(check one-✓)

    (A)   Yes, as a long term arrangement   [ ]
               (with one month pilot)

    (B)   Yes, as a short term promotion   [ ]

    (C)   Yes, either (A) or (B)   [ ]

    (D)   No, they will reject the plan   [ ]

2) **Broadly, what approach did you use to derive your prediction?**

3) **Roughly, how long did you spend on the task of deriving a prediction for this situation?**

[_____] hours

4) **If you have not provided a prediction, please state your reasons?**

5) **Roughly, how many years have you spent as a game theory practitioner or researcher?**

[_____] years

When you have completed this questionnaire, please return either
this document as an email attachment to
kesten.green@vuw.ac.nz
or this page (with your initials printed below) by fax to
(64 4) 499 2080.

Your initials: [_____]

THE 55% PLAN

The collective bargaining agreement between the National Football League (NFL) Management Council and the NFL Players' Association (NFLPA) expires on July 15, 1982. The payers' number one demand is for a fixed percentage of the football clubs' gross revenue. Ed Garvey, the executive director of the NFLPA since 1971, says that 55% of the gross revenue is the players' "bottom line" demand. Jack Dolan, executive director of the Management Council since 1980, rejects the union's demand for a percentage of the gross revenue regardless of what the percentage is.

NFLPA

The players' last contract ended in 1974. The players struck for six weeks during the '74 season but returned to work without a contract to begin playing regular season games. The major issue then was free agency. A free agent is a player who doesn't sign a new contract with his team when the current one expires. Once the contract ends, any team can bid for his services. The owners refused to agree to a free agent system so the NFLPA took the case to federal court and won most of their demands. This decree, however, didn't come until 1976. The final contract was signed in 1977, three years after the last contract expired.

Unrestricted free agency in most sports has caused players' salaries to increase substantially, however this has not happened in football. Of the five hundred players who became free agents since 1977, only six have changed teams. The average salary in the NFL was $78,000 in 1980 compared with $108,000 in hockey, $143,000 in baseball, and $186,000 in basketball. Garvey insists the poor performance of the free agent system in football occurs because the NFL clubs practice "corporate socialism". All clubs share equally in TV revenues and playoff monies. This ensures even a team with a terrible record will be financially sound. There is, therefore, economic incentive to replace more expensive veterans with less expensive rookies. Others feel Garvey, at least in part, is to blame for the slow movement of free agents. In the '77 contract, the union agreed to compensation procedures for a team losing a free agent. Many feel the nature and magnitude of the compensation restricts free agency to the point where no club is willing to sign a good free agent. They point to the case when one of the best running backs became a free agent and not one club made him an offer.

The 55% plan is designed to increase all players' salaries and provide incentive bonuses for excellent playing. Some of the higher paid players, however, may not benefit very much under the proposed method for distributing the wages. This plan will also eliminate the individual contracts negotiated by the teams with each player and, in effect, make the players' agents obsolete. As for the players' unity and their resolve to attain the 55% plan, Garvey feels virtually every player is ready to strike and there is no comparison between the situations in 1974 and 1982.

All players now belong to the NFLPA. Communication with the dispersed membership is vastly improved—players and their wives have been thoroughly briefed on the 55% plan, and there will be a union meeting in mid-March to iron out all the details of the wage distribution. The AFL-CIO has been asked to provide assistance from the entire labour community and a battle cry is being sounded in a 20 minute film featuring popular sports and entertainment personalities.

NFL Management Council

While Dolan is the new kid on the block, he's no rookie. As a senior vice-president of industrial relations for nine years with National Airlines, he negotiated contracts with eight unions and went through four strikes. Dolan consults with six owners who make up the council's executive committee. All, except one, are described as "hard liners".

To the owners, a fixed percentage is equated with control and the owners "are not going to let the union run their business". They deny Garvey's contention they have no incentive to win because of the leagues financial structure. Besides being successful businesspeople who want their clubs to do the best they can, the team's won-lost record effects the price owners can charge for tickets, the demand for luxury boxes, and the amount of money lost from concessions because of no shows. From the union's own estimates, the LA Rams had $7.7 million in profits for 1980 while the Denver Broncos had $2.5 million.

The council feels Garvey will have trouble gaining public sentiment and maintaining rank and file support for such an unorthodox plan. In addition, their pre-contract strategy appears to be to build a lack of confidence in Garvey, so it will be difficult for him to maintain players' support for a strike. People either love Garvey or hate him. His shoot from the hip, wisecracking style often leaves a trail of damning quotes. The NFL Management Council uses a monthly newsletter to "their employees" and the sports press to present Garvey's quotes and actions in the worst light. Dolan is intent on using these weaknesses to stop any attempt to have a fixed percentage of gross revenues as the collective bargaining agreement's wage clause. Estimates of the percentage of gross revenues currently going to the players are between 25% and 45%. Naturally the NFLPA claims the figure is between 25-30% and the NFL Management Council places it near 45%.

Additional Information

- Current compensation losing a free agent is by way of draft choices, i.e. if the player's salary is greater than $x, the team signing the free agent loses two 1st round draft choices. A draft choice is a team's right to select a player from a list of all available college players. Teams take turns to pick, with the weakest team choosing first and so on.
- The percentage of gross revenues will go into one pool, which the NFLPA will distribute based on a base salary plus incentive bonuses.

THE 55% PLAN

Roles: Management & Owner Representatives

You are the executive director and the head owner representative of the NFL Management Council. You are about to meet with the NFLPA to decide the 55% issue. You both hold very strongly to the statements attributed to the owners and view the NFLPA comments as "hogwash" and propaganda. Prepare and defend your position for the upcoming negotiation.

All facts and impressions that are available are given. Extrapolate as necessary.

THE 55% PLAN

Roles: Player Representatives

You are the executive director and the head player representative of the NFLPA. You are about to meet with the Management Council to negotiate and decide the 55% issues. You both hold very strongly to the statements attributed to the NFLPA and view the owners' comments as "hogwash" and propaganda. Prepare and defend your position.

All facts and impressions that are available are given. Extrapolate as necessary.

THE 55% PLAN

1) **Will there be a strike?** *(check one - ✔)*

    (A)   Yes, a long strike   [ ]
                (½ or more of the regular season games will be missed)

    (B)   Yes, a medium length strike   [ ]
                (less than ½ of the regular season games will be affected)

    (C)   Yes, a short strike   [ ]
                (only preseason games missed)

    (D)   No strike will occur   [ ]

2) **Broadly, what approach did you use to derive your prediction?**

3) **Roughly, how long did you spend on the task of deriving a prediction for this situation?**
                                                [____] hours

4) **If you have not provided a prediction, please state your reasons:**

5) **Roughly, how many years have you spent as a game theory practitioner or researcher?**
                                                [____] years

When you have completed this questionnaire, please return either
this document as an email attachment to
kesten.green@vuw.ac.nz
<u>or</u> this page (with your initials printed below) by fax to
(64 4) 499 2080.

Your initials: [_____]

A Pay Dispute Between Capital Coast Health and Nursing Staff

Capital Coast Health Ltd is a government-owned Hospital and Health Service. The company is probably best known to the general public as the operator of Wellington, Kenepuru, and Paraparaumu Hospitals.

The Government funds CCH primarily by purchasing a set level of health services, for example so many hip operations. The level of funding is based on what the Health Funding Authority has calculated are reasonable prices to pay for the services the Government wishes to purchase.

It is the responsibility of the CCH Chief Executive, Margot Mains, to manage the company's affairs so that it can deliver the services it has contracted to provide while remaining financially viable. CCH's 2000/2001 contract with the HFA is worth $258 million, an increase of nearly 10% over the previous year's contract; this reflects both some price increases and some increase in services provided. There are more than 3,000 staff, and payroll costs are more than $140 million per year – over half of the value of the HFA contract. CCH employs nearly 1300 full time equivalent nurses; their pay costs roughly $55 million per year. The collective employment contract many of the nurses are party to has expired and the nurses and management are currently involved in a dispute over their pay.

The New Zealand Nurses Organisation (NZNO) represents more than half of the nurses employed by CCH. The nurses are angry. They consider that they have borne the brunt of budget constraints over the years and that their dedication to their profession has been exploited. Perhaps as a consequence of this, the turnover of nurses is high – as many as one-third of nurses leave CCH during the course of a year, often taking up highly paid nursing jobs overseas.

Some specialist nurses are more difficult to replace than others. Earlier this year, Intensive Care nurses obtained a 7% pay increase from CCH. Junior doctors also obtained a large pay increase in separate negotiations. With these precedents, nurses instructed NZNO to retract the bid they made in February for a 5% pay increase and instead seek 7% on all salaries and allowances for the 12 months from 1 October. In addition, the NZNO are seeking substantial increases in penal rate multipliers, meal, night and weekend allowances, study and long-service leave provisions, and reimbursement of professional association costs.

Margot Mains and the CCH negotiation team, led by Mike Hanson, have now offered the nurses a 5% pay rise with no new contract for 2 years and have stated that this is most the company can afford. At a stop-work meeting nurses rejected the 5% offer and voted almost unanimously in support of holding a 16 hour strike this Sunday, October the 1st. Nurses at the meeting criticised NZNO negotiators for being too willing to compromise. The nurses gave NZNO negotiators Russell Taylor, Alistair Buchan, and the rest of the team clear instructions "not to come back" without an agreement for at least a 7% increase.

The two sides are at odds in their interpretation of CCH's ability to afford a 7% increase. CCH has been struggling to reduce debt levels by reducing costs, increasing efficiency, and only providing services it is funded to provide. The Government requires CCH to break-even, and the CCH negotiators claim that even a 5% pay increase would result in a deficit if cost savings cannot be found elsewhere. The NZNO, on the other hand, maintains that CCH is effectively receiving a bonus from the HFA when they are paid for additional procedures. This is because providing an *additional* procedure of a type CCH is already providing does not cost as much as the standard price per procedure the HFA pays. Moreover, NZNO claims that a 2.5% "creep" in salaries that CCH has budgeted for will not occur due to the current composition of the nursing workforce and the high attrition rate. Creep occurs when staff receive increases in salary that are related to their length of service. The NZNO position is that the "bonus" revenue that will accrue from payments for additional procedures and the absence of creep are sufficient to meet the $1.1m difference between CCH's 5% offer and the nurses' 7% claim.

Both parties have a lot at stake. Margot Mains' performance is assessed, in part, on her financial management of CCH, while the nurses are prepared to go on strike in support of their strongly held belief that they deserve a 7% pay increase. Meanwhile, the Minister of Health Annette King has been sending mixed signals to the two parties about what she sees as an appropriate settlement and has also stated that the Government will *not* become involved in the dispute.

Assume it is now Monday 2 October. The Sunday strike is now over and the two parties are about to recommence negotiations for a settlement of the dispute. The dispute will be resolved, but after how long and with what terms?

If in impasse in negotiations is reached David Hurley, an independent mediator from the new Mediation Service, will be appointed. David will briefly interview each party separately and will then endeavour to mediate an agreement.

Margot Mains, Chief Executive of Capital Coast Health

Margot Mains trained and worked as a nurse, but has an MBA and has for some years held senior management positions. Most recently, before taking up the appointment at CCH, she was the Chief Executive of Mid-Central Health, which is based at Palmerston North Hospital. It was her success in that position that made her a strong candidate for the position she currently holds.

Her role is not easy, and probably rather thankless. Government expects her to run CCH as a business in many ways but, in reality, it is neither fish nor fowl. She is faced with political pressures and with the fact that many people, including many nurses, do not believe that hospitals really are, or should be, businesses. In this ambiguous environment, she is obliged to keep within her budget while ensuring that the hospitals continue to run as smoothly as possible.

Margot has a good working relationship with Mike Hanson, her chief negotiator. She has used him and his company for advice with labour relations and human resource matters for many years, since her days at Mid-Central.

Mike Hanson, Capital Coast Health Negotiator

Mike Hanson is a seasoned industrial negotiator and human resources expert with his own successful business. He has a lot of experience in the health sector.

Mike has a direct and pragmatic style and, like Margot, has quite a lot of sympathy for the nurses. He is, however, employed by CCH and his role is to help Margot Mains by negotiating an agreement with the nurses' representatives that is consistent with achieving a healthy financial position for CCH without undermining CCH's role as a health services provider.

Russell Taylor, New Zealand Nurses Organisation Organiser

Russell Taylor is a seasoned industrial negotiator and union organiser. Before taking up his present job, he worked for the Public Service Association, the main union for public sector employees.

Russell's current role is to represent the interests and wishes of the nurses employed by CCH who are members of the NZNO, the nurses' union. This includes organising the nurses to take effective protest and strike action. Russell and his team negotiate on their behalf in order to get the best possible deal for them. Because he is dealing with the CCH managers and negotiators a lot, he understands their point of view, perhaps better than most members, and has developed a good working relationship with them. At the end of the day, however, he can only make recommendations to the nurses. It is they who will decide whether to accept or reject a pay offer and whether or not they will go on strike or take other action. Feelings are running high in this dispute and there is every sign that the nurses are determined to take further action if necessary.

Russell has an open style and believes in the cause of his members. He is being assisted in the current round of negotiations by Alistair Buchan. Alistair is a senior nurse, and a real asset to the negotiating team.

Alistair Buchan, Nurse and Negotiator

Alistair is an experienced nurse who has relatively recently returned to New Zealand after a number of years abroad. He was very well paid working as a nurse overseas. The contrast with the poor situation of nurses in New Zealand is marked.

As well as working as a nurse for Capital Coast Health, Alistair is currently helping the New Zealand Nurses Organisation, as a negotiator in the pay dispute. He represents the interests and wishes of the nurses employed by CCH who are members of the union.

David Hurley, independent mediator, Mediation Service

David Hurley may be asked to act as mediator in the dispute.
His role as mediator would be to:
- control the process
- ensure that each party states their case
- help the parties to identify issues and options for settlement
- look for a solution that will satisfy the needs & interests of both parties
- record the agreement in writing

The Department of Labour's Mediation Service web-site includes the following information on mediation:
"Mediators will help the parties decide on the process that is most likely to resolve problems as quickly and fairly as possible." *and*

"If the employer and employee cannot reach agreement in mediation, they can agree, in writing, to the mediator making a final and binding decision. The mediator will explain to the parties that once he or she makes a decision, that decision cannot be challenged. The mediator's decision is enforceable…"

CCH / Nurses Dispute Outcomes

1. The outcome of the negotiation was that… *{please tick one box only}*
a. Nurses' demand for an immediate 7% pay rise and a 1-year term was substantially or entirely met [__]
b. CCH's offer of a 5% pay rise and a 2-year term was substantially or entirely accepted [__]
c. A compromise was reached [__]

…

7. Please indicate whether you know more about the situation or the people than has been presented to you and, if you <u>do</u> know more, please state what it is you know and how you know it:
      a. Do you know more?   Yes [___]   No [____]
      b. Briefly, what do you know about the people, the situation, <u>or any similar situation</u>?

---

8. Please describe how you derived your predictions for questions 1 to 3, above:

9. Roughly, how long did you spend deriving a prediction for this situation (including time spent reading the material)?   [____] hours

10. If you have not provided a prediction, please state your reasons:

11. Roughly, how many years have you spent as a game theory practitioner or researcher? [____] years

When you have completed this questionnaire, please return either
this document as an email attachment to
kesten.green@vuw.ac.nz
<u>or</u> this page (with your initials printed below) by fax to
(64 4) 499 2080.

Your initials: [_____]

What's the job worth?

In New Zealand universities most students belong to a students association to which they contribute financially. The University of Nelson Students Association (UNSA) is one such association. Each year, students elect fellow students to form the UNSA Executive. The Executive is responsible for the representation of students' collective and individual interests (services), and for operating the facilities such as venues, recreation centre, book shop, cafes, and bars that are owned by UNSA (trading). Day-to-day, services and trading activities are carried out by the 150 staff employed by UNSA. The members of the Executive are not professional managers. This had caused problems with the management of UNSA's permanent staff in the past. Wishing to avoid such problems in future, the Executive decided to restructure the organisation and employ a professional general manager to take primary responsibility for UNSA trading and services.

The young new General Manager, Dave Sanders, started work on 3 May 2000. Dave's background was in large private sector corporations. On top of learning a new job, he was responsible for restructuring the organisation and renegotiating contracts with major suppliers and the University. He was also straight away involved in pay negotiations with nearly half of the UNSA staff. Some of the staff (five of the eight services staff) belonged to the university staff union — the Association of University Staff or AUS. Services staff contracts had expired in 1998 and Arty Cruickshank, the local AUS representative had written to Dave the day before he started work at UNSA seeking a meeting to discuss renegotiation of the five AUS members' contracts. Dave telephoned Arty on 4 May and they arranged to meet on 9 May.

At the meeting, Arty proposed that his members' salary bands be reviewed. He suggested that either UNSA adopt the University's general staff job evaluation process and the associated salary bands or that UNSA undertake a formal review of the positions. Dave asked for time to consider this matter and to familiarise himself with the roles of the services staff. After further consideration Dave decided that the University's process and salary bands were not appropriate for UNSA staff and, besides, the University HR department weren't prepared to do the evaluation for UNSA. Arty was informed of this by 23 May, and the possibility of getting an independent assessment was discussed by Dave and Arty. Arty wrote that staff were keen on the idea, but suggested that both they and the AUS should be allowed to examine the proposed process in order to assess its fairness.

In early June, Dave gave *all* salaried staff a backdated 2% pay increase in recognition of the delays that had occurred in pay negotiations prior to his arrival. A memo to this effect was circulated on 20 June. The memo told staff to update their position descriptions and forward them to Dave before 30 June 2000 so that a formal review of the positions and performance of staff could be conducted. Arty phoned Dave wondering about progress with the independent assessment and was informed by Dave that he had spoken to an independent HR consultant who had advised him that UNSA first needed to make sure the position descriptions were consistent with the current needs of the organisation. In the event, the full set of position descriptions was not available until mid-August at which point Dave informed Arty that he and UNSA President Gerald Sullivan were in the process of reviewing the descriptions. In late-August, Dave and Gerald met with each staff member individually to discuss their position descriptions before finalising them.

On 19 October, Dave advised Arty that Gerald had undertaken the evaluation by comparing positions and salary bands with those of other student associations. The resulting bands were higher than those of other associations reflecting management's ambition that UNSA should be the top association in NZ. Dave presented the salary bands to Arty and the staff on 9 November and it was agreed that Dave would explain the bands to individual staff. He did this on 23 November and the parties met again on 21 December. Arty told Dave that the staff considered comparison with the other associations was inappropriate and that the salary bands should be wider. Arty proposed comparisons with positions that the staff considered to be more relevant. Dave agreed to consider these submissions.

Dave, Arty, and services staff met again on 8 February 2001 to discuss the comparisons. Some changes were agreed to. Arty complained on behalf of staff over the time taken to deal with this matter and Dave responded by reminding Arty that he had considerable demands on his time, had given a backdated pay increase soon after taking on the job and that he would make a final decision on salary bands within one week. He said that UNSA could not afford to move substantially on salaries. On 14 February, Dave wrote to Arty informing him of the new salary bands that UNSA had decided on. The letter stated the bands had been set after careful deliberation that included consideration of the arguments put forward by Arty as well as UNSA's ability to pay and ability to recruit. Dave's letter noted that the salary band for one position had been increased and the upper limit of all bands had been increased by $2,000 from the figures presented to Arty on 9 November the previous year. Dave met with staff and informed them of the outcome of the review.

Despite the increases, the new upper salary limit for the position held by one senior and very experienced staff member, Education Co-ordinator Freda Thornley, was less than she was currently being paid. Freda's role is to ensure that students' education interests are represented to the university. The role involves advice, advocacy, communication, counselling, liaison, organisation, and research. Freda has held the position for 12 years and has a great deal of institutional knowledge as well as established relationships with the academic staff of the University. Dave did not propose reducing Freda's (or any other employee's) salary, but it was clear that she should not expect any pay rise in future.

Nearly one month later, Arty informed Dave that Freda and her assistant did not consider that the evaluation of their positions had been conducted properly and, consequently, had been undervalued. Arty raised the matter with the Employment Relations Authority. The Authority directed the parties to meet with the Mediation Service's Mel Morrissey on 24 April with the purpose of resolving differences over the position evaluation process. For equity reasons, any new evaluation would need to include all services staff. Dave Sanders will attend the meeting with employment lawyer Linda Lachlan and Freda Thornley will attend with AUS representative Arty Cruickshank.

Descriptions of how the people who will attend the meeting see their roles and the situation are on the next page.

Dave Sanders, General Manager

You were appointed to the new position of General Manager of the University of Nelson Students Association (UNSA) 18 months ago in May 2000. You have previously worked as a manager in large private sector corporations. You had taken on quite a challenge as, on top of familiarising yourself with a very different organisation to the ones you were used to, you were faced with restructuring the organisation, with building relationships with your board (the student Executive), with staff, with the University, and with suppliers.

UNSA's revenue comes partly from trading activities and partly from student fees. Before you took up your new position, the University had negotiated with an earlier student Executive to reduce the fee per student that was paid to UNSA. As a result, when you took on UNSA it had just made its first ever annual loss. You are determined to return UNSA to profitability, and establish it as a model student association run in a professional manner for the benefit of University of Nelson students. You must balance the limitations of a tight budget with the need to pay salaries that will be attractive to capable people. After conducting an evaluation, UNSA is offering salary bands that are higher than other associations. Despite this, a handful of staff want more. You believe this group's bands are generous not only relative to similar positions in other organisations, but also relative to those of other UNSA employees who have more important roles with responsibility for more staff and bigger budgets. You aren't prepared to make exceptions for this group, as that would leave other staff feeling cheated.

Freda Thornley, Education Co-ordinator

You are the Education Co-ordinator of the University of Nelson Students Association (UNSA), a position you have held for many years. You represent students' education interests to the University and provide advice and support to students. Few people in UNSA seem to know or care what it is that you and your assistant do, and yet you perform a key function of UNSA very well with limited resources. You have developed a broad network of relationships with people in the University and a comprehensive institutional knowledge. The better the job you do the more in demand you are and, consequently, the more pressure you are under. You seldom have time for lunch breaks and never for morning or afternoon tea. You do not believe that you are paid adequately for the role you perform.

In the past you reported directly to an elected student President and Executive. This relationship often involved a lot of hand-holding by you as the student representatives are typically inexperienced. This situation changed recently with the appointment of a professional manager.

Arty Cruickshank, union organiser

You are an organiser for the Association of University Staff (AUS). You role is to look after the interests of members working at the University of Nelson. You are also responsible for five members who work for the University of Nelson Students Association, which is on campus but independent of the University.

You represent AUS members in the negotiation of collective agreements with their employer, and you also represent individual members if they have a grievance or become involved in a dispute.

Mediation: You are aware from previous experience that the objective of mediation is to provide each party with the opportunity to state their case, identify the issues behind their employment relationship problem, identify options for settlement and to attempt to reach a solution that satisfies the needs and interests of both parties. A mediator will control the mediation process to ensure that each party is able to participate, and will record the outcome of the mediation if the parties reach a settlement.

