ELSEVIER

International Journal of Forecasting 18 (2002) 321–344

international journal of forecasting

www.elsevier.com/locate/ijforecast

Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement

Kesten C. Green*

School of Business and Public Management, Victoria University of Wellington, and Decision Research Ltd., P.O. Box 5530, Wellington, New Zealand

Abstract

Can game theory aid in forecasting the decision making of parties in a conflict? A review of the literature revealed diverse opinions but no empirical evidence on this question. When put to the test, game theorists' predictions were more accurate than those from unaided judgement but not as accurate as role-play forecasts. Twenty-one game theorists made 99 forecasts of decisions for six conflict situations. The same situations were described to 290 research participants, who made 207 forecasts using unaided judgement, and to 933 participants, who made 158 forecasts in active role-playing. Averaged across the six situations, 37 percent of the game theorists' forecasts, 28 percent of the unaided-judgement forecasts, and 64 percent of the role-play forecasts were correct.

Keywords: Conflict; Expert opinion; Forecasting; Game theory; Judgement; Role-playing; Simulation

1. Introduction

In 1996 the New Zealand government transferred some of the assets of its monopoly electricity generator to a new private sector electricity-generating company, Contact Energy Ltd. It further split the residual into three entities in 1999. Wishing to know how participants in the new competitive market for wholesale electricity would behave following the second split, Contact organised its executives to role-play the generator company managers in a series of electricity trading simulations. The role-play behaviour was so at odds with the executives' own beliefs about how the market participants should and would behave, that they ignored the forecast. Turning to game theory for help, Contact management found it to be ''no help at all . . . the role-playing exercise had already foretold the future, as we were to find out to our cost.''[1]

This anecdote suggests that role-playing may

*Tel.: + 64-4-499-2040; fax: + 64-4-499-2080.
E-mail address: kesten.green@vuw.ac.nz (K.C. Green).

[1] Interview with Toby Stevenson, General Manager Electricity Trading, Contact Energy Limited, 7 December 2000.

0169-2070/02/$ – see front matter
PII: S0169-2070(02)00025-0

be an effective approach to predicting decisions made in conflicts among small numbers of decision makers with much at stake. The primary purpose of the research described in this paper was to investigate the relative accuracy of methods used to forecast decisions made in real conflicts. For this purpose, I defined accuracy as the proportion of forecasts that match the actual decision. Accuracy is commonly regarded as the most important criterion for judging the worth of a forecast (Armstrong, 2001b). The methods I examined were unaided judgement, game theory, and role-playing. I defined game theory as what game theorists do when faced with practical forecasting problems. It was *not* the purpose of the study to investigate other aspects of the methods, such as their value for generating strategic ideas.

While unaided judgement is commonly used to forecast decisions in conflicts, game theory and role-playing are not. Armstrong, Brodie, and McIntyre (1987) surveyed 59 practitioner members of the International Institute of Forecasters. The practitioners were asked about the use, by their respective organisations, of methods for forecasting competitive actions. The authors found that the organisations of 85 percent of practitioners used the opinions of experts with domain knowledge, the organisations of 8 percent of practitioners used formal game theory, and the organisations of 7 percent of practitioners used role-playing. The same study found expert opinion on the relative value of the methods to be at odds with the reported frequency of use by practitioners' organisations. Both marketing and forecasting experts ranked game theory and role-playing more highly than practitioners, although they disagreed about the relative value of the two methods—forecasting experts preferred role-playing over game theory.

Game theory may help practitioners provide more accurate forecasts than unaided judgement because, for example, the discipline of the approach should tend to counter judgemental biases. Indeed, 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''. The entry on game theory in Bullock and Trombley's (1999) dictionary states that game theorists ''hope to produce a complete theory and explanation of the social world''. Given these claims and the fact that game theory is used by forecasting practitioners and is recommended by experts, it is legitimate to ask whether the method can help forecasters make useful predictions for real conflicts.

Opinions on the value of game theory for forecasting real conflicts are diverse. In contrast to the optimistic claims made by Nalebuff and Brandenburger (1996) and in Bullock and Trombley (1999), Shubik (1975, p. xi) described the assumptions behind formal game theory as ''peculiarly rationalistic''. He continued: ''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''. He concluded that while game theory may be applicable to actual games (such as backgammon or chess), and 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'' (p. 14).

The usefulness and realism of role-playing are often contrasted with the limitations of game theory in the game-theory literature. For example, Nalebuff and Brandenburger (1996, p. 62) emphasised the importance and difficulty of

appreciating the perceptions of other parties. In a brief note, they suggested that one way for managers to do this is to ''ask a colleague to role-play by stepping into [another] player's shoes'' (p. 63).

Shubik (1975) dealt with role-playing more comprehensively. He stated 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). Shubik further suggested that game theory is less realistic than role-playing (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).

Schelling (1961) stated that ''part of the rationale of game organization [role-play experimentation] 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'' (p. 47). In contrast, role-plays ''do generate these complexities and, by most reports, do it in a fruitful and stimulating way''.

2. Prior evidence

Although judgement may be adequate for predicting decisions made in routine conflicts, decision makers can be subject to serious judgemental biases or blind spots. Decision makers who are involved in a conflict tend to give ''insufficient consideration of the contingent decisions of others'', as evidenced by such phenomena as winner's curse and non-rational escalation of commitment (Zajac & Bazerman, 1991, p. 43).