Linda Lachlan, employment lawyer

You are an employment lawyer with the local employers association. Your role is to advise member employers on their rights and obligations under the law and, if requested, to advise and represent members should they become involved in a dispute with an employee.

You have formed the view that the employer has conducted a fair process.

Mediation: You are aware from previous experience that the objective of mediation is to provide each party with the opportunity to state their case, identify the issues behind their employment relationship problem, identify options for settlement and to attempt to reach a solution that satisfies the needs and interests of both parties. A mediator will control the mediation process to ensure that each party is able to participate, and will record the outcome of the mediation if the parties reach a settlement.

Mel Morrissey, Mediator

You are a mediator with the Department of Labour's Mediation Service. Your role is to:

    control the mediation process

    ensure that each party states their case

    help the parties to identify issues and options for settlement

    look for a solution that will satisfy the needs & interests of both parties

    record the agreement in writing.

The following process guidelines are provided by the Mediation Service to help mediators:

    Hear each party without interruption look at past and present, no loss of opportunity for both parties to discuss the issues and explore a settlement

    Cross table discussion and an opportunity for clarfication for both parties

    Exploration of key points/issues, where possible identify key points and get future focussed

    Private session if required to go over information(past) that has arisen and a time for reflection(reality test)

    starting the development of options (future focus) and also a plan for joint session (only you will know if something you heard was different or news to you)

    Joint session(if appropriate) on generating options, negotiating, exploring, etc

    Another private session to check agreement, writing up settlement

    Closure

What's the job worth?

1)   **The outcome of the 24 April meeting was?**   *(check one ✓, or %)*

a. Management agree to a new evaluation being conducted before holding further discussions on salary bands [__]

b. Staff accept the salary bands proposed by management on 14 February with few or any modifications   [__]

c. Parties agree to ask a third party (e.g. mediator, independent job-evaluator) to decide on salary bands   [__]

d. Parties fail to reach any agreement   [__]

2)   **Describe how you derived your prediction or, if you have *not* given a prediction, state your reasons:**

3)   **Roughly, how long did you spend on this task?**
{include the time you spent reading the description and instructions}   [____] hours

4)   **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

5)   **Do you recognise the actual conflict described in this file?**   Yes [__]   No [__]

     If so, please identify it: [_____]

6)   **How many people did you discuss this forecasting problem with?**   [____] people

7)   **Roughly, how many years have you spent as a game theory practitioner or researcher?**
     [___] years

8)   **Please rate your experience (out of 10) with conflicts <u>similar to this one</u>**   [____] 0-10

When you have completed this questionnaire, please return
<u>either</u> this document as an email attachment to kesten.green@vuw.ac.nz
<u>or</u> this questionnaire (with your initials at right) by fax to (64 4) 499 2080.   **Your initials:** [_____]

Telco Takeover Bid

In mid-August 2001, the Localville Telco board rejected an unsolicited $6 billion offer for the company from rival telecommunications operator, Expander Telco. The companies are relatively small but profitable players in the huge US market. Both companies specialize in providing services to rural areas that have so far avoided the attentions of the giants. Localville's rejection of the offer is not the end of the matter. Expander's $43-per-share offer is more than 40% above Localville's pre-bid share price of around $30.

Localville grew from small beginnings under the care of current Chairman, Augustus Lovett, who has headed the company for several decades and remains a substantial shareholder. The company's managers and employees (including Chairman Lovett) between them own 30% of Localville's voting stock and, of the 14 board members, five are current or former employees. Localville is based in a small rural town and many of its 7,000 employees live in the area and take pride in participating in community affairs. Chairman Lovett has served as mayor and has sat on local community boards over the years. A takeover by Expander would lead to major layoffs in a community that has benefited from several generations of employment at Localville.

Expander's annual turnover of $7.5 billion is three times that of Localville at $2.5 billion. Although the companies both offer local (landline or wireline) and mobile (wireless or cell-phone) phone services, their service mixes are quite different. Expander has 6.7 million mobile and 2.6 million landline customers, while Localville has 0.8 million mobile and 1.8 million landline customers. Through an arrangement with one of the industry giants, Expander provides national roaming (mobile customers can make and receive calls from anywhere in the US) and can provide flat-rate national calling to mobile customers at cost. Expander has mobile penetration averaging 13% across its territories, and achieves a 61% margin on its mobile operations. Localville does not have a national roaming arrangement, its penetration is less than 10%, its margin is 54%, and its mobile revenues are flat — falling nearly 1.5% in the second quarter. In fact, shortly before Expander made its offer to buy Localville, Localville had approached Expander to assess the rival company's interest in purchasing Localville's mobile operation. Expander rejected the proposition, as company policy is to offer a comprehensive service, including local calls, when moving into a new territory.

Expander's father and son team, Chairman Al Exley and CEO Brad Exley, have acquired more than 250 companies over the last 15 years, and have spent $12 billion on acquisitions during the last three. Localville's footprint (the geographical spread of its customers and potential customers) complements Expander's. Expander's analysis suggests that economies of scale resulting from the purchase of Localville should amount to $100 million (1%) on combined revenue of $10 billion. There would be further gains from lifting the financial performance of Localville operations to the levels achieved by the current Expander operations. As well as these advantages, in the longer term the enlarged Expander may present an attractive acquisition for a major operator willing to pay a premium in order to gain a regional customer base.

Localville has told analysts that the company doesn't need to sell-out to Expander in order to provide a good return to shareholders. Localville maintains that revenue growth from the company's landline business will exceed that of Expander. Nevertheless, the prospects for the mobile side of Localville's business — responsible for roughly 20% of revenue — concerns shareholders. Without national roaming agreements, Localville can't offer the cheap flat-rate national calling deals that customers are increasingly demanding. Localville's strategy of selling the mobile operation and expanding the business as a landline operator is considered by some analysts to offer the best deal for shareholders. The Localville board are adamant that the company is happy to talk to Expander only about selling its mobile operation. It is unlikely that the board could be forced to change its stance by non-employee shareholders who are keen to accept Expander's offer: Localville doesn't have to hold a shareholder meeting for another three months and company rules make it impossible to oust the whole board at one time. Although it is possible under US law to bypass the company board and make an offer directly to shareholders, the process is restricted by regulation and, even if Expander acquired more than 50% of Localville stock, the current board could remain in effective control of the company for some time. Localville released the following statement on mobile services: "...divestiture efforts have been adversely affected by a hostile takeover attempt... and a recent sharp decline in the general mobile market... we continue to believe that divestiture makes strategic sense and we continue to pursue that goal. Nevertheless, mobile generates strong cash flows for Localville and so we do not feel compelled to divest until we are presented with the right offer".

It is now September, 2001. Despite Localville's rejection, Expander hasn't given up on buying the Localville company. Expander knew that the Localville board was not interested in selling the whole company when it first made its bid. Institutional investors and other non-employee shareholders of Localville are not happy with the board's rejection of Expander's offer and this has put pressure on Localville management to find a credible solution to the companies current performance woes. The stand-off between Expander and Localville must be resolved. In which one of the following four ways will this happen?

1. Expander's takeover bid fails completely
2. Expander purchases Localville's mobile operation only
3. Expander's takeover succeeds at, or close to, their August 14 offer price of $43-per-share
4. Expander's takeover succeeds at a substantial premium over the August 14 offer price

Role of Localville Telco Chairman – Augustus Lovett

Several decades ago you became Chairman of Localville, a small telephone company founded by your grandfather. Since then you have built the company into a $US 2.5 billion business. Despite the substantial size of the business, the company head office and many employees remain in the small town where you live and where the company had its beginnings.

Although Localville offers both local and mobile telephone services, mobile users have come to expect to be able to use their phones anywhere in the US for the same (cheap) rate. Localville has not been able to obtain a roaming agreement with one of the major telecommunications companies that would allow the company to offer this service. Recently, Localville approached Expander Telco to explore the possibility of selling Localville's mobile business to Expander. Expander rejected this idea, and offered to buy the whole of Localville. You and the board do not wish to sell the whole business as you believe Localville can do better for shareholders as a provider of local telephone services and, at the same time, protect your employees and the local community. Not all shareholders agree, however, and you and your CEO, Bill Lowe, have a tough fight on your hands to keep Localville safe from Expander's hostile takeover bid.

Role of Localville Telco CEO – Bill Lowe

Localville is small (by US standards) telephone company with a turnover of $US 2.5 billion. Localville offers both local and mobile telephone services, but mobile users have come to expect to be able to use their phones anywhere in the US for the same (cheap) rate. Localville has not been able to obtain a roaming agreement with one of the major telecommunications companies that would allow the company to offer this service. Recently, Localville approached Expander Telco to explore the possibility of selling Localville's mobile business to Expander. Expander rejected this idea, and instead offered to buy the whole of Localville. The Localville board does not wish to sell the whole business as the directors believe Localville can do better for shareholders as a provider of local telephone services and, at the same time, protect Localville employees and the local community. Not all shareholders agree, however, and you and your Chairman, Augustus Lovett, have a tough fight on your hands to keep Localville safe from Expander's hostile takeover bid. In common with many employees, you own shares in Localville.

Role of Expander Telco Chairman – Al Exley

Under you leadership for the past 15 years, telephone company Expander has grown rapidly with the acquisition of more than 250 companies. Recently, Expander was approached by Localville Telco. Localville wished to explore the possibility of selling its mobile phone business to Expander. You rejected the proposition, as Expander's policy is to be able to offer local telephone services as well as mobile when it moves into a new territory. Localville as a whole, however, would be an attractive acquisition (the companies' territories are complementary) and Expander has offered to buy Localville at a price-per-share that was 40% higher than the price prevailing at the time of the offer. So far, the Localville board has rejected this offer, but you and your son (Expander CEO, Brad Exley) believe Localville is a prize worth fighting for.

Role of Expander Telco CEO – Brad Exley

Under the leadership of your father (Chairman Al Exley) for the past 15 years, telephone company Expander has grown rapidly with the acquisition of more than 250 companies. Recently, Expander was approached by Localville Telco. Localville wished to explore the possibility of selling its mobile phone business to Expander. Expander rejected the proposition, as company policy is to be able to offer local telephone services as well as mobile when it moves into a new territory. Localville as a whole, however, would be an attractive acquisition (the companies' territories are complementary) and Expander has offered to buy Localville at a price-per-share that was 40% higher than the price prevailing at the time of the offer. So far, the Localville board has rejected this offer, but you and your father believe Localville is a prize worth fighting for.

Telco Takeover Bid

1) **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   [__]

2) **Describe how you derived your prediction or, if you have *not* given a prediction, state your reasons:**

3) **Roughly, how long did you spend on this task?**
   {include the time you spent reading the description and instructions}   [_____] hours

4) **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

5) **Do you recognise the actual conflict described in this file?**   Yes [__]   No [__]

   If so, please identify it: [_____]

6) **How many people did you discuss this forecasting problem with?**   [_____] people

7) **Roughly, how many years have you spent as a game theory practitioner or researcher?**
                                                                                  [___] years

8) **Please rate your experience (out of 10) with conflicts similar to this one**   [_____] 0-10

When you have completed this questionnaire, please return
either this document as an email attachment to kesten.green@vuw.ac.nz
or this questionnaire (with your initials at right) by fax to (64 4) 499 2080.   **Your initials:** [_____]

International Water Dispute

Today is June 3, 1975. Two poor and arid Asian countries, Midistan and Deltaland, are in dispute over access to the waters of the River Fluvium. The river rises in Uplandia, whose plentiful rain contributes at least 90% of the flow. It then runs through Midistan — where the scanty rainfall makes up the rest of the flow — and then on through Deltaland to the sea. Relations between the two disputants have deteriorated badly, and the Government of Concordia has stepped in, in an attempt to mediate an agreement. Uplandia is not involved in this dispute.

Background

Both Midistan and Deltaland depend heavily on Fluvium water for irrigation. Midistan also uses the river for generating electricity. Deltaland has exploited the waters of the Fluvium since ancient times. Uplandia and Midistan, on the other hand, started to make substantial use of the river's water only about ten years ago. Eighteen months ago, Uplandia began filling its new dam at Updama. A few months later the new Soviet-constructed dam at Mididam in Midistan became operational.

Midistan and Deltaland are ruled by leaders who came to power after military coups. They are loosely aligned to the Soviet Union. Their armed forces are similar in size — both are large and battle-hardened.

Recent developments

On April 7, Deltaland accused Midistan of putting at risk the lives of the three million Deltaland farmers dependent on the water of the Fluvium by diverting excessive volumes from the river. The Deltaland News Agency reported the protest came "as a result of the lack of response by the Midistani Government to all efforts exerted by the Deltaland Government for years to reach an agreement...". Two days later, the Deltaland Government issued a statement saying it would take whatever steps were necessary to ensure access to the waters of the Fluvium and would hold Midistan responsible for any harm to Deltalandish farmers. A congress of Midistani political leaders, on the same day, condemned the Deltaland regime for plotting with enemies of Midistan and betraying the common heritage of the two countries. There were reports that 200 military and civilian leaders had been arrested, in the lead-up to the conference, on charges of plotting against the Midistani Government. Those arrested included the director of a news agency and a former director of Midistani television.

The Midistani Government explained their position on the disagreement over access to Fluvium water in an official statement released on April 19. In the statement, Midistan blamed the current crisis on the Deltaland Government's unwillingness to enter in good faith into tripartite negotiations with Midistan and Uplandia for a permanent agreement over sharing the water. Deltaland instead had conducted secret negotiations with Uplandia. Midistan claimed to have reached provisional agreement with Deltaland two years before for the flow of water during the winter just gone, but had stipulated that the agreed volume would have to be revised when the Updama dam began to fill. Midistan accused Deltaland of avoiding negotiations over this issue when Uplandia had commenced filling the Updama dam in January of last year.

The statement also claimed that, despite substantial reductions in, and interruptions of, the flow of water out of Uplandia, Midistan had allowed 70% of the water received to flow on into Deltaland, had released an additional 200 million m^3 (0.7% of the usual annual inflow) during the middle of last year in response to a request by Deltaland and, during last winter had let 75% of water from Uplandia flow on to Deltaland. Further, Midistan accused Deltaland of failing to modernise its irrigation methods in order to make more effective use of the water it does receive.

In response, Deltaland maintained its claims that more water than was required for electricity generation had been withheld and that only half of the water to which Deltaland was entitled had been received.

Claims and counter-claims by Midistan and Deltaland continued through April and May, as did mediation efforts by neighbouring countries including the wealthy regional leader, Concordia. Midistan accused Deltaland of assassinations and mass executions of dissidents on May 7 and, a week later, closed its airspace to Deltalandish aircraft in response to mistreatment of Midistani airline personnel employed in Deltaland. On May 25, Midistan ordered the immediate closure of one of Deltaland's consulates in Midistan. On May 28 a Midistani military official in Deltaland was stabbed and, on May 29, Midistan accused the Deltalandish government of executing 80 government opponents. By June 2, there were reports that both sides had moved troops to the border between the countries and that Deltaland had threatened to bomb the Mididam dam. In response to the deteriorating situation, Concordia renewed its efforts at mediation and a meeting between ministers from the three countries is to be held.

The meeting

Today, Government ministers and officials from the Midistan, Deltaland, and Concordia will meet to try to resolve the dispute. Those present at the meeting will be a senior Minister from the Kingdom of Concordia, and the Foreign Ministers of the Republics of Midistan and Deltaland each accompanied by a military adviser. A statement will be issued at the end of the meeting. The statement may be one of three alternatives. The gist of these statements are as follows:

(a) Midistan has decided to release additional water in order to meet the needs of the Deltalandish people
(b) Deltaland has ordered the bombing of the dam at Mididam to release water for the needy Deltalandish people
(c) Deltaland has declared war on Midistan.

Role of Republic of Midistan Foreign Minister – Mohammad Fareed

A crisis over access to water is brewing between the poor Asian nation, Midistan, and the neighbouring Republic of Deltaland. As Midistan's Foreign Minister, you are attending a meeting with your Deltaland counterpart (Daud Fawaz) and a senior Minister from the Kingdom of Concordia – a wealthy regional power. The Minister from Concordia (Karim Khalid) will attempt to mediate an agreement. The crises has already led to military preparations by both sides. Before joining the meeting you will discuss objectives and strategy with the military adviser who has accompanied you – General Mustafa Ahmad.

As far as you and your government are concerned, Midistan has acted responsibly in a difficult situation that is not of Midistan's making. You are directly responsible to your President who, along with you and the rest of the Midistan government, came to power after a military coup.

Role of Republic of Midistan Military Adviser – General Mustafa Ahmad

A crisis over access to water is brewing between the poor Asian nation, Midistan, and the neighbouring Republic of Deltaland. You are attending a meeting with Deltaland government representatives and a senior Minister from the Kingdom of Concordia – a wealthy regional power. The Minister from Concordia (Karim Khalid) will attempt to mediate an agreement. The crises has already led to military preparations by both sides – troops have been moved to the vicinity of the common border. Your role is to support and advise your Foreign Minister (Mohammad Fareed) with whom you will discuss objectives and strategy before you both join the meeting.

As far as you and your government are concerned, Midistan has acted responsibly in a difficult situation that is not of Midistan's making. You are directly responsible to your Foreign Minister who, along with you and the rest of the Midistan government, came to power after a military coup.

Role of Republic of Deltaland Foreign Minister – Daud Fawaz

A crisis over access to water is brewing between the poor Asian nation, Deltaland, and the neighbouring Republic of Midistan. Deltaland has a long history of using the waters of the River Fluvium and is heavily dependent on the river for agriculture and drinking water. Midistan has recently built a large dam and the Fluvium's flow into Deltaland has been curtailed. As Deltaland's Foreign Minister, you are attending a meeting with your Midistan counterpart (Mohammad Fareed) and a senior Minister from the Kingdom of Concordia – a wealthy regional power. The Minister from Concordia (Karim Khalid) will attempt to mediate an agreement. The crises has already led to military preparations by both sides. Before joining the meeting you will discuss objectives and strategy with the military adviser (General Dirwar Ali) who has accompanied you.

You are directly responsible to your President who, along with you and the rest of the Deltaland government, came to power after a military coup.

Role of Republic of Deltaland Military Adviser – General Dirwar Ali

A crisis over access to water is brewing between the poor Asian nation, Deltaland, and the neighbouring Republic of Midistan. Deltaland has a long history of using the waters of the River Fluvium and is heavily dependent on the river for agriculture and drinking water. Midistan has recently built a large dam and the Fluvium's flow into Deltaland has been curtailed. You are attending a meeting with Midistan government representatives and a senior Minister from the Kingdom of Concordia – a wealthy regional power. The Minister from Concordia (Karim Khalid) will attempt to mediate an agreement. The crises has already led to military preparations by both sides – troops have been moved to the vicinity of the common border. Your role is to support and advise your Foreign Minister (Daud Fawaz) with whom you will discuss objectives and strategy before you both join the meeting.

You are directly responsible to your Foreign Minister who, along with you and the rest of the Deltaland government, came to power after a military coup.

Role of Kingdom of Concordia Senior Minister – Karim Khalid

Two poor neighbours of the wealthy Kingdom of Concordia appear to be edging closer to war in a dispute over access to the waters of a river that flows from one (Midistan) to the other (Deltaland). Your Kingdom has a traditionally played a paternal role in the region, and has an interest in preserving peace. To that end, you have called a meeting, which is being attended by Midistan and Deltaland representatives, with the hope that your mediation will lead to a peaceful solution to the crisis. Midistan is represented by Foreign Minister Mohammad Fareed and Military Adviser General Mustafa Ahmad. Deltaland is represented by Foreign Minister Daud Fawaz and Military Adviser General Dirwar Ali.

International Water Dispute

1)   **The gist of the statement issued at the end of the meeting was?**   *(check one ✓, or %)*
a. Midistan has decided to release additional water in order to meet the needs of the Deltalandish people   [___]
b. Deltaland has ordered the bombing of the dam at Mididam to release water for the needy Deltalandish people [___]
c. Deltaland has declared war on Midistan   [___]

2)   **Describe how you derived your prediction or, if you have *not* given a prediction, state your reasons:**

3)   **Roughly, how long did you spend on this task?**
              {include the time you spent reading the description and instructions}   [_____] hours

4)   **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

5)   **Do you recognise the actual conflict described in this file?**   Yes [__]   No [__]

       If so, please identify it: [_____]

6)   **How many people did you discuss this forecasting problem with?**   [_____] people

7)   **Roughly, how many years have you spent as a game theory practitioner or researcher?**
                                                                                        [___] years

8)   **Please rate your experience (out of 10) with conflicts _similar to this one_**   [_____] 0-10

When you have completed this questionnaire, please return
either this document as an email attachment to kesten.green@vuw.ac.nz
or this questionnaire (with your initials at right) by fax to (64 4) 499 2080.   **Your initials:** [_____]

ZENITH INDUSTRIAL CORPORATION INVESTMENT DECISION

Zenith is a large manufacturing business based in a populous developed country. Zenith manufactures "product".

The year is 1975, and Zenith has recently been nationalised by a new left-leaning government. Under its previous owners, Zenith had begun to lag behind international competitors as a consequence of disruptions caused by a supplier's ongoing industrial conflicts and the Zenith owners' reluctance to invest in the business. The new government regard Zenith as important because of its size and the widespread use of product by other manufacturers. Unfortunately for the Government, Zenith must reduce costs in order to regain international competitiveness C the corporation is currently suffering losses of $US 10 million per week. A $US 10 billion dollar redevelopment project is underway that will see older plants retired and new plants commissioned. Efficiency gains as a result of the project will result in a net loss of 50,000 jobs, and will have a major impact on a region (Naggaland) that traditionally provides much support for the party that is now in government. Unsurprisingly, the government have made political commitments to Naggaland.

Last year (1974) international demand for product was high and Zenith was not able to keep pace with demand. In that year, the chair of Zenith (Sir Archie Stevenson) had discussions with the owner of Acma, a foreign manufacturer of new technology product-making plant. As an outcome of those discussions, the Zenith board gave approval, in principle, for the purchase of one plant from Acma. The price quoted by Acma was $US 26 million. They have subsequently quoted $US 43 million for two plants. The Acma offers remain current until February 15, 1975.

The Acma plant, or plants, would offer some advantages to Zenith. Firstly, the Acma plant uses a different combination of inputs and would therefore somewhat reduce Zenith's vulnerability to industrial action by a traditional supplier's union. Secondly, the plant would give Zenith experience with a new technology and, thirdly, the plant could be located in Naggaland. Sir Archie was appointed by the Government with the understanding that he would be sympathetic to the Government's political commitment to Naggaland. Sir Archie has made it known that he prefers the two-plant option.

Since the board's original agreement in principle with Acma, the world economy has slid into recession and the demand for product has fallen. Sir Archie has made it known that he wants a $US 50 million reduction in capital spending. With the February 15 deadline looming, and negotiations with Acma underway, Zenith's Policy Committee asks the company's planners to assess the merits of the Acma proposals.

On 5 February the planners came back with a surprising assessment. The production costs of product from the Acma plants would be higher than the production costs from conventional technology plants. The existing technology produces product for $US 112 per unit, whereas a single Acma plant would produce product for $US 126 per unit and two Acma plants would produce product for $US 116 per unit. Not only that, the planners' demand projections indicated that currently planned production capacity would be sufficient to meet demand in all but the first 2 years of the 6 year forecast period. The planners pointed out that this projected slight shortfall in production could be met by delaying the retirement of a number of older plants while their capacity was still required. Sacrificing the Acma plant would also appear to be a sensible way of meeting the Chairman's expressed desire to cut capital expenditure by $US 50 million. Sir Archie has asked Herbert Lumley (Director of Planning and Capital Development) to present the planners' findings to the Policy Committee on Tuesday 11 February.

The 11 February meeting is to be chaired by Sir Archie. Herbert Lumley will present the planners' findings that are outlined above. The purpose of the meeting is to decide whether to commission one Acma plant, two Acma plants, or not to commission any Acma plants. The Committee must reach a decision on this question at the meeting. The Committee's recommendation (one, two, or zero Acma plants) will almost certainly be accepted by the Board C most Board members are also on the Committee.

Zenith Policy Committee

Sir Archie Stevenson (Chairman of Zenith Board and Policy Committee)

Mark Stepman QC (Deputy Chairman of Zenith Board)

Robert Revell (Company secretary)

James Drywall (Corporate strategy)

Lord Gratton (Executive Director)

Lionel Hunt (Finance)

Frank Holdall (Supplies & Transport)

Herbert Lumley (Planning & Capital Development Director)

John Grove-White (Chief Planner)

Ron Able (Capital Projects Manager & Chief Negotiator)

Sir Archie Stevenson
(Chairman of Zenith Board and Chairman of Policy Committee)

You were appointed by the Government with the expectation that you would oversee the transformation of Zenith in such a way as to minimise any damage to the Government's reputation, particularly among its natural constituencies. The Government know that their Naggaland constituents will be badly affected by plant closures and staff layoffs by Zenith. You hail from Naggaland yourself, and retain interests there.

You are not averse to using your authority to take control of a meeting. At the same time, as chairman, you must be prepared to let all parties have their say, and prefer to reach a consensus decision. Your responsibilities as the Chairman of the Board of this very large industrial company, are considerable. It is you that must be able to see the "bigger picture". You must balance the demands of the various parties that have an interest in a decision that is to be made, and at the same time retain the support of your fellow Committee members and the Board.

You are kept well-informed about developments in the company as they occur C this gives you something of an advantage over others in the company, as their information is often incomplete.

Mark Stepman QC (Zenith Deputy Chairman)

As a senior officer of Zenith, you make many decisions based on material that is presented to you by other officers of the company. You must use your judgment in weighting the evidence that is provided and in incorporating your view of the wider aims of the company.

Robert Revell (Zenith Company Secretary)

As a senior officer of Zenith, you make many decisions based on material that is presented to you by other officers of the company. You must use your judgment in weighting the evidence that is provided and in incorporating your view of the wider aims of the company.

James Drywall (Zenith Managing Director for Corporate Strategy)

As a senior officer of Zenith, you make many decisions based on material that is presented to you by other officers of the company. You must use your judgment in weighting the evidence that is provided and in incorporating your view of the wider aims of the company.

Lord Gratton
(Zenith Executive Director)

As a senior officer of Zenith, you make many decisions based on material that is presented to you by other officers of the company. You must use your judgment in weighting the evidence that is provided and in incorporating your view of the wider aims of the company.

Lionel Hunt
(Zenith Finance Director & member of Zenith Board)

As Zenith Industrial Corporation's Finance Director it is your responsibility to insure that the company is run a way that is financially sound. You must ensure that new capital investment can be expected to deliver a satisfactory return, either singly or as part of a diversified portfolio. You expect to be presented with a well-argued case for any new expenditure, and are not readily impressed by unsupported rhetoric.

Herbert Lumley
(Zenith Planning & Capital Development Director)

As the Director of Planning and Capital Development for Zenith Industrial Corporation, both the Chief Planner, John Grove-White, and the Manager of Capital Projects, Ron Able, report to you. Ron also has the role of chief negotiator for the purchase of some major new plant for Zenith. Ron has been lobbying you hard to gain your support for the purchase, and emphasises that the company's Chairman, Sir Archie Stevenson, is keen for the purchase to go ahead.

The Policy Committee must meet to decide whether or not the plant purchase should go ahead, and have asked you to get your planning team, headed by John Grove-White, to assess the merits of the proposal. You expect John and his team to conduct a rigorous and impartial assessment. You are expected to present their findings to the Committee. Both John, and Ron Able will be with you at the meeting.

John Grove-White
(Chief Planner)

Your manager, Herbert Lumley, asked you and your team to assess the merits of a plan to purchase some major new plant. You are aware that there has already been a commitment of sorts to the new plant but, as a matter of personal and professional pride, you and your team will endeavour to conduct a rigorous assessment of the plan. Consequently, you expect your findings to be taken seriously.

Although you might consider yourself a cut above your colleagues, both intellectually and socially, you are decidedly a junior participant in the situation that is unfolding. You are an ambitious young man and, as such, are conscious that your decisions can influence your prospects.

Ron Able
(Capital Projects Manager & Chief Negotiator)

You are the chief negotiator for the purchase of some major new plant for your company. This puts you in a powerful position, as you control the information about the progress of the negotiation that is passed-on to your colleagues. As you are also Capital Projects Manager, you have a vested interest in the purchase going ahead – the more capital investment, the greater your responsibilities. You believe that the Chairman is strongly supportive of the purchase.

You are a WWII veteran, and a practical no-nonsense sort of man.

ZENITH INDUSTRIAL CORPORATION INVESTMENT DECISION

1)   **Which option will the Zenith Policy Committee choose?**

(check one - ✓)

(A)   One ACMA plant   [ ]

(B)   Two ACMA plants   [ ]

(C)   No ACMA plants   [ ]

2)   **Broadly, what approach did you use to derive your prediction?**

3)   **Roughly, how long did you spend on the task of deriving a prediction for this situation?**

[_____] hours

4)   **If you have not provided a prediction, please state your reasons:**

5)   **Roughly, how many years have you spent as a game theory practitioner or researcher?**

[_____] years

When you have completed this questionnaire, please return either
this document as an email attachment to
kesten.green@vuw.ac.nz
or this page (with your initials printed below) by fax to
(64 4) 499 2080.

Your initials: [_____]

Appendix 3: Zenith Investment questionnaires provided to participants: Unaided judgement (novice, expert), structured analogies (expert), and simulated interaction(novice)

[_____]

ZENITH INDUSTRIAL CORPORATION INVESTMENT DECISION

1) **Which option will the Zenith Policy Committee choose?**   *(please check one - ✓)*
   - (A)   One ACMA plant   [ ]
   - (B)   Two ACMA plants   [ ]
   - (C)   No ACMA plants   [ ]

2) **Broadly, what approach did you use to derive your prediction?**

3) **Roughly, how long did you spend on the task of deriving a prediction for this situation?**

[____] minutes

4) **If you have not provided a prediction, please state your reasons:**

For the next 3 questions, please rate the situation & roles using the 11-point scales provided.

5) **Overall, role descriptions were:**
   Indistinct   0-1-2-3-4-5-6-7-8-9-10   **Clear**   [___]

6) **Initially, the opposing parties took stances that were:**
   Incompatible   0-1-2-3-4-5-6-7-8-9-10 **Congruent**   [___]

7) **Overall, the characters seemed likely to take stances that would be:**
   Rigid 0-1-2-3-4-5-6-7-8-9-10 **Flexible**   [___]

8) **Please indicate whether you recognised the situation, and provide the identity of the situation where this is appropriate.**

   i. Recognise: [___] Yes   [___] No

   ii. Situation identity: _____

	Strongly Agree		Not Sure		Strongly Disagree

9. I've never experienced an actual situation like the one described in this material [ ]   [ ]   [ ]   [ ]   [ ]

10. I was born in 19[____]   11. I am   male [__]   female [__]
12. My principal occupation is [_____] {eg. Student, salesperson, etc}
13. If a student, major subjects [_____] {eg. ENGL, LAWS, MGMT, etc}
14. The industry I work in is   [_____] {eg. Manufacturing, retailing, etc.}
15. The sector I work in is   [_____] {eg. Private, central government, etc}
16. My major qualification is   [_____] {eg. School Cert, BA, etc}

17. Please record your OBSERVATIONS on the situation described in this material and the roles of the people involved, here and on the back of this page.

When you have completed this questionnaire,
please fold this side in and return it to the person supervising the session.
Please do not discuss this exercise with anyone until after all questionnaires have been returned.
Thank you for your help with this research.

ZENITH INDUSTRIAL CORPORATION INVESTMENT DECISION

1)   **Which option will the Zenith Policy Committee choose?**

(check one ✓ , or %)

     (A)   One ACMA plant   [ ]

     (B)   Two ACMA plants   [ ]

     (C)   No ACMA plants   [ ]

2)   **Broadly, what approach did you use to derive your prediction?**

3)   **If you have *not* given a prediction, please state your reasons:**

4)   **Roughly, how long did you spend on this task?**
{include the time you spent reading the description and instructions}

          [_____] hours

5)   **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

6)   **Do you recognise the actual conflict described in this file?**   Yes [__] No [__]

     If so, please identify it:

     [_____]

7)   **How many people did you discuss this forecasting problem with?**   [_____]
people

8)   **Roughly, how many years experience do you have as a conflict management specialist?**   [_____] years

9)   **Please rate your experience (out of 10) with conflicts <u>similar to this one</u>**
          [_____] 0-10

ZENITH INDUSTRIAL CORPORATION INVESTMENT DECISION

1) **(A) In the table below, please briefly describe**
  (i) your analogies,
  (ii) their source (e.g. your own experience, media reports, history, literature , etc.), **and**
  (iii) the main similarities and differences between your analogies and this situation.
  (B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).
  (C) Enter the responses from question 2 (below) closest to the outcomes of your **analogies.**

(A) (i) description, (ii) source, (iii) similarities & differences	(B) Rate	(C) Q2
a.		
b.		
c.		
d.		
e.		

2) **Which option will the Zenith Policy Committee choose?**

(check one ✓ , or %)

  (A)   One ACMA plant   [ ]

  (B)   Two ACMA plants   [ ]

  (C)   No ACMA plants   [ ]

3) **If you have *not* given a prediction, please state your reasons:**

4) **Roughly, how long did you spend on this task?**
{include the time you spent reading the description and instructions}
[_____] hours

5) **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

6) **Do you recognise the actual conflict described in this file?**   Yes [__] No [__]

If so, please identify it:

[_____]

7) **How many people did you discuss this forecasting problem with**
[_____] people

8) **Roughly, how many years experience do you have as a conflict management**
specialist?   [_____] years

9) **Please rate your experience (out of 10) with conflicts <u>similar to this one</u>**
[_____] 0-10

When you have completed this questionnaire, please return
<u>either</u> this document as an email attachment to kesten.green@vuw.ac.nz
  <u>or</u> this questionnaire (with your initials at right) by fax to (64 4) 499 2080.   **Your initials:** [_____]

195

Sir Archie Stevenson
(Chairman of Zenith Board and Chairman of Policy Committee)

You were appointed by the Government with the expectation that you would oversee the transformation of Zenith in such a way as to minimise any damage to the Government's reputation, particularly among its natural constituencies. The Government know that their Naggaland constituents will be badly affected by plant closures and staff layoffs by Zenith. You hail from Naggaland yourself, and retain interests there.

You are not averse to using your authority to take control of a meeting. At the same time, as chairman, you must be prepared to let all parties have their say, and prefer to reach a consensus decision. Your responsibilities as the Chairman of the Board of this very large industrial company, are considerable. It is you that must be able to see the "bigger picture". You must balance the demands of the various parties that have an interest in a decision that is to be made, and at the same time retain the support of your fellow Committee members and the Board.

You are kept well-informed about developments in the company as they occur—this gives you something of an advantage over others in the company, as their information is often incomplete.

Zenith Policy Committee: See question 17, overleaf, for a list of Committee members
_____

Please answer the next 2 questions <u>at the end of the role-play</u> but while "in-character".
1. My judgment is that Zenith should choose: 0 [   ]   1 [   ]   2 [   ] Acma plants
2. The Policy Committee decided to recommend: 0 [   ]   1 [   ]   2 [   ] Acma plants

3. I was born in 19[_____]   **4.** I am   male [___]   female [___]

5. My principal occupation is [_____] {*eg. Student, salesperson, etc*}

6. If a student, degree &
       major subjects   [_____] {*eg. BA in ENGL, BCA in ACCY, etc*}

7. The industry I work(ed) in [_____] {*eg. Agriculture forestry fishing,*
       Mining & quarrying, Manufacturing, Electricity gas & water, Construction,
       Wholesale & retail trade restaurants & hotels, Transport storage & communication,
       Business & financial services, community social & personal services.}

8. The sector I work in is   [_____] {*eg. Private enterprise, Central*
       government-trading, Local government-trading, Producer boards, Financial intermediaries
       (banks, insurance, etc), Central government non-trading, Local government non-trading.}

9. My major qualification is [_____] {*eg. School Cert, BA, etc*}

10. Please record your OBSERVATIONS on the situation, the roles of the people involved, and the way a decision was arrived at.

☛**PTO…**

Please answer the next questions about your experience with the role-play.	Strongly Agree		Not Sure		Strongly Disagree
11. I'm probably quite similar to the person whose character I played	[]	[]	[]	[]	[]
12. I didn't find the other role-players very convincing in their roles	[]	[]	[]	[]	[]
13. I've never experienced an actual situation like the one we role-played	[]	[]	[]	[]	[]
14. My role coincided with my personal beliefs and attitudes	[]	[]	[]	[]	[]
15. Overall our role-play seemed realistic	[]	[]	[]	[]	[]
16. The role-players seemed to care about the outcome	[]	[]	[]	[]	[]

17. For each of the policy committee members listed below, please indicate (using the scale provided) to what extent you agree that the person playing the member was CONVINCING IN THE ROLE

a. Sir Archie Stevenson (Chairman of Zenith Board and Policy Committee)	[]	[]	[]	[]	[]
b. Mark Stepman QC (Deputy Chairman of Zenith Board)	[]	[]	[]	[]	[]
c. Robert Revell (Company secretary)	[]	[]	[]	[]	[]
d. James Drywall (Corporate strategy)	[]	[]	[]	[]	[]
e. Lord Gratton (Executive Director)	[]	[]	[]	[]	[]
f. Lionel Hunt (Finance)	[]	[]	[]	[]	[]
g. Frank Holdall (Supplies & Transport)	[]	[]	[]	[]	[]
h. Herbert Lumley (Planning & Capital Development Director)	[]	[]	[]	[]	[]
i. John Grove-White (Chief Planner)	[]	[]	[]	[]	[]
j. Ron Able (Capital Projects Manager & Chief Negotiator)	[]	[]	[]	[]	[]

18. Before this game, I already knew the person who was playing…

a.	[]	[]	[]	[]	[]
b. Mark Stepman QC	[]	[]	[]	[]	[]
c. Robert Revell	[]	[]	[]	[]	[]
d. James Drywall	[]	[]	[]	[]	[]
e. Lord Gratton	[]	[]	[]	[]	[]
f. Lionel Hunt	[]	[]	[]	[]	[]
g. Frank Holdall	[]	[]	[]	[]	[]
h. Herbert Lumley	[]	[]	[]	[]	[]
i. John Grove-White	[]	[]	[]	[]	[]
j. Ron Able	[]	[]	[]	[]	[]

19. My real-world relationship with the person playing…

a.	[]	[]	[]	[]	[]
b. Mark Stepman QC is similar to the relationship between the characters	[]	[]	[]	[]	[]
c. Robert Revell is similar to the relationship between the characters	[]	[]	[]	[]	[]
d. James Drywall is similar to the relationship between the characters	[]	[]	[]	[]	[]
e. Lord Gratton is similar to the relationship between the characters	[]	[]	[]	[]	[]
f. Lionel Hunt is similar to the relationship between the characters	[]	[]	[]	[]	[]
g. Frank Holdall is similar to the relationship between the characters	[]	[]	[]	[]	[]
h. Herbert Lumley is similar to the relationship between the characters	[]	[]	[]	[]	[]
i. John Grove-White is similar to the relationship between the characters	[]	[]	[]	[]	[]
j. Ron Able is similar to the relationship between the characters	[]	[]	[]	[]	[]

When you have completed this questionnaire,
please put it in "Sir Archie's" envelope – "Sir Archie" will give the envelope to the supervisor of the session.
Please do not discuss this exercise with anyone until after all questionnaires have been returned.
Thank you for your help with this research.

Appendix 4: Information Sheet and Informed Consent form

INFORMATION SHEET

Simulation Games for Forecasting

Background and purpose

Simulation games have for many years been used as method for predicting and understanding human behaviour. Despite this, there is no unified literature on simulation games and many questions about the influences on the predictive accuracy of such games remain unanswered. For example, much research on simulation games has used student subjects, but no direct comparisons have been made between the forecasting accuracy of simulations using students and that of simulations using other people. The purpose of my research is to compare the forecasting accuracy of simulation games using people with different backgrounds and experience.

What you will be asked to do

Before starting your game, you will be allocated to a small group of people who will take part in the same game as you. You will be given the description of one character in the game and his or her role. When you have read about your character you will be expected to behave as you believe your character would behave until your simulation is over.

When you are "in-character", you will be given a description of the situation your character is confronted with. The person supervising your simulation will tell you what to do next. You will be expected to continue with your simulation game until an outcome is achieved, a time-limit is reached, or you are asked to stop.

At the end of your simulation you will be asked to provide information about the outcome, and to comment on this. You may also be asked to provide some basic information about yourself. Your name will not be associated with any of the information you provide, so your anonymity will be preserved.

In order for your simulation to work well, you will need to take your role seriously. You will need to give your best endeavours to playing the part of someone who may or may not be like you and who may or may not appear likeable. People tend to become emotionally involved in their simulation. This is good, because it is likely to help make the simulation realistic. Nevertheless, because the situation your character is confronted with will probably involve a conflict of some type, you should realise that behaviour that occurs during the simulation may be unsettling for you and for other people in your group.

Thank you for considering participation in this research.

Kesten Green
School of Business and Public Management
Victoria University of Wellington
Email: kesten.green@vuw.ac.nz

INFORMED CONSENT

Simulation Games for Forecasting

Informed consent is a standard requirement of the Victoria University of Wellington Human Ethics Committee. Ethics Committee approval is required for any research conducted under the auspices of the University. In order to fulfil the Ethics Committee's informed consent requirement for you to participate in this research, you must read the following material and sign your agreement at the bottom of the page.

I have read the INFORMATION SHEET about this research project and consider that I have an adequate understanding of it.

I have had the opportunity to seek clarification and elaboration on the nature of the project and my role in it, and consider that any questions I have asked have been answered to my satisfaction.

I understand that, in the course of the simulation game that I am to participate in, both I and other participants are likely to get emotionally involved. In fact, this is expected of me. I know that it is possible that some verbal exchanges between participants might be upsetting to me at the time, but accept that this may be a necessary part of the simulation task I am set.

I understand that I retain the right to withdraw myself, and any information that I have provided, prior to completing the task I am set and without providing a reason or being penalised in any way. I also understand that I will be asked to confirm my consent, by signing this form a second time, on completion of the tasks that are set for me.

I understand that nothing will be retained by the researcher that identifies me with any of the information or opinions I provide in the course of my participation in this research.

I hereby agree to take part in the research project:

Signed: _____   Date: __/__/02   {Sign *before* taking part}
Name: _____
Address: _____

I hereby confirm my agreement to the researcher retaining (in an anonymous form) the information and opinions I have provided in the course of my participation in this project:

Signed: _____   Date: __/__/02   {Sign *after* taking part}

Appendix 5: Text of email appeal for unaided-judgement participants (IACM solo version)

Subject: Using judgement to predict the outcomes of conflicts

Dear Dr X

I am writing to you because you are an expert on conflicts. I am engaged on a research project on the accuracy of different methods for predicting the outcomes of conflicts. At this stage, I'm investigating expert judgement for forecasting.

What I would like you to do is to read the attached descriptions of some real (but disguised) conflict situations and to predict the outcome of each conflict. If you can't read the attachments, please let me know and I'll send the material in your preferred format if I'm able.

Each attached file contains a conflict description and a short questionnaire. Please follow these steps for each conflict:
        1/ Read the description and
        2/ Fill-in the questionnaire (electronically if you can)
                a) make your prediction (either pick an outcome or assign probabilities)
                b) record the total time you spent on all tasks
                c) return the questionnaire.

One of the objectives of this research is to assess the effect of collaboration on forecast accuracy. You have been allocated to the *no*-collaboration treatment, so please do *not* discuss these forecasting problems with other people, as it's important that you give an individual response.

Although I intend to acknowledge the help of all of the people who assist with this research, my report will not associate any prediction with any individual.

Your prompt response is very important to the successful completion of my project. Please help me to prove the sceptics wrong about the level of cooperation I get!

Best regards,
Kesten Green
School of Business and Public Management,
Victoria University of Wellington
e-mail: kesten.green@vuw.ac.nz
Ph: (64 4) 499 2040   Fx: (64 4) 499 2080
PO Box 5530, Wellington, New Zealand

Appendix 6: Text of email appeal for structured-analogies participants (IACM solo version)

Subject: Using analogies to predict the outcomes of conflicts

Dear Dr X

I am writing to you because you are an expert on conflicts. I am engaged on a research project on the accuracy of different methods for predicting the outcomes of conflicts. At this stage, I'm investigating the formal use of "analogies" for forecasting. That is, forecasting on the basis of the outcomes of similar conflicts that are known to the forecaster.

What I would like you to do is to read the attached descriptions of some real (but disguised) conflict situations and to predict the outcome of each conflict. If you can't read the attachments, please let me know and I'll send the material in your preferred format if I'm able.

Each attached file contains a conflict description and a short questionnaire. Please follow these steps for each conflict:

1/ Read the description and
2/ try to think of several analogous situations and
3/ about how similar your analogies are to the conflict.
4/ Fill-in the questionnaire (electronically if you can)
   a) describe your analogies
   b) rate your analogies
   c) make your prediction (either pick an outcome or assign probabilities)
   d) record the total time you spent on all tasks
   e) return the questionnaire.

One of the objectives of this research is to assess the effect of collaboration on forecast accuracy. You have been allocated to the *no*-collaboration treatment, so please do *not* discuss these forecasting problems with other people, as it's important that you give an individual response.

Although I intend to acknowledge the help of all of the people who assist with this research, my report will not associate any prediction with any individual.

Your prompt response is very important to the successful completion of my project. Please help me to prove the sceptics wrong about the level of cooperation I get!

Best regards,
Kesten Green
School of Business and Public Management,
Victoria University of Wellington
e-mail: kesten.green@vuw.ac.nz
Ph: (64 4) 499 2040   Fx: (64 4) 499 2080
PO Box 5530, Wellington, New Zealand

Appendix 7: Text of email appeal for game-theorist participants

Subject: Using Game Theory to predict the outcomes of conflicts

Dear Dr X

I am writing to you because you are an expert in game theory. I am engaged on a research project which investigates the accuracy of different methods for predicting the outcomes of conflicts.

What I would like you to do is to read each of the 5 attached descriptions of real conflict situations and to predict the outcomes of each conflict. The files contain both descriptions of the situations and of the individuals or parties involved.

Each file includes a short questionnaire. Space is provided for your prediction, and for a short description of the method you used to derive your prediction. You may assign probabilities to possible outcomes, rather than picking a single outcome, if you consider this to be appropriate. If you are unable to provide a prediction for a situation, please state why in the space provided in the questionnaire. The sixth file contains a questionnaire only. Please complete it when you have finished with the 5 situations.
I would appreciate it if you do not discuss the situations with other people, as I'd rather each participant provided an independent response.

Although I intend to acknowledge all of the people, such as yourself, who help me with this research, my report will not associate any prediction with any individual.
Your prompt response is very important to the successful completion of my project.
Please help me to prove the sceptics wrong about the level of cooperation I get!

Best regards,
Kesten Green

School of Business and Public Management
Victoria University of Wellington
e-mail: kesten.green@vuw.ac.nz
Ph: (64 4) 499 2040
Fx: (64 4) 499 2080
PO Box 5530
Wellington
New Zealand

Appendix 8: Game theorist responses: A copy of Appendix 3 from Green (2002a)

<u>Not appropriate to apply game theory to the problems provided (6 responses)</u>
One stated that the aim of what he does is 'not to predict what shall happen here. This depends on the psychology of the players, which is not the object of mathematics. It is to give one of [the players] the quantitative tools that will let him act optimally according to his perceived interests'.

Another was of the opinion that 'most/many theorists see GT as prescriptive rather than descriptive', and therefore not an appropriate technique for 'predict[ing] actual behaviour'. This respondent asserted that 'game theory is a mediocre predictor of actual behaviour', and that 'I believe that the long-term ambition of (most) game theory to find optimal solutions to any decision problem is fundamentally misconceived'. Further, 'finding a situation that can be well modelled by a game of chicken tells you a number of interesting things ... it does not give you a prediction of what the outcome will be with real players ... nor ... how idealised rational players should react'.

A third game theorist stated that he 'did not see why [predicting decisions] is a game theory question'. In particular, the respondent objected that the request to predict the decision made in the one situation he had looked at, presumably the Panalba Drug Policy, 'seemed ... a question about my opinions on company ethics'.

In a fourth expert's opinion 'the role of game theory in practical situations is not so much in computing the equilibrium, but rather a useful help in thinking the situation through'. His brief outline of what this would involve had a similar flavour to the approach recommended by Nalebuff and Brandenburger (1996). The respondent went on to write that 'Game theory is a tool in understanding complex situations ... which forces you to think of the strategic aspects of the situation, but people do not always behave strategically, and one has to take that into account also'. In sum: 'The best game theory ... can offer is to explain some phenomena, but I don't see how it can predict the outcomes of real life situations'. This response was echoed by a fifth game theorist: 'I am afraid our theoretical knowledge is not straightforwardly applicable to real-life problems'. A sixth wrote that she was 'a game "theorist" and not a strategic planner', and further that she failed to 'see any "game theory" in [the] project'.

<u>Insufficient information to derive a prediction (4 responses)</u>
One stated: 'You have not provided sufficient information about preferences and institutions for me to identify a game-theoretic model and make a prediction from that'. The respondent was concerned that in order to 'predict what "really" might happen (rather than what a theoretical model would predict), I would need to know a lot more about the context in which the problems arose'.

<u>Unresolved responses (51)</u>
Twenty-five respondents stated that they could not read the MS-Word documents that contained the information on the situations and the summary questionnaire. I sent these respondents the information in the form they requested. Seventeen did not respond, four refused to participate or were on leave, and four returned completed questionnaires.

Eight respondents asked for more information about the researcher and the research. I sent replies to all those who had asked for more information but provided little extra information because in doing so I would have risked responses from this group being different from those of other respondents. As it happens, none of this group returned completed questionnaires.

As many as 44 experts responded promising help with the research; 13 of these respondents did so, and five later refused.

Appendix 9: Delphi panel appeal and part 1: Rating the importance of criteria for selecting forecasting methods

First email message for part 1

Subject: Decisions in conflicts: Delphi panel on choosing a method to predict decisions

Dear All

I now have, thanks in part to help from IACM members, data on the accuracy of four methods for forecasting decisions in conflicts. My findings on three of these have been published in the International Journal of Forecasting 18(3).

My findings on the fourth method, and on the effects on accuracy of collaboration and expertise will, I hope, be published in the same journal in due course.

Although accuracy is important when choosing a forecasting method, other factors such as timeliness and cost can also be important. The purpose of my research programme is to be able to offer useful advice to managers who face problems of forecasting behaviour in conflicts. To this end, I wish to know: 1) how important the various criteria are that one might consider when selecting a method for forecasting conflicts; 2) how you rate the four methods on the basis of the criteria; and 3) how likely it is that you would apply the methods to forecasting decisions in a real and important conflict.

I plan to gather this information using a Delphi Panel, and I wonder if you would be willing to take part. Participation won't be onerous and should be fun. All I need at this stage is for you to rate the 16 criteria, below, for importance on a 1 to 7 scale where 1=unimportant and 7=important, and to describe briefly your reason for your rating. When I have enough responses, I'll send participants an email summarising the responses, including the reasons, and will ask everyone to repeat the task taking account of the feedback. I will make sure that responses are kept anonymous.

The task is included below.

I do hope you are able to take part -- I would really appreciate your help.

Best regards
Kesten
kesten.green@vuw.ac.nz

==
Part 1, round 1:
CRITERIA FOR SELECTING A METHOD
FOR FORECASTING DECISIONS IN CONFLICTS
==

The questions below are concerned with selecting methods to predict decisions made in conflicts. Decisions such as to:
  * Declare war over water dammed by an upstream neighbour
  * Fight vigorously a hostile takeover bid
  * Go ahead with a threatened hospital nurses strike.

For each of the 16 criteria listed below, please indicate how important you think it is in selecting a method of prediction (1=unimportant; 7=important) & give a brief reason for your rating.
----

1. Accuracy
      Rate: [ ]
      Reason:

2. Timeliness in providing forecasts
      Rate: [ ]
      Reason:

3. Cost savings resulting from improved decisions
      Rate: [ ]
      Reason:

4. Ease of interpretation
      Rate: [ ]
      Reason:

5. Flexibility
      Rate: [ ]
      Reason:

6. Ease in using available data
      Rate: [ ]
      Reason:

7. Ease of use
      Rate: [ ]
      Reason:

8. Ease of implementation
      Rate: [ ]
      Reason:

9. Ability to incorporate judgemental input
      Rate: [ ]
      Reason:

10. Reliability of confidence intervals
      Rate: [ ]
      Reason:

11. Development cost (computer, human resources)
      Rate: [ ]
      Reason:

12. Maintenance cost (data storage, modifications)
      Rate: [ ]
      Reason:

13. Theoretical relevance
      Rate: [ ]
      Reason:

14. Ability to compare alternative policies
      Rate: [ ]
      Reason:

15. Ability to examine alternative environments
      Rate: [ ]
      Reason:

16. Ability to learn (experience leads forecasters to improve procedures)
      Rate: [ ]
      Reason:

Finally, had you read my paper, or any commentary on my research on this topic, before completing this task?
      Yes: [ ]
      No: [ ]

Thank you.
Please send your response to me at...
kesten.green@vuw.ac.nz

Second email message for part 1 – example message to one panellist

Subject: Decisions in conflicts: Delphi panel feedback

Dear X

Thanks again for your responses.

At the end of this message is a copy of the questionnaire you received before, with the addition of summaries of the responses from the seven panelists. I've provided the median, minimum, and maximum responses for each criterion as well as one or two of the reasons that panelists provided for a high rating and one or two of the reasons for a low rating.

Please consider the summary information, and type in your final rating for each of the criteria and your reason for the rating.

I look forward to seeing your new responses.

Best regards
Kesten
kesten.green@vuw.ac.nz

==
Part 1, round 2:
CRITERIA FOR SELECTING A METHOD
FOR FORECASTING DECISIONS IN CONFLICTS
==
The questions below are concerned with selecting methods to
predict decisions made in conflicts. Decisions such as to:
  * Declare war over water dammed by an upstream neighbour
  * Fight vigorously a hostile takeover bid
  * Go ahead with a threatened hospital nurses strike.

For each of the 16 criteria listed below, please indicate how
important you think it is in selecting a method of prediction
(1=unimportant; 7=important) & give a brief reason for your rating.

1. Accuracy
Reasons for high rating:
   Not sure what the point of a forecasting method is if it is not accurate
   Not useful if inaccurate
Reasons for low rating:
   Conflicts are dynamic and accuracy may assume a single correct option rather
        than a set of possible options.
   Foolish to think that any predicting model could be 100%, so should not rely
        on it only.
            Median 7
            Max   7
            Min   5
            Your initial rating   6
            Your final rating [_____]
            Your reason:

2. Timeliness in providing forecasts
Reasons for high rating:
   If it is not timely then no practical application to it
   More important than accuracy in that deadlines are liberating while accuracy
constricts.
Reasons for low rating:
   It can wait (usually).
   Too late is useless, but deadlines are usually flexible.

Median 6
Max   7
Min   4
Your initial rating   4
Your final rating [_____]
Your reason:

3. Cost savings resulting from improved decisions
Reasons for high rating:
   This seems another important reason to forecast
   Always one of the considerations in business. But not the most important
Reasons for low rating:
   Costs are seldom the issue. Value is the issue.
   Psychological 'saving' as important as cost savings
        Median 4
        Max   7
        Min   3
        Your initial rating   4
        Your final rating [_____]
        Your reason:

4. Ease of interpretation
Reasons for high rating:
   Seems unlikely to be useful if one can't make sense of the forecast
   Forecasting should provide enhanced description rather than explanation. The
        greater the lucidity of the description the greater the flexibility of the
        interpretation.
Reasons for low rating:
   If accuracy is high, my ease is not so important
   Depends on how often and who has to use it.
        Median 5
        Max   7
        Min   2
        Your initial rating   5
        Your final rating [_____]
        Your reason:

5. Flexibility
Reasons for high rating:
   Yeah, don't want to add an "attachment" for each different conflict.
   Related to #6 - the method must be adaptable to the available data
Reasons for low rating:
   Predicting decisions should set standards that move towards certainty.
        Variables which enhance flexibility reduce certainty.
   Despite the specificity of usage, I should have thought there was value in
        using tools that could inform future related (but not identical)
        problems and also provide a degree of pre-fabricated structure on
        which to hang one's thinking.
        Median 5.5
        Max   6
        Min   3
        Your initial rating   6
        Your final rating [_____]
        Your reason:

6. Ease in using available data
Reasons for high rating:
   Most businesses would not engage in model use if the process itself required
        something more complicated than the use of what they already have
   A good forecasting model should be adaptable to the data
Reasons for low rating:
   I'd have thought that this shouldn't be over-emphasised since it is the
        unknown and the hard to predict that will frequently determine these

situations: available data will tend to confirm settled views.
If accuracy is high, my ease is not so important
 Median 6
 Max 6
 Min 2
 Your initial rating 6
 Your final rating [_____]
 Your reason:

7. Ease of use
Reasons for high rating:
   Simply because in a serious situation you don't want time wasted learning on
       the job.
   I'm basically as lazy as the next guy
Reasons for low rating:
   If it is accurate a specialist should be able to take the time to learn how
       to use it
   If accuracy is high, my ease is not so important
 Median 5
 Max 7
 Min 2
 Your initial rating 7
 Your final rating [_____]
 Your reason:

8. Ease of implementation
Reasons for high rating:
   A reality of business
   Time and cost of organizing has to be taken into account
Reasons for low rating:
   All else being equal this seems less important than accuracy.

 Median 5
 Max 7
 Min 3
 Your initial rating 5
 Your final rating [_____]
 Your reason:

9. Ability to incorporate judgemental input
Reasons for high rating:
   Dynamic systems are always more accessible to input than static systems.
   It is important that a third party not only comment on the likely outcome
       but also indicate whether this is an appropriate, just, effective outcome
       or if there is another intervention process that would likely result in a
       better outcome
Reasons for low rating:
   I probably would in any case anyway.
   Would be helpful
 Median 7
 Max 7
 Min 4
 Your initial rating 5
 Your final rating [_____]
 Your reason:

10. Reliability of confidence intervals
Reasons for high rating:
   Otherwise, I would suspect the instrument each time - not good!
   Like accuracy, good reliability is critical
Reasons for low rating:
   Often too statistically static which gives a false positive.

Median 6
Max   7
Min   4
Your initial rating   7
Your final rating [_____]
Your reason:

11. Development cost (computer, human resources)
Reasons for high rating:
   The greater the up-front cost, the more secure in the forecasting
      variables.
   Yes important as if conflicts are not common then investing in development
      costs may not be seen to be value for money
Reasons for low rating:
   Just seems less important than flexibility, accuracy, etc. if it results in
      long-term cost savings
   I guess must be considered to get it done
         Median 5.5
         Max   7
         Min   3
         Your initial rating   6
         Your final rating [_____]
         Your reason:

12. Maintenance cost (data storage, modifications)
Reasons for high rating:
   This rating depends on whether or not re-use/reinterpretation is envisaged.
      If not, lower.
   No one will use it if has high costs to maintain, harder to justify
Reasons for low rating:
   An externality that is difficult to compute
   If all else is good, will find $ for maintenance
         Median 5
         Max   6
         Min   2
         Your initial rating   5
         Your final rating [_____]
         Your reason:

13. Theoretical relevance
Reasons for high rating:
   Theory grows from practice not the other way around. Predictions are too
      often caste based on the idiosyncratic life histories of a small group of
      practitioners. Prediction should be inductive.
   I'd still say this was reasonably important since theoretical parameters
      provide one useful set of benchmarks
Reasons for low rating:
   In business I think this is less considered
   Sufficient if it works
         Median 4
         Max   7
         Min   3
         Your initial rating   6
         Your final rating [_____]
         Your reason:

14. Ability to compare alternative policies
Reasons for high rating:
   Should be able to compare different outcomes that might result from
      different solutions, conflict management methods - show people options
   This is vital for the exercise of judgemental input.
Reasons for low rating:
   While important not sure business always takes this approach

A really nice feature - but not essential if the primary purpose works.

    Median 6
    Max   7
    Min   4
    Your initial rating   6
    Your final rating [_____]
    Your reason:

15. Ability to examine alternative environments
Reasons for high rating:
   Environmental context or conflict arenas have a higher productivity than the
      weighing of the conflict skills of the participants.
   Many similar types of conflicts occur in different contexts, this would be
      very valuable
Reasons for low rating:
   Depends on likelihood of re-use I suppose
   A really nice feature - but not essential if the primary purpose works.

    Median 6
    Max   7
    Min   3
    Your initial rating   6
    Your final rating [_____]
    Your reason:

16. Ability to learn (experience leads forecasters to improve procedures)
Reasons for high rating:
   Should lead to improved future conflict management - what we are all
      striving for.
   Hopefully, it's easier to learn/retain than Microsoft Access!
Reasons for low rating:
   This depends very much on the field. The less likely it is that
      circumstances will re-produce themselves, the less important - if
      earning (as your parentheses imply) is about adapting the methodology
      for further use.

    Median 6
    Max   7
    Min   4
    Your initial rating   6
    Your final rating [_____]
    Your reason:

Appendix 10: Delphi panel part 2: Rating the forecasting methods against the selection criteria

First email message for part 2

Subject: Four methods for forecasting decisions in conflicts

Dear X

Thanks again for taking part in the Delphi Panel for Part 1. I look forward to seeing your responses for Part 2 -- this is where the task gets more specific and interesting.

The questionnaire for Part 2 is pasted below. Before you complete the questionnaire, please read the attached file. The file contains information that you will need to know on four forecasting methods for conflicts.

Delphi - descriptions of

metho...

Kind regards
Kesten

==
Part 2, round 1:
RATING OF METHODS
FOR FORECASTING DECISIONS IN CONFLICTS
==

The questions below ask you to judge how four methods for
forecasting decisions in conflicts measure-up against the 16
criteria you are now familiar with.

The methods and evidence on their performance are described in
the attached 3-page MS-Word document. Please read the
document before answering the questions. The four methods are:

        * game theory (GT)

        * simulated interaction (SI)

        * structured analogies (SA)

        * unaided judgement (UJ)

For the 16 criteria listed below, please rate each of the methods
with a score of between zero and 10 (0=inadequate; 10=excellent)
and give a brief reason for your rating.

For some of the criteria, there is no direct information on the
methods in the MS-Word document. In order to fill in the gaps
in the information, please think of how the methods could be
used for conflict forecasting problems you are familiar with.
----

1. Accuracy

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

2. Timeliness in providing forecasts

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

3. Cost savings resulting from improved decisions

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

4. Ease of interpretation

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

5. Flexibility

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

6. Ease in using available data

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

7. Ease of use

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

8. Ease of implementation

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

9. Ability to incorporate judgemental input

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

10. Reliability of confidence intervals

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

11. Development cost (computer, human resources)

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

12. Maintenance cost (data storage, modifications)

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

Reason:

13. Theoretical relevance

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

14. Ability to compare alternative policies

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

15. Ability to examine alternative environments

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

16. Ability to learn (experience leads forecasters to improve procedures)

 Rate: GT[ ]   SI[ ]   SA[ ]   UJ[ ]

 Reason:

Thank you.

Please send your response to me at...
kesten.green@vuw.ac.nz

Four methods for forecasting decisions in conflicts

1. Decisions in conflicts

Managers often wish to know how a conflict will unfold. Whether a conflict is industrial, commercial, civil, political, diplomatic, or military, predicting the decisions of others can be difficult. Yet it is important that managers plan for likely eventualities and seek effective strategies. Failure to accurately predict the decisions of others can lead to needless strikes, losses, protests, reversals, wars, and defeats. This document describes findings from research on methods for forecasting decisions made in particular conflicts. Conflicts that involve interaction between few parties.

Eight diverse conflicts were used in the research. The conflicts were all real situations that were disguised for the purposes of the research. In brief, the conflicts and the decisions to be forecast were:

Artists' Protest: Members of a rich nation's artists' union occupied a major gallery and demanded generous financial support from their government. What will be the final resolution of the artists' sit-in?
Employee grievance: An employee demanded a meeting with a mediator when her job was down-graded after an evaluation by her new manager. What will the outcome of the meeting with the mediator be?
55% Pay Plan: Professional sports players demanded a 55 percent share of gross revenues and threatened to go on strike if the owners didn't concede. Will there be a strike and, if so, how long will it last?
Telco takeover bid: An acquisitive telco, after rejecting an offer to buy the mobile business of another, made a hostile bid for the whole corporation. How will the stand-off between the companies be resolved?
Distribution Channel: An appliance manufacturer proposed to a supermarket chain a novel arrangement for retailing their wares. Will the management of the supermarket chain agree to the plan?
Nurses Dispute: Angry nurses increased their pay demand and threatened more strike action after specialist nurses and junior doctors received big increases. What will the outcome of their negotiations be?
Water dispute: Troops from neighbouring nations moved to their common border and the downstream nation threatened to bomb a new upstream dam. Will the upstream neighbour agree to release additional water and, if not, how will the downstream nation's government respond?
Zenith Investment: A large manufacturer evaluated an investment in expensive new technology in the face of political pressure. How many new manufacturing plants will the corporation decide to commission?

2. Forecasting methods

The accuracy of forecasts from four conflict forecasting methods has been compared using the eight conflicts described. The four methods were: game theory, simulated interaction, structured analogies, and unaided judgement. These are the principal methods for forecasting decisions in conflicts in that they are either commonly used or are recommended by experts, or both.

Game theory [i]. Formal analysis of the behaviour of two or more parties with divergent interests in situations that can be described by rules. For example, Prisoner's Dilemma is one of the more popular of the games that have been studied. Game theoretic analysis seems to provide insight into historical situations involving conflict and cooperation. To be useful, however, analysis must be done in advance of the outcome, and this is likely to be difficult. In effect, a game theorist must describe a game that is analogous to the target situation.

Simulated interaction (SI) [ii]. Simulated interaction involves the acting out of interactions among people or groups who have roles that are likely to lead to conflict. SI is a subset of role playing, as role playing is also used for situations that do not involve interaction. An unaided expert must try to think through several rounds of interaction in order to make a forecast. In contrast, SI can realistically simulate interactions. SI can be used to forecast the effect of different strategies.

Structured analogies (SA) [iii]. Analogies are commonly used in an informal way when people make judgmental forecasts. Structured analogies involves domain experts selecting situations that are similar to a target situation, describing the similarities and differences, and providing an overall similarity rating for each similar (analogous) situation. The outcomes of the analogous situations are then used to forecast the outcome of the target situation. SA depends on the availability of situations that are similar to the target.

Unaided judgement (UJ). Managers mostly rely on their own judgement for forecasting decisions in conflicts; either entirely or in conjunction with the judgemental predictions of others who know about the situation. In some situations it may be practical to ascertain the judgements of the other party or parties to a conflict, in the form of their stated intentions – for example, Taleban leaders stated they would increase terrorist attacks on US targets in response to any bombing of Afghanistan. Managers can incorporate such information into their own judgemental forecasts of the behaviour of another party. The term "unaided judgement" is intended to be self-explanatory – it is judgement without recourse to a formal method.

3. Evidence from research

The research reported here is the only formal evidence available on the relative accuracy of forecasts of decisions in real conflicts from the four methods.

Method

Research participants were given descriptions of one or more of the eight conflicts and forecast, using one of the four forecasting methods, the decisions that were actually made. There are between 2 and 6 possible decisions for the conflicts. Selecting randomly, one would expect to be correct 31 percent of the time.

Conflict descriptions were the same for all forecasting methods. Descriptions of all roles were provided to all *expert* participants, to some *non-experts* using unaided judgement, but not to any *simulated interaction* participants. Participants choose a single decision from a list or assigned probabilities to decisions.

Game theory: Descriptions of the conflicts and questionnaires were provided via email. Game theory experts were informed that their participation in the research involved "using game theory to predict the outcomes of conflicts" as part of a project that "investigates the accuracy of different methods for predicting the outcomes of conflicts".

Simulated interaction: Participants assembled in a lecture theatre or similar venue to simulate a single conflict. Each was given the description of one role and was asked to adopt that role for the duration of the session. Where possible, they were allocated to roles with which they were most able to empathise. Participants then read a description of the conflict and simulated interactions between the parties to the conflict until a decision was made or time ran out. The simulation decision was the forecast. Sessions typically lasted 50 minutes including role assignment, briefing, reading, role-playing, and questionnaire completion.

Structured analogies: Descriptions of the conflicts and questionnaires were provided via email. Participants were asked to describe several analogous conflicts, rate their similarity to the target conflict, and (for each analogy) nominate the target conflict decision closest to the decision that was actually made in the analogous conflict.

Unaided judgement: Participants using this method were given no special instructions or were structured analogies participants who were unable to think of any analogies.

Expertise

Non-experts (novices) were mostly university students.

Game theory experts were recruited using personalised email appeals to Game Theory Society members and to recipients of the International Society of Dynamic Games E-Letter.

Experts in conflict management, forecasting, decision making, marketing, and employment relations were recruited using personalised email appeals to members of the International Association of Conflict Management and the International Institute of Forecasters' Research Associates lists, and using impersonal email appeals to the CMD-Net, JDM, Decision, ELMAR, HR-Net, IERN-L, and PRIR-L lists. Personal appeals were also made to members of a convenience sample of experts. Marketing experts were sent the Distribution Channel and Telco Takeover Bid material, employment relations experts were sent the Artists' Reprieve, Employee Grievance, 55% Pay Plan, and Nurses Dispute material, while other experts were sent material for all eight conflicts.

Findings

Overall, simulated interaction using novice role-players was the method that provided the most accurate forecasts at 64 percent correct (Table 1 – bold figures are the best forecasts for each of the conflicts). Game theory experts' forecasts (37 percent correct) were no more accurate than those of either experts or novices using unaided judgement. Nor were they better than chance. Structured analogies forecasts were 49 percent correct.

Simulated interaction provided the most consistently accurate forecasts. Forecasts from the method were the most accurate forecasts for five of the eight conflicts and were not significantly different in the case of the other three.

Table 1: Summary of Findings
Accuracy: Percent correct predictions (number of forecasts)

		Novice*	Expert			Novice*
	Chance	UJ	UJ	GT	SA	SI
Artists' Protest	17	5 (39)	9 (22)	6 (18)	**30** (10)	29 (14)
55% Pay Plan	25	27 (15)	14 (14)	29 (17)	31 (13)	**60** (10)
Telco Takeover Bid	25	10 (10)	0 (9)	0 (6)	14 (7)	**40** (10)
Distribution Channel	33	5 (42)	35 (20)	31 (13)	50 (11)	**75** (12)
Nurses Dispute	33	68 (22)	71 (14)	50 (14)	69 (13)	**82** (22)
Water Dispute	33	45 (11)	40 (5)	60 (5)	**100**(3)	90 (10)
Zenith Investment	33	29 (21)	40 (15)	22 (18)	38 (8)	**59** (17)
Employee Grievance	50	67 (9)	75 (4)	**100**(5)	59 (11)	80 (10)
Average (unweighted)	31	32 (169)	36 (103)	37 (96)	49 (76)	**64** (105)

Forecasts from non-experts for Artists' Protest, 55% Pay Plan, and Distribution Channel reported in Armstrong (2001)[iv] except for 13 UJ findings: Artists' (1 correct / n=8); Distribution (1/5). All other forecasts from Green (2002)[v] and other (as yet unpublished) findings by Green.

[i] This is an edited version of a definition in the online Forecasting Dictionary: http://www.forecastingdictionary.com .

[ii] ditto

[iii] ditto

[iv] Armstrong, J. S. (2001). Role playing: a method to forecast decisions. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 15-30.

[v] Green, K. C. (2002a). Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement. *International Journal of Forecasting*, *18*, 321-344.

Second email message for part 1 – example message to one panellist

Subject: Forecasting methods: Delphi panel feedback

Dear X

This is the last round for the panel -- thank you very much for persevering.

Following this message is a copy of the questionnaire you received before, with the addition of summaries of the responses from all the panelists: median, minimum, and maximum for each criterion for each method. I've also provided edited reasons for responses.

Please consider the summary information, and type in your final ratings for the methods and your reasons for the ratings.

I look forward to seeing your final set of responses.

Best regards
Kesten
kesten.green@vuw.ac.nz

```
==============================================
Part 2, round 2:
RATING OF METHODS
FOR FORECASTING DECISIONS IN CONFLICTS
==============================================
```

The questions below ask you to judge how four methods for forecasting decisions in conflicts measure-up against the 16 criteria you are now familiar with.

The methods and evidence on their performance are described in the 3-page MS-Word document that you have read.

The four methods are:

> * game theory (GT)
>
> * simulated interaction (SI)
>
> * structured analogies (SA)
>
> * unaided judgement (UJ)

For the 16 criteria listed below, please rate each of the methods with a score of between zero and 10 (0=inadequate; 10=excellent) and give a brief reason for your rating.

For some of the criteria, there is no direct information on the methods in the MS-Word document. In order to fill in the gaps in the information, please think of how the methods could be used for conflict forecasting problems you are familiar with.
```
-------------------------------------------------------------
```

1. Accuracy

The research outcomes you have identified.
SI offers opportunity to place oneself within the situation - yielding better insight and decision.
Zero for GT & UJ on the basis that something no better than chance is inadequate (i.e. adequacy is something that is better

217

than chance).
The cases I think of contain too many dimensions for game theory;
SI, I am learning from you, may be excellent for them; I can't think
of analogies that capture very much useful; and I used to think UJ
(my way) was more useful than your statistics seem to bear out.

*	GT	SI	SA	UJ
Median	4	7	3	3
Max	5	9	5	6
Min	0	6	3	0
Your initial ratings	4	7	5	3
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

2. Timeliness in providing forecasts

Where there is less need for external assistance there is a quicker
response.
The fewer steps to accomplish the quicker the results.
There was no discussion of how long these take, but it seems that
all require some time and effort to collect information or opinions or
allow for setting up and enactment of a simulation   .
I can't see that any of these methods is particularly time-
consuming (even assembling people for SI).
For GT one has to know too much before one can predict; how
does one simulate before knowing the later scenes; if analogies
work, they work early; ditto, UJ.
NOTE: SI ROLE-PLAYERS ARE GIVEN SAME INFORMATION AS
PROVIDED FOR OTHER METHODS & IMPROVISE
INTERACTIONS UNTIL A DECISION IS MADE
-- IE. THEY ARE NOT GIVEN A SCRIPT.

*	GT	SI	SA	UJ
Median	5	5	8	8
Max	8	8	8	9
Min	1	2	5	6
Your initial ratings	1	5	8	9
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

3. Cost savings resulting from improved decisions

Not likely as incidence of predicted outcome unreliable.
Limited time invested (total)yields cost savings.
Only SI had good enough accuracy to result in much cost savings.

I assume savings correlate with accuracy since in the absence
of accuracy there will be no impact on the quality of decisions.

SI is relatively expensive; the others are relatively cheap.
NOTE: GT, SA, UJ REQUIRES EXPERTS. SI CAN USE STUDENTS
AS ROLE-PLAYERS.

*	GT	SI	SA	UJ
Median	3	3	3	3
Max	8	8	8	8
Min	0	2	2	0
Your initial ratings	3	3	3	3
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

4. Ease of interpretation

I am assuming business context here. So I suppose the more
direct methods will be deemed better.
Less "expertism" = easier interpretations.
Seems that the results of SI or UJ would be easy to interpret,
while GT is more complicated and SA seems difficult because
it may be hard to find truly analogous situations.
SA seems to me the only one that poses some problems
since analogues are never going to be exact so there is an
ambiguity that won't exist with the precise spuriousness of
GT or the precise outcomes of the role play or unaided
judgements.

*	GT	SI	SA	UJ
Median	5	5	5	8
Max	9	10	7	10
Min	2	2	2	6
Your initial ratings	2	5	5	6
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

5. Flexibility

I think judgement will be dependent on the individual's ability to
think laterally SI, SA because of their approach will allow greater
variation and responsiveness.
Most flexible when only one mind is involved.
Again, simulations can always be created and unaided judgment
can always be done, but the others require locating an analogous
situation or making it fit the game theory model.
GT at 5 because of the difficulty in describing all situations in
terms of decision rules.
The virtue of GT is rigor, ie its INflexibility.

*	GT	SI	SA	UJ
Median	4	9	7	9
Max	5	10	9	10
Min	0	5	2	5
Your initial ratings	0	8	7	5
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

6. Ease in using available data

None are perfect - each has advantages.
Simulations can always be created and unaided judgment can
always be done, but the others require locating an analogous
situation or making it fit the game theory model.
I like to write role plays so turning data into a role play is
relatively easy.

*	GT	SI	SA	UJ
Median	4.5	8.5	5.5	8.5
Max	6	10	7	10
Min	2	6	2	6
Your initial ratings				
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

7. Ease of use

Less complexity in setting up will rate SA & UJ higher.
An individual has the upper hand with "ease."
Simulations can always be created and unaided judgment can
always be done, but the others require locating an analogous
situation or making it fit the game theory model. Nevertheless,
SI is not so easy as UJ since you have to set up the simulation
and find participants, etc.
This assumes availability of experts for GT and SA.
The simulation is the most complex to prepare and stage.

*	GT	SI	SA	UJ
Median	4	4	6	9
Max	8	8	8	10
Min	0	2	2	8
Your initial ratings	0	2	7	8
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

8. Ease of implementation

An individual decision maker has the easiest time of it.
GT seems complicated to me, but I suppose if you have an expert
on hand it would be easier to implement (or just as easy) as the
others. SA seems most difficult because of the need to find an
analogous situation. SI and UJ are rated lower here because they
both require cooperation of others which may not always be easy
in actual implementation.
Can't see any reason why any of these should be difficult given
appropriate expertise.
The simulation is the most complex to prepare and stage.

*	GT	SI	SA	UJ
Median	7.5	5.5	4.5	7.5
Max	8	8	8	9
Min	5	2	2	7
Your initial ratings				
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

9. Ability to incorporate judgemental input

A less structured approach is more responsive.
Unaided judgement is most able (consciously or unconsciously).

GT seems most interested in probable outcomes without
commentary. In SI it seems you could get judgments from
participants which may be relevant. Assuming one can find an
analogous situation, it would be easy to discuss the advantages
and disadvantages of what had happened and how it could have
been improved, so this and UJ seem most conducive to
judgmental input.
I assume GT is more problematic here since once you've decided
on your decision rules you're stuck with them (but then I suppose
you can make allowance for that - sorry, not well enough
informed here).
SI and UJ are the most flexible for this.

*	GT	SI	SA	UJ
Median	5	8	8	10
Max	7	10	10	10
Min	2	5	5	9
Your initial ratings	5	8	8	10
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

10. Reliability of confidence intervals

Pre-structure will aide outcomes.
Don't feel comfortable rating this one.
I assume this correlates with accuracy.
NOTE: PERHAPS THINK OF IT THIS WAY… IF A FORECASTER
USING METHOD X SAYS "I AM 75% CONFIDENT THAT
DECISION Y WILL OCCUR", WILL HE-OR-SHE BE CORRECT
75% OF THE TIMES HE-OR-SHE MAKES SUCH AN ASSERTION?

*	GT	SI	SA	UJ
Median	4	7	3.5	1
Max	8	8	4	2
Min	0	6	3	0
Your initial ratings				
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

11. Development cost (computer, human resources)

Unaided judgements appear to cost least in terms of development
costs.
All seem to require significant time or resources except UJ.

*	GT	SI	SA	UJ
Median	4	6	7	9
Max	7	9	8	10
Min	2	2	5	2
Your initial ratings	2	2	7	10
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

12. Maintenance cost (data storage, modifications)

Not sure what maintenance costs would be with these models.

Models must be developed/maintained and administrated.

All seem to require significant time or resources except UJ.

I assume SI can't be stored since its reliant on mobile human
capital that can't always be reassembled.

*	GT	SI	SA	UJ
Median	7	6	7	8
Max	8	7	8	10
Min	6	5	4	2
Your initial ratings				
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

13. Theoretical relevance

GT & SA high as both will be useful in developing predictive models, even if only applicable in specific incidents. SA probably more useful in a broader application.
The "purer" the process - the more theoretical (I would suppose).

Only GT is directly derived from theory, others could involve and certainly could inform theory but hard to say whether they would. Can't judge the level but I'd put them all the same except UJ.

NOTE: IS THE THEORY THAT THE MORE REALISTICALLY A METHOD CAN REPRESENT THE SITUATION, THE MORE ACCURATE IT WILL BE, RELEVANT HERE?

*	GT	SI	SA	UJ
Median	8	5	6	3.5
Max	10	7	7	7
Min	6	2	5	2
Your initial ratings	7	4	7	5
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

14. Ability to compare alternative policies

The unaided judges would think they can compare.
Presumably one could simulate the conflict with several policies and compare the results, seems most difficult to do in UJ.
I assume SA is going to be limited by the search for useful analogies.
The GT will lend itself most readily.

*	GT	SI	SA	UJ
Median	8	7	5	5
Max	10	9	9	8
Min	7	2	2	2
Your initial ratings	8	7	7	4
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

15. Ability to examine alternative environments

Very hard to compare "fights" to "disputes."
GT and SI seems to lend themselves to that since there is control over the set up of the scenario/game/simulation. Again SA requires finding an appropriate analogy and UJ would be harder unless someone had experience in many different environments (or access to advisors who do).

*	GT	SI	SA	UJ
Median	6	6	6	4
Max	10	10	8	6
Min	4	5	4	4
Your initial ratings	4	5	6	4
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

16. Ability to learn (experience leads forecasters to improve procedures)

Maybe the higher level of participation in the process allows for

greater learning.

It would seem that the unaided judge is well placed to learn and learn quickly - but is this so?

I am assuming that experience with any of the methods would be beneficial and improve future performance.

Again my instinct is that three of these lend themselves to easy improvement whereas SA leaves forecasters in the hands of the knowledge and experience of domain experts and the existence (or not) of analogies.

SI, as I understand it, has no memory as there are new players each time.

*	GT	SI	SA	UJ
Median	8	6	8	9
Max	10	10	10	10
Min	5	2	5	8
Your initial ratings	5	6	8	8
Your final ratings	[___]	[___]	[___]	[___]
Your reasons:				

Thank you.

Please send your response to me at...
kesten.green@vuw.ac.nz

Appendix 11: Delphi panel part 3: Likelihood that methods would be used or recommended by panellists

Email message for part 3

Subject: Results of Delphi Panel on conflict forecasting methods

Dear Panelists

Thank you again for your help with this task.

The panel started with seven of you who participated in rating the importance of forecasting method selection criteria in the first part of the task.

Five of you continued on the panel (for part 2 round 1) to rate four methods against the criteria. Three of you responded to the feedback from 2/1 and modified your ratings or not, as you saw fit, in part 2 round 2.

I have included in the final averages both the ratings from the three who responded in 2/2 and the ratings from 2/1 for the two who did not.

After weighting each method's ratings for each criteria by the importance of the criteria, the overall ratings for the methods are (out of 10):

Simulated interaction:	6.6
Unaided judgement:	6.4
Structured analogies:	6.0
Game theory:	5.4

I'm sorry to bother you again after all your help, but would you mind answering four quick questions? With the time you have spent considering the problem of forecasting decisions in conflicts, your responses would be extremely valuable.

Having taken part in this panel, how likely would you be to use or recommend each of the methods the next time you are faced with an important conflict forecasting problem? Specifically, what are the chances out of 10 (where 10 indicates you are certain or practically certain, and zero indicates there is no chance or practically no chance) that you would use or recommend for an important conflict forecasting problem:

Game theory?   [   ] 0-10

Simulated interaction?   [   ] 0-10

Structured analogies?   [   ] 0-10

Unaided judgement?   [   ] 0-10

{Note that your four responses don't have to add to 10 as you may expect to use more than one of the methods}.

Best regards, Kesten
kesten.green@vuw.ac.nz

Appendix 12: Number of forecasts, by conflict, method, and forecast decision [a]

Conflict	Method	Expertise		Forecast A	B	C	D	E	F	X[b]
Artists Protest	Unaided judgement	novices	solo	**1**	5	1	1	0	0	
		experts	solo	**2**	9	4	3	2	0	
			joint	**0**	1	1	2	0	0	
	Game theorist	experts	solo	**1**	5	7	2	0	2	1
	Structure analogies	experts	solo	**1**	1	2	0	1	0	
			joint	**2**	0	1	0	1	0	
	Simulated interaction	novices	joint							
Distribution Channel	Unaided judgement	novices	solo	**1**	3	*0*	1			
		experts	solo	**2**	5	*8*	3			
			joint	**0**	1	2	0			
	Game theorist	experts	solo	**2**	7	2	2			
	Structure analogies	experts	solo	**3**	2	3	1			
			joint	**2**	1	*0*	0			
	Simulated interaction	novices	joint							
55% Pay Plan	Unaided judgement	novices	solo							
		experts	solo	**2**	5	4	1			
			joint	**0**	0	4	0			
	Game theorist	experts	solo	**5**	5	5	2			
	Structure analogies	experts	solo	**3**	4	0	1			
			joint	**2**	1	0	2			
	Simulated interaction	novices	joint							
Nurses Dispute	Unaided judgement	novices	solo	4	3	**15**				
		experts	solo	3	1	**10**				
			joint	0	0	**1**				
	Game theorist	experts	solo	4	3	**7**				
	Structure analogies	experts	solo	1	1	**6**				
			joint	1	1	**3**				
	Simulated interaction	novices	joint	3	1	**18**				
Personal Grievance	Unaided judgement	novices	solo	2	**4**	1	2			
		experts	solo	1	**2**	1	0			
			joint							
	Game theorist	experts	solo	2	**3**	0	0			
	Structure analogies	experts	solo	1	**5**	4	2			
			joint	1	**1**	0	0			
	Simulated interaction	novices	joint	2	**6**	2	0			
Telco Takeover	Unaided judgement	novices	mixed	2	**1**	1	6			
		experts	solo	4	**0**	3	2			
			joint							
	Game theorist	experts	solo	1	**0**	1	5			
	Structure analogies	experts	solo	1	**1**	1	4			
			joint	0	**0**	2	0			
	Simulated interaction	novices	joint	2	**4**	1	3			
Water Dispute	Unaided judgement	novices	solo	**5**	4	2				
		experts	solo	**3**	2	0				
			joint	**0**	0	1				
	Game theorist	experts	solo	**4**	0	2				
	Structure analogies	experts	solo	**3**	0	1				
			joint	**1**	0	0				
	Simulated interaction	novices	joint	**9**	1	0				
Zenith Investment	Unaided judgement	novices	solo	3	**6**	12				
		experts	solo	4	**4**	6				
			joint	0	**2**	0				
	Game theorist	experts	solo	4	**4**	10				
	Structure analogies	experts	solo	3	**3**	1				
			joint	1	**0**	0				
	Simulated interaction	novices	joint	5	**10**	2				

Notes

Figures in bold are numbers of accurate forecasts and are scored as 1 for proportion correct calculations. Figures in italics (Distribution Channel forecast C) are scored ½, as the forecast option chosen included both the actual decision and another decision. Probabilistic forecasts are included if the highest probability was given to a single decision option.

a   Forecasts new to this research; i.e. forecasts reported in Armstrong (2001a) are not shown

b   Probabilistic forecasts that were inaccurate (the actual decision was given a low or zero probability) but probabilities were too evenly distributed to choose a single decision option.

Appendix 13: Comparison of Brier scores and *PFAR* scores

In order to compare the characteristics of the Brier score and *PFAR* score measures, I examined the scores for an illustrative subset of the set of probabilistic forecasts that would be likely to lead a decision maker to anticipate a particular outcome (Table 47). That is, forecasts where one option is allocated a probability that is greater than the probabilities allocated to each of the other options.

The *PFAR* score is a ratio, and therefore aggregations of the score should properly be calculated as the sum of all absolute errors for a set of forecasts divided by the sum of all naïve forecast errors for that set. Nevertheless, the denominator of the score varies only with the number of outcome options, and an average (mean) of individual scores is mathematically equivalent to the calculation of a grand ratio when the denominator of the individual scores is the same. This is the case in the illustration I use for comparing the measures, and for each conflict in my research. I therefore use averages in my illustration and, with actual data, for the calculation of aggregate scores for each conflict. I use medians of the conflict averages to compare forecasting methods.

Because the Brier score squares errors, the measure tends to favour an even allocation of probabilities and penalise a polarised distribution. Here are three illustrations of problems that can arise from this characteristic of the Brier score, accompanied by comparisons with the *PFAR* measure. First, the Brier score for a naïve forecast of the outcome of a situation with four defined outcome options $BS(3^*,3,3,3)$ is 0.75. I have circumscribed scores that are greater than or equal to this value in the upper half of the table. An anomaly is revealed. Forecasts that would have led a decision maker who found the forecasts credible to have correctly anticipated the actual outcome one time in two, have a worse (higher) Brier score than the naïve forecast – correct one time in four. Forecasts with high-probability first choices (the options assigned the highest probability for each conflict) score little better than forecasts with lower-probability first choices when they are accurate. When they are not accurate, forecasts with high-probability first choices are heavily penalised.

Whereas the Brier score for a naïve forecast varies with the number of options, the *PFAR* score for a naïve forecast is always 1.0. I have circumscribed *PFAR* scores that are greater than or equal to 1.0 in the lower half of Table 47. *PFAR* scores lead to the

sensible conclusion that any set of forecasts likely to lead a decision maker to correctly anticipate actual outcomes more frequently than chance, should be preferred to the naïve forecast.

Table 47
Effect of assignment of probabilities on average error measures for many forecasts [a]
(Forecasts for four outcome options)

BRIER SCORES

Forecast probabilities[b]	Frequency with which first choice is correct (percent)					
	100	90	75	50	25	0
(**1.00**, 0.00, 0.00, 0.00)	0.00	0.20	0.50	1.00	1.50	2.00
(**0.95**, 0.02, 0.02, 0.02)	0.00	0.19	0.47	0.94	1.40	1.87
(**0.95**, 0.05, 0.00, 0.00)	0.01	0.19	0.47	0.94	1.41	1.87
(**0.75**, 0.08, 0.08, 0.08)	0.08	0.22	0.42	0.75	1.08	1.42
(**0.75**, 0.25, 0.00, 0.00)	0.13	0.26	0.46	0.79	1.13	1.46
(**0.50**, 0.17, 0.17, 0.17)	0.33	0.40	0.50	0.67	0.83	1.00
(**0.30**, 0.23, 0.23, 0.23)	0.65	0.67	0.69	0.72	0.75	0.79

PFAR SCORES

Forecast probabilities[b]	Frequency with which first choice is correct (percent)					
	100	90	75	50	25	0
(**1.00**, 0.00, 0.00, 0.00)	0.00	0.13	0.33	0.67	1.00	1.33
(**0.95**, 0.02, 0.02, 0.02) (**0.95**, 0.05, 0.00, 0.00)	0.07	0.19	0.38	0.69	1.00	1.31
(**0.75**, 0.08, 0.08, 0.08) (**0.75**, 0.25, 0.00, 0.00)	0.33	0.42	0.56	0.78	1.00	1.22
(**0.50**, 0.17, 0.17, 0.17)	0.67	0.71	0.78	0.89	1.00	1.11
(**0.30**, 0.23, 0.23, 0.23)	0.93	0.94	0.96	0.98	1.00	1.02

a Effectively, error measures for an infinite series of forecasts where the option assigned the highest probability (first choice) is correct with the frequency shown in the titles. Other options are assumed to be correct in equal proportion

b In some cases, the probabilities displayed do not add to one due to rounding. The number of significant digits used in the error measure calculations, however, was limited only by software and hardware constraints.

Second, for forecasts that would lead a decision maker to anticipate actual outcomes at the rate of chance or less, Brier scores decrease (get better) as the probability assigned to the first choice decreases. If one had to choose between similarly inaccurate forecasting methods, the Brier scores would sensibly lead one to prefer the method that signalled the uncertainty most strongly. This preference is reflected in the Brier scores for forecasts with first choices that are accurate one time in four (chance) or less often. On the other

hand, for forecasts that would lead a decision maker to correctly anticipate actual outcomes more often than by chance, a forecasting method that offered a decision maker confidence (high probability for first choices) ahead of uncertainty (lower probability), would be preferable. For forecasts that are always correct, Brier scores *do* increase (get worse) as the probability assigned to actual outcomes decreases. This ordering is violated, however, for accuracy levels greater than chance but less than 100 percent.

Unlike Brier scores, *PFAR* scores do reward high probabilities for first choices in accurate forecasts and penalise them in forecasts that are less accurate than chance. *PFAR* scores do not distinguish between forecasts that are as accurate as chance – they are all scored as 1.0.

Third, Brier scores are sensitive to the allocation of probabilities to options that fail to occur, while *PFAR* scores are not.

There is another difference between the measures that is not illustrated in Table 47. Brier scores do not directly distinguish between the accuracy of forecasts of situations for which many possible outcomes are defined and those for which few are defined. For example, a forecast for a situation with two defined outcomes, and where one of the outcomes is allocated a probability of 1.0, will earn a Brier score of 2.0 if the forecast is wrong. But so too will such a forecast for a situation with 20 defined outcomes. It is arguable that Brier scores are biased towards forecasting problems with more, rather than fewer, options. As the number of options increases, the opportunity for forecasters to spread probabilities and likelihood they would do so increases. As discussed, the greater the spread of probabilities, the smaller is the Brier-score penalty that is incurred. Thus lower (better) Brier scores are likely to be more common for situations with more outcome options than for situations with fewer. *PFAR* scores, on the other hand, do account for *a priori* difficulty arising from the number of options. This occurs because the denominator in the *PFAR* formula (error from a naïve forecast) increases with the number of options towards an asymptote at 2.0.

I calculated Brier scores and *PFAR* scores from participants' probabilities (Table 49). In order to assess whether using the probability assessments of forecasters would be likely to improve forecast accuracy (in cases where the necessary data were available) I also calculated Brier scores and *PFAR* scores from probabilities derived in two other ways.

The first of these was to set participants' first choices to 1.0, and other options to 0.0. The second was to use the information structured-analogies participants provided about their analogies to derive probabilities using the following rule (one of many possible rules). The probability of each decision is equal to the highest rating given to an analogy that suggests the decision plus one-third of the sum of ratings given to any other analogies that suggest that decision, all divided by the sum of these aggregates across all decisions. Decisions not suggested by any of a participant's analogies are given a probability of 0.0.

For example, assume a fictitious participant used structured analogies to forecast a conflict with three decision options. Table 48 shows the similarity ratings he provided (out of ten) for his three analogies. The ratings are shown in the columns corresponding to the decision options (A, B, or C) suggested by those analogies. The bottom line of the figure shows the probabilities, derived using the rule just described, for the decision options.

Table 48
Deriving probabilities from structured analogies data using a rule

| | Decision options | | | |
	A	B	C	Sum
Analogy 1 rating		7		
Analogy 2 rating			5	
Analogy 3 rating			3	
Probability of decision (derived using rule)	0 /13 = **0.00**	(7 + 0/3) /13 = **0.54**	(5 + 3/3) /13 = **0.46**	(7 + 0/3) + (5 + 3/3) = **13**

Where participants considered that their analogies suggested more than one decision, the ratings were ascribed to each of the decisions that were suggested. In practice, there were never more than two decisions suggested by a single analogy. In such cases, the ratings were counted twice.

Table 49

Brier and *PFAR* scores for cases in which solo experts provided probabilistic forecasts, by derivation[a] of probabilities and forecasting method

	Brier scores					*PFAR* scores				
	One option set to 1.0 [b]		Participants' probabilities		Rule	One option set to 1.0 [b]		Participants' probabilities		Rule
	Unaided judgement	Structured analogies	Unaided judgement	Structured analogies	Structured analogies	Unaided judgement	Structured analogies	Unaided judgement	Structured analogies	Structured analogies
Artists Protest	2.00		1.34			1.20		1.14		
	2.00		1.50			1.20		1.20		
	2.00		1.26			1.20		1.20		
Average	**2.00**		**1.37**			**1.20**		**1.18**		
Average 2										
Distribution Channel			*0.50*					*0.75*		
	2.00	*0.00*	0.78		*0.45*	1.50	*0.00*	1.05		*0.80*
	2.00	*0.00*	1.13		*0.32*	1.50	*0.00*	1.13		*0.60*
		0.50		0.56	1.10		0.75		0.86	1.11
Average	**2.00**	**0.17**	**0.96**	**0.56**	**0.62**	**1.50**	**0.25**	**1.09**	**0.86**	**0.84**
Average 2		**0.50**		**0.56**	**1.10**		**0.75**		**0.86**	**1.11**
55% Pay Plan		2.00		1.61	1.34		1.33		1.33	1.33
	2.00		1.82			1.33		1.33		
Average	**2.00**	**2.00**	**1.82**	**1.61**	**1.34**	**1.33**	**1.33**	**1.33**	**1.33**	**1.33**
Average 2		**2.00**		**1.61**	**1.34**		**1.33**		**1.33**	**1.33**
Personal Grievance		2.00		1.52	2.00		1.33		1.33	1.33
		0.00		0.06	0.54		0.00		0.27	0.80
	0.00		0.32			0.00		0.53		
Average	**0.00**	**1.00**	**0.32**	**0.79**	**1.27**	**0.00**	**0.67**	**0.53**	**0.80**	**1.07**
Average 2		**1.00**	**0.32**	**0.79**	**1.27**		**0.67**		**0.80**	**1.07**
Nurses Dispute	0.00		0.24			0.00		0.60		
		0.00	*0.50*		*0.00*		*0.00*	*0.75*		*0.00*
Average	**0.00**	**0.00**	**0.24**		**0.00**	**0.00**	**0.00**	**0.60**		**0.00**
Average 2										
Telco Takeover	2.00	*2.00*	1.52		*1.51*	1.33	*1.33*	1.33		*1.33*
	2.00		1.36			1.33		1.27		
		2.00		1.50	2.00		1.33		1.33	1.33
	2.00		0.72			1.33		0.80		
		0.50			*1.52*		*0.80*			1.33
Average	**2.00**	**2.00**	**1.20**		**1.76**	**1.33**	**1.33**	**1.13**		**1.33**
Average 2		**2.00**		**1.50**	**2.00**		**1.33**		**1.33**	**1.33**
Water Dispute	0.00		0.24			0.00		0.60		
	2.00		0.88			1.50		1.13		
			0.67					*1.00*		
Average	**1.00**		**0.56**			**0.75**		**0.87**		
Average 2										
Zenith Investment	2.00	0.00	0.78		0.18	1.50	0.00	1.05		0.45
		0.00		0.08	0.00		0.00		0.30	0.00
Average	**2.00**	**0.00**	**0.78**		**0.09**	**1.50**	**0.00**	**1.05**		**0.23**
Average 2		**0.00**		**0.08**	**0.00**		**0.00**		**0.30**	**0.00**
Total [c]						**1.27**	**0.46**	**1.07**		**0.96**
Conflicts, number						8	6	8		6
Observations, number						14	11	14		11
Total 2 [c]							**0.75**		**0.86**	**1.11**
Conflicts, number							5		5	5
Observations, number							6		6	6

Notes:

Figures in italics are excluded from the calculation of "Average 2" and "Total 2" in all cases, and from "Average" and "Total" in some. The criterion for exclusion was lack of a matching figure from the same method but an alternative derivation.

a For a single participant for a single conflict, the probabilities from which the Brier scores and *PFARs* were calculated were derived in three different ways: 1/ see note b; 2/ the participants' own probabilities were used unchanged; 3/ probabilities were derived from participants' analogy decisions and ratings using the rule described in subsection 4.1.1. In all cases, forecasts of C for Distribution Channel were re-coded with 0.5 allocated to A and 0.5 to B

b When a participant allocated the highest probability for a conflict to a single option, that option was re-coded as one and the rest as zero. When participants' probabilities were inconsistent with their own analogies, any forecasts from their probabilities were coded to unaided judgement and, where this was reasonable, single forecasts were derived from the analogies

c Medians of conflict averages.

Table 49 includes only forecasts for which participants provided probabilities for several outcomes. Few such forecasts were provided. Consequently, interpretation of the Table 49 findings can only be tentative – none of the differences are statistically significant. As measured by Brier scores and *PFAR* scores, the relative accuracy of forecasts from unaided judgement and structured analogies is consistent with comparisons of percent correct. That is, structured-analogies forecasts were more accurate. Brier scores suggest that participants' probabilities from unaided judgement and from structured analogies were more accurate than forecasts which set the participants' first choices to 1.0 and other options to 0.0. In contrast, when the comparisons are made using *PFAR* scores, this relationship holds only for unaided-judgement forecasts. The rule for deriving probabilistic forecasts from participants' analogies data appears to offer little or no advantage in accuracy over first-choice-set-to-one or over participants' probabilities.

Table 49 is more interesting for the examples it provides, with real forecast data, of the different conclusions that can be implied by Brier scores and *PFAR* scores. First, the Brier scores for participants' probabilities and rule-based probabilities for one conflict, 55% Pay Plan, are all lower (better) than the first-choice-set-to-one forecasts, which were all completely inaccurate ($BS = 2.0$). The *PFAR* scores for the same set of data were uniformly 1.33 – completely inaccurate. In other words the Brier scores suggest that the first-mentioned two approaches to deriving probabilities were superior to setting the first-choice option to one, whereas the *PFAR* scores suggest they were not. In all of these forecasts, the actual outcome was assigned a probability of 0.0. The difference between them was that the probabilities were spread between two or three options in the case of the forecasts with lower Brier scores.

Second, a similar phenomenon can be seen in the Artists Protest and Telco Takeover figures. For those conflicts, however, non-zero probabilities were assigned to the actual outcomes in the case of some forecasts and so the *PFAR* scores also improve for these, but to a much lesser extent than the Brier scores.

Finally, in the case of the Water Dispute figures, one would conclude from the average Brier scores for the conflict that participants' probabilities were more accurate than the strategy of setting the probability of the first choice to 1.0. One would conclude the opposite from looking at the *PFAR* scores. There are two sets of forecasts in these

averages. Both measures agree that accuracy improves between first-choice-set-to-one and participants' probabilities for one set and declines for the other. The opposite conclusions from the average scores arises because: a) *PFAR* punishes heavily a reduction in the probability assigned to the actual outcome from 1.0 to 0.6 (*PFAR* from 0.00 to 0.60) whereas the Brier score does not (*BS* from 0.00 to 0.24), and b) the Brier score rewards strongly an increase in the probability assigned to the actual outcome from 0.0 to 0.25 (*BS* from 2.00 to 0.88) whereas the *PFAR* does not (1.50 to 1.13).

Appendix 14: Assessment of *a priori* judgements of predictability: Approach and response

Approach

Undergraduate students were told to use their judgement to rate the predictability, on a five-point scale, of each of a set of three conflicts (Figure 2).

Figure 2
A priori predictability rating question

Having read the preceding description of a real (but disguised [a]) conflict, please rate the chances that a knowledgeable forecaster's prediction of the decision that will be made will be an accurate one:

{please tick one box only}

Very likely	Likely	Not sure	Unlikely	Very unlikely
[ ]	[ ]	[ ]	[ ]	[ ]

a All but one of the conflicts (55% Pay Plan) was provided in a disguised form to the raters.

While it is more usual to ask experts than students such questions and it would, therefore, have been more representative to do so, a meta-analysis by Armstrong (1985) showed that expertise, beyond a mere acquaintance with the subject of a forecast, does not result in improved forecasting accuracy. It is reasonable to argue that university students do have an acquaintance with conflicts through experience and formal study. Indeed, in several of the studies examined by Armstrong, the forecasts of students were used as a benchmark against which recognised experts' forecasts fall short. Experts tend to be more confident than novices and, unless that confidence is based on a history of predicting similar cases together with immediate and unambiguous feedback, experts' confidence tends to lead them to ignore important information and consequently to make predictions that are less accurate than those of novices (Arkes, 2001; Fischhoff, 2001). As the conditions for accurate expert forecasts are generally not met in the domain of conflict forecasting, experts ratings may be no better than novices' for this task. This assertion is supported by my findings on the effect of expertise on forecast accuracy (subsection 4.2.1).

I did not provide role information to raters. Neither did I provide the decision options that had been provided to forecasters – these were clear from the description, however, in the case of Zenith Investment. Instead of decision options, the forecasting problem for

each conflict was presented as a single question at the end of each conflict description (Table 50). The question for the raters to answer (Figure 2) was included after this.

Table 50
Forecasting problem for each conflict

Artists Protest:	What will be the final resolution of the artists' sit-in?
Distribution Channel:	Will the management of a supermarket chain accept the CRTP in their stores?
55% Pay Plan:	Will there be a strike and, if so, how long will it last?
Nurses Dispute:	What will the outcome of the negotiation be?
Personal Grievance:	What will the outcome of the 24 April meeting be?
Telco Takeover:	How will the stand-off between Expander and Localville be resolved?
Water Dispute:	Will Midistan agree to release additional water and, if not, how will Deltaland respond?
Zenith Investment:	How many new ACMA plants will the Committee recommend?

Response

The student raters took, on average, eight minutes to rate each conflict for *a priori* predictability. I obtained 10 ratings for four of the conflicts and 11 ratings for the other four. The students used their knowledge of conflict prediction to rate the conflicts. As I was not seeking a representative sample of opinion or behaviour, the validity of the average ratings are not likely to be improved by the addition of more raters (Ashton, 1986; Hogarth, 1978; Libby and Blashfield, 1978). I excluded people who were previously aware of my research from participating in order to avoid the possibility of bias from prior experience with the conflicts.

Appendix 15: Questionnaire for obtaining forecast usefulness ratings

NOTE: Each of the figures in the questionnaire response boxes below is the usefulness rating for the adjacent decision option. These were derived from the survey responses.

Rating forecast usefulness

This rating task is part of a larger project to investigate the accuracy of forecasts from different methods for predicting decisions made by parties in conflict.

Following this note are sets of options that were judged by experts to be the decisions that might have been made in eight real conflicts. (Although they are all real, most of the conflicts are disguised.) Although every effort was made to ensure that the options were complete and mutually exclusive, in some cases the decision that was actually made does not match exactly any single decision option. Further, from a decision-maker's point of view, a forecast need not always be spot-on to be useful. For example, a forecast in August 2001 that attacks on the West's oil supplies by al Qaeda were imminent would not have been strictly accurate but could have been useful had governments and businesses responded by increasing security.

For each conflict, please read the brief description of the actual outcome, then use your judgement to rate (on a scale of zero to 10) each of the decision options that were provided to forecasters. Using the September 11th example, a forecast that al Qaeda would *never* attack targets in the US should be given a score of *zero*, whereas a forecast that, within one month, they would use passenger planes to attack targets in New York and Washington should be given a score of *10*. The "attack on oil supplies" forecast described in the previous paragraph should be given a usefulness score *between* zero and 10. Note that your ratings of the decision options for any one conflict don't have to add to 10, or to any other number. In other words: rate each option independently.

You have about three minutes for each conflict.

1. Personal Grievance

Forecasters were asked: The outcome of the 24 April meeting was?

Actual decision: At the meeting, management agreed to a new evaluation provided it was of trivial cost and covered all service employees. The employees' union arranged this. The new evaluation recommended LOWER salary bands than management had proposed. The bands proposed by management on 14 February were accepted reluctantly by staff as the best deal that was available.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

a. Management agree to a new evaluation being conducted before holding further discussions on salary bands   [**6**]

b. Staff accept the salary bands proposed by management on 14 February with few or any modifications   [**7**]

c. Parties agree to ask a third party (e.g. mediator, independent job-evaluator) to decide on salary bands   [**6**]

d. Parties fail to reach any agreement   [**0**]

Briefly, what are your reasons for your ratings?

2. Nurses Dispute

Forecasters were asked: The outcome of the negotiation was?

Actual decision: A pay deal was struck that had a monetary value half way between the cost of the nurses demand and the value of the employer's offer.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

a. Nurses' demand for an immediate 7% pay rise and a 1-year term was substantially or entirely met [**2**]

b. CCH's offer of a 5% pay rise and a 2-year term was substantially or entirely accepted   [**2**]

c. A compromise was reached   [**8**]

Briefly, what are your reasons for your ratings?

3. Zenith Investment

Forecasters were asked: Which option will the Zenith Policy Committee choose?

Actual decision: The Committee chose to commission two ACMA plants.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

(A)   One ACMA plant   [**5**]

(B)   Two ACMA plants   [**10**]

(C)   No ACMA plants   [**0**]

Briefly, what are your reasons for your ratings?

4. Distribution Plan

Forecasters were asked: Will the management of a supermarket chain accept CRTP in their stores?

Actual decision: The supermarket chain agreed to a long-term arrangement commencing with a one month pilot.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

(A)   Yes, as a long term arrangement   [**10**]
(with one month pilot)
(B)   Yes, as a short term promotion   [**5**]

(C)   Yes, either (A) or (B)   [**7**]

(D)   No, they will reject the plan   [**0**]

Briefly, what are your reasons for your ratings?

5. Water Dispute

Forecasters were asked: The gist of the statement issued at the end of the meeting was?

Actual decision: In the statement, Midistan recognised the plight of the Deltalandish people and undertook to release extra water into Deltaland. Neither war nor dam-bombing occurred.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

  a. Midistan has decided to release additional water in order to meet the needs of the Deltalandish people   [**8**]

  b. Deltaland has ordered the bombing of the dam at Mididam to release water for the needy Deltalandish people   [**0**]

  c. Deltaland has declared war on Midistan   [**0**]

Briefly, what are your reasons for your ratings?

6. Telco Takeover

Forecasters were asked: How was the stand-off between Localville and Expander resolved?

Actual decision: Expander withdrew their bid for all of Localville and agreed to purchase Localville's mobile operation only, which is what Localville had originally proposed.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

  a. Expander's takeover bid failed completely   [**0**]

  b. Expander purchased Localville's mobile operation only   [**10**]

  c. Expander's takeover succeeded at, or close to, their August 14 offer price of $43-per-share   [**0**]

  d. Expander's takeover succeeded at a substantial premium over the August 14 offer price   [**0**]

Briefly, what are your reasons for your ratings?

7. Artists Protest

Forecasters were asked: What will be the final resolution of the artists' sit-in?

Actual decision: The government decided to relax entrance rules for the programme and to allow artists to remain in the programme indefinitely.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

(A)	The government will relax entrance rules and allow an artist to remain in the programme for an indefinite period.	[**10**]
(B)	The government will extend an artist's time in the programme to 2 or 3 years.	[**3**]
(C)	The government will extend the programme 2 or 3 years and relax entrance rules.	[**5**]
(D)	The government will relax the entrance requirements only	[**4**]
(E)	The government will make no change in the programme.	[**0**]
(F)	The government will end the programme completely.	[**0**]

Briefly, what are your reasons for your ratings?

8. 55% Pay Plan

Forecasters were asked: Will their be a strike?

Actual decision: NFL players went on strike for most of the regular season.

Please rate how useful each of these decision options would have been as a forecast

(Usefulness: 0 to 10)

(A)	Yes, a long strike (½ or more of the regular season games will be missed)	[**10**]
(B)	Yes, a medium length strike (less than ½ of the regular season games will be affected)	[**5**]
(C)	Yes, a short strike (only preseason games missed)	[**3**]
(D)	No strike will occur	[**0**]

Briefly, what are your reasons for your ratings?

9. Were you aware of Kesten Green's research on decisions in conflicts prior to reading this material?

Yes [___]   How did you become aware of it?

✎_____

No [___]

Appendix 16: Delphi panellists' ratings of conflict forecasting method criteria

Round 1	**Round 2**

1. Accuracy

Round 1	Round 2
7 Not useful if inaccurate	7 What is important is the aim is accuracy
7 Not sure what the point of a forecasting method is if it is not accurate	7 While there are multiple possible outcomes, forecasting should provide an accurate prediction regarding likeliness of action (e.g., war, strike)
7 If it is not accurate, of what value is it?	7 Accuracy is essential.
7 given the potentially costly nature of the outcome. Clearly, trying to model alternative outcomes would place much less emphasis on accuracy.	7
6 The reason to use this method.	6 Feel comfortable with range of panel responses visa vi mine
6 Important but could not be a 7 as it would be foolish to think that any predicting model could be 100%, so should not rely on it only.	5 I still think it is folly to believe in scientific methods alone
5 Conflicts are by nature dynamic and what constitutes accuracy may assume a single correct option rather than he setting of a range of options.	6 There are more reasons for forecasting than accuracy.

2. Timeliness in providing forecasts

Round 1	Round 2
7 More important than accuracy in that deadlines are liberating while accuracy constricts.	7 time is an absolute
7 If it is not timely then no practical application to it	7 I agree with the other comments around timeliness. once you have the info you can then choose what to do with it. this is empowering
7 given the fact that the outbreak of conflict is not within the control of the party seeking to choose an analytical methodology. There is no use in forecasts that have arrived after the event they are designed to illuminate.	7
6 Not helpful if not delivered in timely manner	5 Maybe deadlines are flexible, but for your conflict examples given above they are probably not. Any way I am persuaded to reduce. If 5.5 was an option I would take that rather than 5
5 Too late is useless, but deadlines are usually flexible.	3 Timeliness is overrated
4 It can wait (usually).	6 Persuaded by too late is useless comment.
4 Could be important if the matter is urgent (e.g., war or strike) but accuracy is more important	6 If the likely outcome is war or strike as in above examples, timeliness would be important to try to prevent such outcomes

3. Cost savings resulting from improved decisions

Round 1	Round 2
7 This seems another important reason to forecast	6 In business settings cost savings are crucial but in other settings I agree that other "costs" may be as or more important
5 Always one of the considerations in business. But not the most important	6 costs are always a factor
4 Costs are seldom the issue. Value is the issue. Driving to cost/benefit ends to favor end costs over up-front costs.	4 issue is value not cost.
4 Nice to know but "power" is usually persuasive.	4 Costs are interesting but probably more of a "selling" point than other
4 Psychological 'saving' as important as cost savings	4 I'm staying with 4 as the human side is as (or really more) important as the money side
3 Usually, not always, cost saving is a cover explanation for something else.	3 Cost is only occasionally central
? not clear on this question: are we talking about improved methodology selection decisions or decisions consequent upon the outbreak of conflict?	

4. Ease of interpretation

Round 1	Round 2
7 Forecasting should provide enhanced description rather than explanation. The greater the lucidity of the description the greater the flexibility of the interpretation.	6 ease of interpretation bends to timeliness
6 Has to be accessible and transparent	6 Access to info is imperative to encourage use and confidence
6 Seems unlikely to be useful if one can't make sense of the forecast	6 Same as original response - must be able to make sense of it
5 Depends on how often and who has to use it.	5 Comfortable with my response as ease only one of many factors.
4 assuming this question refers to interpretation being possible (if costly) as distinct from being difficult regardless of expertise etc (in which case I might go up a notch: but , interpretative difficulty may be a source of analytical richness so...	4
4 Should not be too complex	4 We can train people in interpretation
2 If accuracy is high, my ease is not so important	2 Ease means I can do it if I work harder. So I do.

5. Flexibility

6 Yeah, don't want to add an "attachment" for each different conflict.

6 Related to #6 - the method must be adaptable to the available data

6 Because life situations demand this

5 despite the specificity of usage, I should have thought there was value in using tools that could inform future related (but not identical) problems and also provide a degree of pre-fabricated structure on which to hang one's thinking.

4 I guess it should be flexible, but am unsure of what 'flexibility' means in this context

3 Predicting decisions should set standards that move towards certainty. Variables which enhance flexibility reduce certainty.

? I don't know what this might mean.

6 The predictor must be flexible!

6 Each conflict is so different there must be ability to tailor the model

4 I take on board the comments about certainty and the danger of introducing too much uncertainty

5

5 I think there is confusion here between flexibility of the method, and flexibility of the usage. If method I would stay with 4, if usage I would go to 5

4 I agree that adaptability is the issue and probably synonymous to flexibility in this context.

I think people interpreted this question very differently.

6. Ease in using available data

6 This is ambiguous in that "using" data mixes the idea of information and the idea of data. If you mean using data to construct informed decisions than it's a 6. If ease means access to data than it's a 3.

6 Most businesses would not engage in model use if the process itself required something more complicated than the use of what they already have

6 Yup, don't need hours of research to "dig up" the "right stuff."

6 A good forecasting model should be adaptable to the data

5 Workable ease is required

3 I'd have thought that this shouldn't be over-emphasized since it is the unknown and the hard to predict that will frequently determine these situations: available data will tend to confirm settled views.

2 If accuracy is high, my ease is not so important

6 We got this right.

6 reasons have not changed

6 It must be easy to use on current data or it would not be used.

6 Want to be able to use available data

5 Ease of use is important for busy people

3

2 My ease is not important.

7. Ease of use

7 A reality of business

7 I'm basically as lazy as the next guy

6 simply because in a serious situation you don't want time wasted learning on the job.

4 Workable ease required (but unsure what is difference in this and Q6

3 If it is accurate a specialist should be able to take the time to learn how to use it

2 If accuracy is high, my ease is not so important

? I am unclear what this refers to

6 I think other factors will influence this

5 I'm persuaded that accuracy a specialist will overcome any "ease of use questions.

6

5 Agree don't want to waste time

4 Leaning slightly more in favor of need for low learning curve, but still believe this can be overcome by having a specialist armed with this knowledge.

2 If accuracy is high, my ease is not so important

3 While of some importance, relative to the other variable this rates a 3.

8. Ease of implementation

7 A reality of business

7 organizing time and cost is an essential piece as well as understanding opportunity costs of not implementing.

5 If this implies that it is harder for other people, it may make it less useful.

5 Not sure if this is outcome or process?

4 Again - all else being equal this seems less important than accuracy, etc.

3 Not so critical

? not sure how implementation of a method differs from use.

7 I think other factors will influence this

6 I accept that ease demands "dumbing-down" and as a goal short circuits the other variable.

7 If accuracy is high, my ease is not so important

6 This usually demands "others" cooperation and therefore must be reasonable to them to implement - I'll stand pat.

7 Again, leaning more toward practicality for business purposes but still don't think this is highest priority

6 OK I guess this is more important as per Q7

4

9. Ability to incorporate judgemental input

7 dynamic systems are always more accessible to input than static systems.

7 I don't believe scientific approach alone is good or "true"

7 It is important that a third party not only comment on the likely outcome but also indicate whether this is an appropriate, just, effective outcome or if there is another intervention process that would likely result in a better outcome

7 Provided it's transparent.

6 Almost any method in this field has to allow for this.

5 I probably would in any case anyway.

4 Would be helpful

7 The human factor of contextual judgment is paramount.

7 reasons have not changed

7 Same as above answer - want to get at underlying interests and determine an outcome best for everyone, not just a "right or wrong" answer

7

6 Judgment is essential in this field.

6 Yup, I'm persuaded that judgement is more important than I first thought.

5 Its a question of what is 'judgemental'. I wouldn't want to increase uncertainty

10. Reliability of confidence intervals

7 Like accuracy, good reliability is critical

7 clearly much more important that absolute values

7 Otherwise, I would suspect the instrument each time - not good!

5 Sure.

4 Guessing only

4 Often too statistically static which gives a false positive.

?

7 Same as earlier, want good reliability

7

6 I will also yield from the "extreme" position toward a bit more flexible one.

5

6 Maybe I missed the point first time?

4 This is still ambiguous depending on whether one is referring to a statistical manipulation or level of judgmental reliability.
not sure if I want to comment on this

11. Development cost (computer, human resources)

7 The greater the up-front cost, the more secure in the forecasting variables.

6 Yes important as if conflicts are not common then investing in development costs may not be seen to be value for money

6 Low of course!

5 One would like to downgrade this (faced with a serious breakdown) but the reality is that costs are a real factor.

4 I guess must be considered to get it done

3 Just seems less important than flexibility, accuracy, etc. if it results in long-term cost savings

? If resources are available, not important; if not, then important.

7 You can't dance if no one pays for the band.

6 no change

6 Close enough!

5

4 Ambiguity sets in... My rating is made not on the actual quantum of development cost required, but on how "impactful" this would be on going ahead

5 Once again thinking more practically for corporate settings

3 Usually manageable.

12. Maintenance cost (data storage, modifications)

6 This rating depends on whether or not re-use/reinterpretation is envisaged. If not, lower.

5 If worthwhile then this would not be such an issue. Flows from position on above

5 Everything is treadmill update today - this is probably no different.

5 No one will use it if has high costs to maintain, harder to justify

3 an externality that is difficult to compute

2 If all else is good, will find $ for maintenance

2 If resources are available, not important; if not, then important.

6

5 no change

5 All answers helpful but I'll stay where I am.

5 Same as earlier response

4 this cost is inversely related to developmental costs.

3 Okay I'll come up a bit, but in many situations the maintenance cost just is a reality and not such a big factor if using something that is of proven value

2

13. Theoretical relevance

7 Theory grows from practice not the other way around. Predictions are too often caste based on the idiosyncratic life histories of a small group of practitioners. Prediction should be inductive.

6 it is, isn't it?

4 Doesn't seem like a primary goal but of course it would be nice!

4 I'd still say this was reasonably important since theoretical parameters provide one useful set of benchmarks

4 Sufficient if it works

3 In business I think this is less considered

3 for my work not so important

7 Every conflict worth its salt has theoretical relevance.

6 Still think it must be "sound" to be accepted.

4 Same

4

4 Its of moderate relevance - more important is does it work

4 while business maybe less concerned about this, credibility in the wider academic community would require this

3 For me, not important. Others will differ.

241

14. Ability to compare alternative policies

7 Should be able to compare different outcomes that might result from different solutions, conflict management methods - show people options

7 This is vital for the exercise of judgemental input.

6 Alternatives provide the referent point for difficult decisions. The comparison with unformed policies is even better.

6 A really nice feature - but not essential if the primary purpose works.

6 All this work is comparing alternatives.

5 While important not sure business always takes this approach

4 Helpful

7 Still think looking at options is important

7

6 If this is correct then our theoretical relevance score is naively low. You can't get to 14 without 13.

6 The "what if" is vital.

6 Accuracy is central, and comparing is essential to determining accuracy.

4 agree not the primary purpose

4 Comparison can be helpful, but its not the raison d'etre

15. Ability to examine alternative environments

7 Environmental context or conflict arenas have a higher predictivity than the weighing of the conflict skills of the participants.

7 Many similar types of conflicts occur in different contexts, this would be very valuable

6 While important not sure business always takes this approach

6 A really nice feature - but not essential if the primary purpose works.

4 Depends on likelihood of re-use I suppose

3 Not so necessary

? I'm not sure enough that I know what this means.

7 The power of any analysis is based on its utility versus salient alternatives.

6 May not be as important as originally thought since each user will only be concerned with one environment

6 maybe here is where the flexibility would be useful

6 Not useful to me now - I'll stay.

4

4 Useful but not an essential ingredient

16. Ability to learn (experience leads forecasters to improve procedures)

7 Should lead to improved future conflict management - what we are all striving for.

7 I assume this is part of accuracy.

6 Should be an ideal in any business

6 Hopefully, it's easier to learn/retain than Microsoft Access!

5 All things being equal, the choice which allows for expansion of interpretation of prior or anticipated conditions is desirable.

5 Must learn from what do

4 This depends very much on the field. The less likely it is that circumstances will re-produce themselves, the less important - if learning (as your parentheses imply) is about adapting the methodology for further use.

7 Still think the goal of learning from previous conflicts is crucial

7

7 it should always be about learning!!

6 No reason to change

6 The question is how to gather , interpret, then disseminate the learned experience.

6 I agree its very important to be able to learn

4

Appendix 17: Delphi panellists' ratings of forecasting methods against criteria

<table>
<tr><td colspan="5">Round 1</td><td colspan="5">Round 2</td></tr>
</table>

A. Accuracy

SI	UJ	SA	GT		SI	UJ	SA	GT	
9	6	3	2	The cases I think of contain too many dimensions for game theory; SI, I am learning from you, may be excellent for them; I can't think of analogies that capture very much useful; and I used to think UJ (my way) was more useful than your statistics seem to bear out.	9	4	2	2	
7	3	5	4	That is the research outcomes you have identified	7	3	5	4	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
7	3	3	5	Seems consistent with the findings of the study described	7	3	3	3	GT and SA seem very limited, SI seems best based on your research results. UJ seems like educated guessing.
7	4	4	5	SI offers opportunity to place oneself within the situation - yielding better insight and decision.	6	4	4	5	Additional information is helpful in determining modified responses.
6	0	3	0	0 on the basis that something no better than chance is inadequate (i.e. adequacy is something that is better than chance)	6	0	3	0	Didn't participate in round 2. Round 1 responses pasted to include in final averages.

B. Ability to incorporate judgemental input

SI	UJ	SA	GT		SI	UJ	SA	GT	
10	10	10	7	I assume GT is more problematic here since once you've decided on your decision rules you're stuck with them (but then I suppose you can make allowance for that - sorry, not well enough informed here)	10	10	10	7	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
9	9	5	2	SI and UJ are the most flexible for this.	4	9	5	1	
8	10	10	2	GT seems most interested in probable outcomes without commentary. In SI it seems you could get judgments from participants which may be relevant. Assuming one can find an analogous situation, it would be easy to discuss the advantages and disadvantages of what had happened and how it could have been improved, so this and UJ seem most conducive to judgmental input.	10	10	10	2	Same as earlier with realization that SI is also very conducive
8	10	8	5	a less structured approach is more responsive	8	10	8	5	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
5	10	5	5	Unaided judgement is most able (consciously or unconsciously).	5	10	7	5	Modify my responses slightly due to others ratings.

C. Ability to learn (experience leads forecasters to improve procedures)

SI	UJ	SA	GT		SI	UJ	SA	GT	
10	10	10	10	I am assuming that experience with any of the methods would be beneficial and improve future performance.	9	9	9	9	Other arguments persuade me to be less optimistic, but I still believe any experience should result in learning
8	8	5	8	Again my instinct is that three of these lend themselves to easy improvement whereas SA leaves forecasters in the hands of the knowledge and experience of domain experts and the existence (or not) of analogies.	8	8	5	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
6	8	8	5	maybe the higher level of participation in the process allows for greater learning	6	8	8	5	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
5	9	7	5	It would seem that the unaided judge is well placed to learn and learn quickly - however, The unaided judges would think they can compare - this might be a study in itself.	6	9	7	6	OK, slightly more due to others input.
2	9	8	9	SI, as I understand it, has no memory as there are new players each time.	7	7	7	8	

D. Timeliness in providing forecasts

SI	UJ	SA	GT		SI	UJ	SA	GT	
8	8	8	8	I can't see that any of these methods is particularly time-consuming (even assembling people for SI)	8	8	8	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
6	6	5	5	There was no discussion of how long these take, but it seems that all require some time and effort to collect information or opinions or allow for setting up and enactment of a simulation	5	7	8	5	Have to agree with others that GT and SI would take longer, if analogy exists it would be quick, UJ could also be done more quickly but not sure it is that important if not accurate
5	9	8	1	where there is less need for external assistance there is a quicker response	5	9	8	1	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
4	6	6	8	The fewer steps to accomplish the quicker the results.	4	6	6	8	Not persuaded to modify my responses from data given.
2	8	8	2	For GT one has to know too much before one can predict; how does one simulate before knowing the later scenes; if analogies work, they work early; ditto, UJ.	3	9	7	7	

E. Reliability of confidence intervals

SI	UJ	SA	GT		SI	UJ	SA	GT	
8	2	4	8	Pre-structure will aide outcomes.	8	2	4	8	I'll stay at the "high" scores
6	0	3	0	I assume this correlates with 1 above	6	0	3	0	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
?	?	?	?	Don't feel comfortable rating this one	7	3	3	3	Rated same as accuracy based on earlier response and your research report.

F. Ability to compare alternative policies

SI	UJ	SA	GT		SI	UJ	SA	GT	
9	7	9	10	Presumably once could simulate the conflict with several policies and compare the results, seems most difficult to do in UJ	9	8	9	10	Same as earlier, except that if the same person is doing the UJ they may have enough experience to compare alternatives as well
8	8	4	8	I assume SA is going to be limited by the search for useful analogies	8	8	4	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
7	5	5	7	The unaided judgement would think they can compare - this might be a study in itself.	7	5	5	7	No reason to change
7	4	7	8		7	4	7	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
2	2	2	9	The GT will lend itself most readily.	3	6	6	8	

G. Ability to examine alternative environments

SI	UJ	SA	GT		SI	UJ	SA	GT	
10	6	8	10	GT and SI seems to lend themselves to that since there is control over the set up of the scenario/game/simulation. Again SA requires finding an appropriate analogy and UJ would be harder unless someone had experience in many different environments (or access to advisors who do)	10	7	8	10	Same as earlier, except that if the same person is doing the UJ they may have enough experience to compare alternatives as well
6	4	4	6	Very hard to compare "fights" to "disputes."	6	4	4	6	I don't share the confidence which others seem to find.
5	4	6	4		5	4	6	4	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
?	?	?	?	Sorry I find it very hard to place meaningful values here					

H. Development cost (computer, human resources)

SI	UJ	SA	GT		SI	UJ	SA	GT	
9	2	5	2	Again the role play is most time consuming to prepare.	9	6	7	2	
7	10	7	7	All seem to require significant time or resources except UJ	7	10	7	7	Same as earlier
6	9	8	4		6	9	8	4	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
6	3	5	6	Unaided judgements appear to cost least in terms of development costs.	6	3	5	6	No reason to change
2	10	7	2	self explanatory	2	10	7	2	Didn't participate in round 2. Round 1 responses pasted to include in final averages.

I. Flexibility

SI	UJ	SA	GT		SI	UJ	SA	GT	
10	10	2	4	Again, simulations can always be created and unaided judgment can always be done, but the others require locating an analogous situation or making it fit the game theory model	10	10	6	4	Same as earlier - SA probably be more flexible than I originally thought since using that info to interpret a new situation
9	9	7	5	GT at 5 because of the difficulty in describing all situations in terms of decision rules	9	9	7	5	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
9	9	9	2	The virtue of GT is rigor, ie its inflexibility.	9	6	7	2	
8	5	7	0	I think judgement will be dependent on the individuals ability to think laterally SI, SA because of their approach will allow greater variation and responsiveness	8	5	7	0	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
5	7	7	5	Most flexible when only one mind is involved.	5	7	7	5	Believe mine to be effective responses in circumstances.

J. Ease in using available data

SI	UJ	SA	GT		SI	UJ	SA	GT	
10	10	2	4	Again, simulations can always be created and unaided judgment can always be done, but the others require locating an analogous situation or making it fit the game theory model	10	10	4	4	Same as earlier, again I am rating SA a bit more favorably as I see a bit greater potential than originally conceived
9	9	5	2	I like to write role plays so turning data into a role play is relatively easy.	6	6	6	9	
8	8	7	5		8	8	7	5	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
6	6	6	6	None are perfect - each has advantages.	6	6	6	6	Not persuaded to change responses.

K. Ease of implementation

SI	UJ	SA	GT		SI	UJ	SA	GT	
8	8	8	8	Can't see any reason why any of these should be difficult given appropriate expertise	8	8	8	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
7	7	2	7	GT seems complicated to me, but I suppose if you have an expert on hand it would be easier to implement (or just as easy) as the others. SA seems most difficult because of the need to find an analogous situation. SI and UJ are rated lower here because they both require cooperation of others which may not always be easy in actual implementation.	7	7	4	7	Same as earlier with again a bit more of a nod for SA for reasons discussed above
4	7	4	5	An individual decision maker has the easiest time of it.	6	8	5	7	Believe my scores could be increased slightly based upon others.
2	9	5	8	The simulation is the most complex to prepare and stage.	3	7	7	9	

L. Ease of interpretation

SI	UJ	SA	GT		SI	UJ	SA	GT	
10	10	4	5	Seems that the results of SI or UJ would be easy to interpret, while GT is more complicated and SA seems difficult because it may be hard to find truly analogous situations	10	10	7	5	Same as earlier except that interpreting SA would probably be easy assuming availability of an analogous situation
8	8	6	8	SA seems to me the only one that poses some problems since analogues are never going to be exact so there is an ambiguity that won't exist with the precise spuriousness of GT or the precise outcomes of the role play or unaided judgements.	8	8	6	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
5	6	5	2	I am assuming business context here. so i suppose the more direct methods will be deemed better	5	6	5	2	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
3	7	7	3	Less "expertism" = easier interpretations.	3	7	7	3	Comfortable with initial ratings.
2	8	2	9		7	5	4	9	

M. Theoretical relevance

SI	UJ	SA	GT		SI	UJ	SA	GT	
7	7	7	10	Only GT is directly derived from theory, others could involve and certainly could inform theory but hard to say whether they would	7	7	7	10	Still believe GT is most obviously theoretically relevant...others could have theoretical implications if data is collected and analyzed in a way that permits generalizations to be made. Keeping records of how forecasting methods have proved accurate, effective, etc. could certainly be relevant to whether theoretical relevance results from their use.
6	2	5	6	The "purer" the process - the more theoretical (I would suppose).	6	2	5	6	My scores are "accurate enough."
4	5	7	7	GT & SA high as both will be useful in developing predictive models, even if only applicable in specific incidents. SA probably more useful in a broader application	4	5	7	7	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
2	2	5	9	The GT will lend itself most readily.	7	7	7	9	
?	?	?	?	Can't judge the level but I'd put them all the same except UJ					

N. Cost savings resulting from improved decisions

SI	UJ	SA	GT		SI	UJ	SA	GT	
8	3	2	2	Only SI had good enough accuracy to result in much cost savings	7	2	2	2	Still feel that SI is most economical, especially since it is also most accurate; if forecasting is accurate there shouldn't be much cost savings
6	0	3	0	I assume savings correlate with accuracy since in the absence of accuracy there will be no impact on the quality of decisions.	6	0	3	0	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
3	6	6	3	Limited time invested (total)yields cost savings.	3	6	6	3	Answer spread too wide - leave my answers as is.
3	3	3	3	not likely as incidence of predicted outcome unreliable	3	3	3	3	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
2	8	8	8	SI is relatively expensive; the others are relatively cheap.	8	2	2	2	

O. Ease of use

SI	UJ	SA	GT		SI	UJ	SA	GT	
8	8	6	8	This assumes availability of experts for GT and SA	8	8	6	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
8	10	2	4	Again, simulations can always be created and unaided judgment can always be done, but the others require locating an analogous situation or making it fit the game theory model, but SA is not particularly easy since you have to set up the simulation and find participants, etc. UJ seems easiest to use.	5	10	5	4	SI could be a bit more complicated to arrange, SA not as difficult if analogy is readily available
4	9	8	4	An individual has the upper hand with "ease."	4	9	8	4	Believe my scores are accurate enough.
2	9	5	8	The simulation is the most complex to prepare and stage.	7	6	7	9	
2	8	7	0	again less complexity in setting up will rate these approaches higher	2	8	7	0	Didn't participate in round 2. Round 1 responses pasted to include in final averages.

P. Maintenance cost (data storage, modifications)

SI	UJ	SA	GT		SI	UJ	SA	GT	
7	10	7	7	All seem to require significant time or resources except UJ	7	10	7	7	Same as earlier - all seem to require time of experts or resources except UJ
6	2	4	6	Models must be developed/maintained and administrated.	6	6	6	6	Persuaded to "up" my scores slightly.
5	8	8	8	I assume SI can't be stored since its reliant on mobile human capital that can't always be reassembled	5	8	8	8	Didn't participate in round 2. Round 1 responses pasted to include in final averages.
?	?	?	?	not sure what maintenance costs would be with these models					

References

Adam, H., & Moodley, K. (1993). Forecasting scenarios for South Africa. *Futures*, *25*(4), 404-413.

Anastasakis, L., & Mort, N. (2001). The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH). Research Report No. 813, Department of Automatic Control & Systems Engineering, The University of Sheffield, United Kingdom.

Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 495-515.

Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, *83*(June/July), 18-24.

Armstrong, J. S. (1985). *Long-range forecasting*. New York: John Wiley.

Armstrong, J. S. (1987). Forecasting methods for conflict situations. In Wright, G., & Ayton, P. (Eds.), *Judgmental Forecasting*. Chichester: Wiley, 157-176.

Armstrong, J. S. (1991). Prediction of consumer behaviour by experts and novices. *Journal of Consumer Research*, *18*, 251-256.

Armstrong, J. S. (1997). Why can't a game be more like a business?: a review of Co-opetition by Nalebuff and Brandenburger. *Journal of Marketing*, *61* (April), 92-95.

Armstrong, J. S. (2001a). Role playing: A method to forecast decisions. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 15-30.

Armstrong, J. S. (2001b). Judgmental bootstrapping: inferring experts' rules for forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 171-192.

Armstrong, J. S. (2001c). Selecting forecasting methods. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 365-386.

Armstrong, J. S. (2001d). Combining forecasts. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 417-439.

Armstrong, J. S. (2001e). Evaluating forecasting methods. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 443-472.

Armstrong, J. S. (2001f). Standards and practices for forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 679-732.

Armstrong, J. S. (2001g). The forecasting dictionary. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 761-824.

Armstrong, J. S. (2002). Assessing game theory, role playing, and unaided judgement. *International Journal of Forecasting*, *18*, 345-352.

Armstrong, J. S., & Brodie, R. J. (1999). Forecasting for marketing. In Hooley, G. J., & Hussey, M. K. (Eds.), *Quantitative methods in marketing*, 2nd ed.. London: International Thompson Business Press, 92-119.

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., & Collopy, F. (1998). Integration of statistical methods and judgement for time series forecasting: principles from empirical research. In Wright, G. & Goodwin, P. (Eds.), *Forecasting with judgement*, Chichester: John Wiley, 269-293.

Armstrong, J. S., & Hutcherson, P. D. (1989). Predicting the outcome of marketing negotiations. *International Journal of Research in Marketing*, *6*, 227-239.

Armstrong, J. S., & Walker, H. S. (1983). Validation of role playing as a predictive technique for conflict situations. *World Future Society Bulletin, 17*(4), 15-22.

Ashton, A. H. (1986). Combining the judgments of experts: how many and which ones? *Organizational Behavior and Human Decision Processes*, *38*, 405-414.

Austen-Smith, D., & Banks, J. S. (1998). Social choice theory, game theory, and positive political theory. *Annual Review of Political Science, 1*, 259-287.

Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review*, *85*(5), 1337-1343.

Batson, C. D., & Ahmad, N. (2001). Empathy-induced altruism in a prisoner's dilemma II: what if the target of empathy has defected? *European Journal of Social Psychology*, *31*(1), 25-36.

Bazerman, M. H. (1998). *Judgment in managerial decision making*, 4th ed.. New York: Wiley.

Beatty, R. P., Riffe, S. M., & Thompson, R. (1999). The method of comparables and tax court valuations of private firms: an empirical investigation. *Accounting Horizons, 13*(3), 177-199.

Beer, J. d. (c2000). Dealing with uncertainty in population forecasting. Department of Population, Statistics Netherlands. Retrieved December 11, 2002, from www.cbs.nl/nl/publicaties/publicaties/maatschappij/ bevolking/papers/dealing-with-uncertainty.pdf.

Bennett, P. G. (1995). Making decisions in international relations: game theory and beyond. *Mershon International Studies Review*, *39*, 19-52.

Bennett, P. G., & Huxham, C. S. (1982). Hypergames and what they do: a "soft O.R." approach. *Journal of the Operational Research Society*, *33*(1), 41-50.

Bennett, P. G., & McQuade, P. (1996). Experimental dramas: prototyping a multiuser negotiation simulation. *Group Decision and Negotiation*, *5*, 119-136.

Berg, T. L. (1970). *Mismarketing: case histories of marketing misfires.* New York: Doubleday, 87-131.

Bernstein, S., Lebow, R. N., Stein, J. G., & Weber, S. (2000). God gave physics the easy problems: adapting social science to an unpredictable world. *European Journal of International Relations*, *6*(1), 43-76.

Binmore, K. (1990). *Essays on the foundations of game theory*. Cambridge, MA: Basil Blackwell.

Blume, A., DeJong, D. V., Kim, Y. G., & Sprinkle, G. B. (2001). Evolution of communication with partial common interest. *Games and Economic Behavior, 37*(1), 79-120.

Bolton, G. E. (2002). Game theory's role in role-playing. *International Journal of Forecasting*, *18*, 353-358.

Bowander, B. Muralidharan, B., & Miyake, T. (1999). Forecasting technological change: insights from theories of evolution. *Interdisciplinary Science Reviews*, *24*(4), 275-288.

Boyle, R. H. (1982). The 55% solution. *Sports Illustrated*, 1 February, 30.

Brams, S. J., & Togman, J. M. (2000). Agreement through threats: the Northern Ireland case. In Miroslav, N., & Lepgold, J. (Eds.), *Being useful: policy relevance and international relations theory*. Ann Arbor, MI: University of Michigan Press: 325-342.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1-3.

Bueno de Mesquita, B. & Stokman, F. N., (1994). Models of exchange and of expected utility maximisation: a comparison of accuracy. In Bueno de Mesquita, B., & Stokman, F. N. (Eds.), *European community decision making: models, applications, and comparisons*. New Haven, CN: Yale University Press, 214-228.

Bullock, A., & Trombley, S. (Eds.), (1999). *The new Fontana dictionary of modern thought*, 3rd ed.. London: Harper Collins.

Cannon, W. T., & Reed Consulting Group. (1999). The appropriate return on equity for the Transco and Disco Business operations of the Ontario Hydro Services Company. The Ontario Energy Board Staff. Retrieved 11 December, 2002, from http://www.oeb.gov.on.ca/cases/RP-1998-0001/ROEsummary.doc.

Carayannis, E. G., & Alexander, J. (2001). Virtual, wireless mannah: a co-operative analysis of the broadband satellite industry. *Technovation, 21* (12), 759-766.

Castro, D. C., Lubker, B. B., Bryant, D. M., & Skinner, M. (2002). Oral language and reading abilities of first-grade Peruvian children: associations with child and family factors. *International Journal of Behavioral Development, 26*(4), 334-344.

Chambers, J. C., Mullick, S. K., & Smith, D. D. (1971). How to choose the right forecasting technique. *Harvard Business Review, July/August*, 44-69.

Chang, S-C. (1999). A study on traffic-flow forecasting using time-series analysis and artificial neural network: the application of judgmental adjustment. Thesis for Master of Science Degree, Department of Mechatronics, Kwangju Institute of Science and Technology. Retrieved 11 December, 2002, from moon.kjist.ac.kr/papers/ms_thesis (changsc).pdf.

Collopy, F., Adya, M., & Armstrong, J. S. (2001). Expert systems for forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 285-300.

Cyert, R. M., March, J. G., & Starbuck, W. H., (1961). Two experiments on bias and conflict in organisational estimation. *Management Science*, *7*, 254-264.

Diekmann, A. (1993), Cooperation in an asymmetric volunteers dilemma game: theory and experimental evidence. *International Journal of Game Theory*, *22*(1), 75-85.

Dixit, A., & Skeath, S. (1999). *Games of strategy*. New York: Norton.

Doggett, K. (1998). Glossary of verification terms (revised June, 1998). National Oceanic and Atmospheric Administration. Retrieved November 13, 2002, from http://www.sel.noaa.gov/forecast_verification/verif_glossary2.html.

Drabble, M. (Ed.), (1995). *The Oxford companion to English literature*, 5th ed., revised. Oxford, England: Oxford University Press.

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*(May 1977), 119-127.

Eliashberg, J., LaTour, S. A., Rangaswamy, A., & Stern, L. W. (1986). Assessing the predictive accuracy of two utility-based theories in a marketing channel negotiation context. *Journal of Marketing Research*, *23*, 101-110.

Erev, I., Roth, A. E., Slonim, R. L., & Barron, G. (2002). Predictive value and the usefulness of game theoretic models. *International Journal of Forecasting*, *18*, 359-368.

Feder, S. A. (1987). Factions and Policon: new ways to analyze politics. *Studies in Intelligence*, *31*(1), 41-57.

Fischhoff, B. (2001). Learning from experience: coping with hindsight bias and ambiguity. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 543-554.

Fraser, N. M. (1986). Political and social forecasting using conflict analysis. *European Journal of Political Economy*, *2*(2), 203-222.

Fraser, N. M., & Hipel, K. W. (1984). *Conflict analysis: models and resolutions*. New York: North-Holland.

Fuller, S. (2000). Verification: probability forecasts. *NWP Gazette, December 2000*. Retrieved November 10, 2002, from http://www.met-office.gov.uk/research/nwp/publications/nwp_gazette/dec00/verification.html.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.), (2001). *The analogical mind: perspectives from cognitive science*. Cambridge, MA: Bradford Books.

Ghemawat, P., & McGahan, A. M. (1998). Order backlogs and strategic pricing: the case of the US large turbine generator industry. *Strategic Management Journal*, *19*(3), 255-268.

Ghosh, M., & John, G. (2000). Experimental evidence for agency models of salesforce compensation. *Marketing Science*, *19*(4), 348-365.

Gibbons, R., & Van Boven, L. (2001). Contingent social utility in the prisoners' dilemma. *Journal of Economic Behavior and Organisation*, *45*(1), 1-17.

Glantz, M. H. (1991). The use of analogies in forecasting ecological and societal responses to global warming. *Environment*, *33*(5), 11-15 & 27-33.

Gonzalez, S. (2000). Neural networks for macroeconomic forecasting: a complementary approach to linear regression models. Working Paper 2000-07, Department of Finance, Canada. Retrieved 11 December, 2002, from www.fin.gc.ca/wp/2000-07e.pdf.

Goodwin, P. (2002). Forecasting games: can game theory win? *International Journal of Forecasting*, *18*, 369-374.

Graef, R. (1976). Decision Steel. Grenada Colour Productions.

Graham, L. D. (1991). Predicting academic success of students in a master of business administration program. *Educational and Psychological Measurement*, *51*(3), 721-727.

Green, K. C. (2001). The effect of mediation: Capital Coast Health nurses pay dispute. Decision Research Limited, unpublished report commissioned by the Employment Relations Service of the New Zealand Department of Labour.

Green, K. C. (2002a). Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement. *International Journal of Forecasting*, *18*, 321-344.

Green, K. C. (2002b). Embroiled in a conflict: who do you call?. *International Journal of Forecasting*, *18*, 389-395.

Green, K. C. (2002c). The effect of mediation and information: a personal grievance. Decision Research Limited, unpublished report commissioned by the Employment Relations Service of the New Zealand Department of Labour.

Gruca, T. S., Kumar, K. R., & Sudharshan, D. (1992). An equilibrium-analysis of defensive response to entry using a coupled response function model. *Marketing Science*, *11*(4), 348-358.

Haddad, C. (2001, September 3). The telecom small fry that ate the boonies. *Business Week Online*. Retrieved May 3, 2002, from http://www.businessweek.com/magazine/content/01_36/b3747057.htm.

Hargreaves Heap, S. P., & Varoufakis, Y. (1995). *Game theory: a critical introduction*. New York: Routledge.

Harvey, N. (2001). Improving judgement in forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 59-80.

Henderson, H. (1998). Viewing "the new economy" from diverse forecasting perspectives. *Futures*, *30*(4), 267-275.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, *21*, 40-46.

Howard, N. (1994a). Drama theory and its relation to game theory. Part 1: dramatic resolution vs. rational solution. *Group Decision and Negotiation*, *3*, 187-206.

Howard, N. (1994b). Drama theory and its relation to game theory. Part 2: formal model of the resolution process. *Group Decision and Negotiation*, *3*, 207-235.

Jehiel, P. (1998). Repeated games and limited forecasting. *European Economic Review*, *42*(3-5), 543-551.

Johnson, G., & Scholes, K. (2002). *Exploring corporate strategy*, 6th ed.. Harlow, Essex, UK: Pearson Education.

Kadoda, G., Cartwright, M., & Shepperd, M. (2001). Issues on the effective use of CBR technology for software product development. *Case-based reasoning research and development, Proceedings, 2080*, 276-290.

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: biases and corrective measures. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgement under uncertainty: heuristics and biases.* New York: Cambridge University Press, 414-421.

Keesing's Contemporary Archives (1975). Iraq: Syria-Iraq dispute. August 18-24, 27284-27285.

Keser, C., & Gardner, R. (1999). Strategic behavior of experienced subjects in a common pool resource game. *International Journal of Game Theory*, *28*(2), 241-252.

Kharif, O. (2001, August 21). A small town vs. a very big deal. *Business Week Online*. Retrieved May 3, 2002, from http://www.businessweek.com/bwdaily/dnflash/aug2001/nf20010821_745.htm.

Khong, Y. F. (1992). *Analogies at war: Korea, Munich, Dien Bien Phu, and the Vietnam decisions of 1965*. Princeton NJ: Princeton University Press.

Kirshenbaum, J. (1982). Right destination, wrong track. *Sports Illustrated*, 1 February, 7.

Kliot, N. (1994). *Water resources and conflict in the Middle East*. London: Routledge.

Kolodner, J. L. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann.

Kungliga Vetenskapsakademien The Royal Swedish Academy of Sciences (1994, October 11). *Press release: The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel for 1994*. Retrieved September 16, 2002, from http://www.nobel.se/economics/laureates/1994/press.html.

Langdon, C. (2000a). Nurses vote today on strike. *The Dominion*, Edition 2, 20 September, 3.

Langdon, C. (2000b). Nurses support call for strike. *The Dominion*, Edition 2, 21 September, 3.

Langdon, C. (2000c). Nurses' pay boosted, strike off. *The Dominion*, Edition 2, 6 December, 3.

Lawson, C. T. (1998). Household travel/activity decisions. Dissertation for PhD in Urban Studies, Portland State University. Retrieved 11 December, 2002, from www.upa.pdx.edu/CUS/publications/docs/SR035.pdf.

Leeflang, P. S. H., & Wittink, D. R. (2000). Building models for marketing decisions: past, present and future. *International Journal of Research in Marketing, 17*(2-3), 105-126. Retrieved December 10, 2002, from www.ub.rug.nl/eldoc/som/f/00F20/00f20.pdf.

Libby, R., & Blashfield, R. K. (1978). Performance of a composite as a function of the number of judges. *Organizational Behavior and Human Performance*, *21*, 121-129.

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: the state of the art to 1980. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgement under uncertainty: heuristics and biases*. New York: Cambridge University Press, 306-334.

London, S. (2002). Games or serious business?. *Financial Times*, 26 March, 16.

Liu, J. H., Pham, L. B., & Holyoak, K. J. (1997). Adjusting social inferences in familiar and unfamiliar domains: the generality of response to situational pragmatics. *International Journal of Psychology*, *32*(2), 73-91.

MacGregor, D. G. (2001). Decomposition for judgmental forecasting and estimation. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 107-123.

Mazlish, B. (Ed.) (1965). *The railroad and the space program: an exploration in historical analogy*. Cambridge, MA: M.I.T. Press.

McAfee, P., & McMillan, J. (1996). Analyzing the airwaves auction. *Journal of Economic Perspectives, Winter, 8*, 159-175.

McCabe, K. A., & Smith, V. L. (2000). A comparison of naïve and sophisticated subject behavior with game theoretic predictions. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(7), 3777-3781.

McCarthy, B. (2002). New economics of sociological criminology. *Annual Review of Sociology, 28*, 417-442.

Mentzas, G. (1997). Intelligent process support for corporate decision making. *Journal of Decision Systems, 6*(2), 117-138.

Mildenhall, P. T., & Williams, J. S. (2001). Instability in students' use of intuitive and Newtonian models to predict motion: the critical effect of the parameters involved. *International Journal of Science Education, 23*(6), 643-660.

Morwitz, V. G. (2001). Methods for forecasting from intentions data. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 33-56.

Nalebuff, B. J., & Brandenburger, A. M. (1996). *Co-opetition*. London: Harper Collins.

Neslin, S. A., & Greenhalgh, L. (1983). Nash's theory of cooperative games as a predictor of the outcomes of buyer-seller negotiations. *Journal of Marketing Research*, *20*, 368-379.

Neustadt, R. E., & May, E. R.. (1986). *Thinking in time: the uses of history for decision makers*. New York: Free Press.

Newman, B. (1982). Artists in Holland survive by selling to the government. *The Wall Street Journal*, 7 January, 1.

Organski, A. F. K. (2000). The outcome of the negotiations over the status of Jerusalem: a forecast. In Miroslav, N., & Lepgold, J. (Eds.), *Being useful: policy relevance and international relations theory*. Ann Arbor, MI: University of Michigan Press: 343-359.

Pfister, H. R., & Konerding, U. (1996). Explaining and predicting behavior with uncertain consequences: inferences from behavioral decision research for attitude research. *Zeitschrift für Sozialpsychologie*, *27*(1), 90-99.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104,* 1-15.

Radio New Zealand Limited (2000a, September 20). Brenda Wilson (Chief Executive, New Zealand Nurses Organisation) interviewed by Geoff Robinson. *Morning Report*, Transcript: Newztel News Agency Ltd.

Radio New Zealand Limited (2000b, September 20). Rae Lamb (Health Correspondent, Radio New Zealand) interviewed by Mary Wilson with exerted material from interviews with Annette King (Minister of Health), Susan Rolls (Emergency Nurse at Wellington Hospital), and Russell Taylor (Wellington Nurses Union Organiser). *Checkpoint*, Transcript: Newztel News Agency Ltd.

Radio New Zealand Limited (2000c, September 22). Margot Mains (Chief Executive Officer, Capital Coast Health) interviewed by Geoff Robinson. *Morning Report*, Transcript: Newztel News Agency Ltd.

Raiffa, H. (1982). *The art and science of negotiation*. Cambridge, MA: The Belknap Press of Harvard University Press.

Rapoport, A., & Orwant, C. (1962). Experimental games: a review. *Behavioral Science*, *7*(1), 1-37.

Reisman, A., Kumar, A., & Motwani J. G. (2001). A meta review of game theory publications in the flagship US-based OR/MS journals. *Management Decision*, *39*(2), 147-155.

Rey, S. J. (2000). Integrated regional econometric and input-output modelling: issues and opportunities. *Papers in Regional Science, 79*, 271-292.

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: the role of the Delphi technique. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 125-144.

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 405-416.

Sandholm, W. H. (1998). History-independent prediction in evolutionary game theory. *Rational Society*, *10*(3), 303-326.

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: game theory with a human face. *Journal of Economic Psychology, 22*(5), 617-640.

Schelling, T. C. (1961). Experimental games and bargaining theory. *World Politics*, *XIV*(1), 47-68.

Schrodt, P. A. (2002). Forecasts and contingencies: from methodology to policy. Paper presented at the American Political Science Association meetings, Boston, 29 August – 1 September. Retrieved December 10, 2002, from www.ku.edu/~keds/pdf.dir/ Schrodt.APSA02.pdf.

Schwenk, C. R. (1995). Strategic decision-making. *Journal of Management*, *21*(3), 471-493.

Shakespeare, W. (1606). Macbeth, act 1, sc. 3. *Complete Moby™ Shakespeare*. Retrieved January 1, 2003, from http://the-tech.mit.edu/Shakespeare/macbeth/ macbeth.1.3.html.

Shefrin, H. (2002). Behavioral decision making, forecasting, game theory, and role-play, *International Journal of Forecasting*, *18*, 375-382.

Shubik, M. (1975). *Games for society, business and war*. Amsterdam: Elsevier.

Siegel, S., & Castellan, N. J. Jr. (1988). *Nonparametric statistics for the behavioral sciences*, 2nd ed.. Singapore: McGraw-Hill.

Sigall, H., Aronson, E., & van Hoose, T. (1970). The cooperative subject: myth or reality. *Journal of Experimental Social Psychology, 6*, 1-10.

Singer, A. E., & Brodie, R. J. (1990). Forecasting competitors' actions: an evaluation of alternative ways of analyzing business competition. *International Journal of Forecasting, 6*, 75-88.

Smith, V. L. (1994). Economics in the Laboratory. *Journal of Economic Perspectives, 8*(1), 113-131. Retrieved December 10, 2002, from www.ices-gmu.org/Pdfs/ JEP1994.pdf.

Sonnegard, J. (1996). Determination of first movers in sequential bargaining games: an experimental study. *Journal of Economic Psychology*, *17*(3), 359-386.

Souder, W. E., & Thomas, R. J. (1994). Significant issues for the future of product innovation. *Journal of Product Innovation Management*, *11*(4), 344-353.

Statman, M., & Tyebjee, T. T. (1985). Optimistic capital budgeting forecasts: an experiment. *Financial Management*, *Autumn*, 27-33.

Stewart, T. R. (2001). Improving reliability of judgmental forecasts. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers, 81-106.

Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002) Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11537-11542.

Suleiman, R. (1996). Expectations and fairness in a modified Ultimatum game. *Journal of Economic Psychology*, *17*(5), 531-554.

Tesfatsion, L. (2002). Agent-based computational economics: growing economies from the bottom up. *Artificial Life, 8*(1), 55-82.

Tesfatsion, L. (2003 – forthcoming). Agent-based computational economics. To be published in Luna, F., Perrone, A., & Terna, P. (Eds.), *Agent-based theories, languages, and practices*. Routledge. Retrieved December 10, 2002, from www.econ.iastate.edu/tesfatsi/acewp1.pdf.

Tetlock, P. E. (1992). Good judgment in international politics: three psychological perspectives. *Political Psychology*, *13*(3), 517-539.

Tetlock, P. E. (1999). Plausible pasts and probable futures in world politics: are we prisoners of our preconceptions. *American Journal of Political Science, 43*(2), 335-366.

Walker, M. & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review, 91*, 1521-1538.

Winkler, R. L. (1983). The effect of combining forecasts and the improvement of the overall forecasting process. *Journal of Forecasting, 2*, 293-294.

Wong, F. & Tan, P. Y. (c1992). Neural networks and genetic algorithm for economic forecasting. Institute of Systems Science, National University of Singapore. Retrieved 11 December, 2002, from http://sunsite.bilknet.edu.tr/pub/security/cerias/doc/genetic_ algorithms/apps/GA-Financial-Forecasting.ps.gz.

Wright, G. (2002). Game theory, game theorists, university students, role-playing and forecasting ability. *International Journal of Forecasting*, *18*, 383-387.

Wrolstad, J. (2002, March 19). Alltel pays $1.65b for CenturyTel Wireless. *Wireless NewsFactor*. Retrieved May 3, 2002, from http://www.wirelessnewsfactor.com/perl/ story/16833.html.

Yokum, T., & Armstrong, J. S. (1995). Beyond accuracy: comparison of criteria used to select forecasting methods. *International Journal of Forecasting*, *11*, 591-597.

Zajac, E. J., & Bazerman, M. H. (1991). Blind spots in industry and competitor analysis: implications of interfirm (mis)conceptions for strategic decisions. *Academy of Management Review*, *16*(1), 37-56.