Babcock, Loewenstein, Issacharoff, and Camerer (1995) demonstrated that judgement tends to be biased by the role of the judge. They asked participants to provide an estimate of a fair judgement in a dispute between two parties. Before being presented with identical briefing material, half the participants were told they were to take the role of a lawyer for the complainant and the other half were told they were to represent the defendant. Estimates of a fair judgement diverged between the two groups of 'lawyers' and the authors show that the two groups interpreted the briefing material in different (self-serving) ways. Similarly, Cyert, March, and Starbuck (1961) found that role-players produced divergent forecasts from identical sets of numbers depending on whether they were told they were 'cost analysts' or 'sales analysts'. Statman and Tyebjee (1985) replicated the study, and obtained results consistent with the earlier research.

Given the foregoing evidence, a manager wanting to forecast a decision in a conflict may consider asking someone *not* involved in the conflict for his or her considered opinion on what decision is most likely. Yet independent judges are also subject to influences that lead to inaccurate forecasts. For example, they may be subject to biases arising from the use of common judgemental heuristics, or to overconfidence. Bazerman (1998, p. 6) identified three broad classes of heuristic (availability, representativeness, anchoring) that can engender judgemental biases. Arkes (2001) examined the evidence and concluded that experience often leads forecasters using unaided judgement to ignore base rate data and shun decision aids to the detriment of forecast accuracy. Evidence on the forecasting accuracy of independent judges

is provided by Armstrong (2001a) who found that research participants exercising their unaided judgement performed no better than chance when predicting decisions made in conflicts.

There is little evidence on the predictive validity of game theory for real conflicts. Reisman, Kumar, and Motwani (2001) reviewed all game-theory articles published in the leading US OR/MS journals and found that an average of less than one article per year addressed a real-world application. Armstrong (1997) stated ''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).

Such evidence as there is tends to be indirect and incomplete, typically comparing game-theory predictions with decisions made in role-play experiments rather than with those from actual conflicts. For example, Armstrong and Hutcherson (1989) found two studies (Neslin & Greenhalgh, 1983; Eliashberg, LaTour, Sangaswamy, & Stern, 1986) that involved the use of game theory to predict decisions made in negotiations. In both studies the negotiations were role-plays rather than actual negotiations, with the implication that the role-play decisions (which game theory predicted imperfectly) were equivalent to actual negotiation agreements. Similar approaches have been widely used by game theorists. Shubik (1975) notes that ''experimental gaming'' is employed to examine the ''validity of various solution concepts'' developed by game theorists (p. 20). Rapoport and Orwent (1962) conducted a comprehensive review of the use of experimental games to test game-theory hypotheses. They concluded this review with a suggestion that ''game theory is not descriptive and will not predict human behavior, especially in games with imperfect information about the payoff matrices'' (p. 34).

I sent an e-mail letter to 474 e-mail addresses asking for empirical evidence on the predictive validity of game theory for real conflicts. I had previously used the addresses to recruit game-theorist research participants. Included in the letter was a URL for a draft of this paper. I received 18 responses, none of which provided substantive evidence on predictive validity.

One response was an automatic message from an ISP informing of an invalid address and five were automatic messages generated for addressees who were on leave or otherwise unavailable. A single respondent claimed to have no relevant expertise, and a further four either asked for the draft paper in another format or stated that they would look at the material later. Six respondents provided comments on the paper or information on game theory. One of these 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''. The second of these respondents pointed to his own work on the civil conflict in Northern Ireland as offering evidence of predictive validity for a variety of game theory (Brams & Togman, 2000). He stated ''I think our predictions . . . have for the most part been borne out by events''.[2] Brams and Togman suggested that game theory and the theory of moves (an extension of game theory) give insights that ''can help political leaders predict both the dynamics of conflict and the conditions that can provide an escape from it'' (p. 337). The book in which their paper appears also contains claims of accuracy for a forecasting model using game-theory and decision-theory analysis of the civil conflict over Jerusalem (Organski, 2000). A third respondent pointed to his own work on analysing play in an international sporting competition. His paper was not relevant to this research. Finally, four of the

[2]Personal communication from Steven Brams, 13 June 2001.

respondents who had offered comment or information did not address the question of predictive validity.

A search of the Social Science Citation Index (SSCI) for the period 1978 to 7 July 2001 using the phrases 'game theory' and 'forecasting' yielded two articles. Substituting 'prediction' for 'forecasting' yielded a further 14 articles. Jehiel (1998) turned up in both searches, but the article was purely speculative and theoretical. Of the remainder, one concerned a behavioural pharmacology study using mice (Parmigiani, Ferrari, & Palanza, 1998), a second was a review of research on strategic decision making (Schwenk, 1995), and a third concerned the predictive validity of agency theory (Ghosh & John, 2000).

Aside from Henderson (1998), which offered only opinions on the usefulness of game theory, the remaining 11 articles all provided evidence for the predictive accuracy of game theory relative to another method, relative to a decision made in a real conflict, or both. Predictions were compared to the outcomes of experiments in six of the articles, and this was the only comparison that was made in two of these articles (Diekmann, 1993; Sonnegard, 1996). Batson and Ahmad (2001) compared the predictions of classic game theory with predictions from the theory of rational choice and the empathy–altruism hypothesis. Gibbons and Van Boven (2001) compared game-theory predictions with the outcomes of experiments and the stated preferences of role-players. McCabe and Smith (2000) compared game-theory predictions with the outcomes of experiments and with predictions based on theories of 'social exchange'. Suleiman (1996) conducted experiments to test hypotheses about why classic game theory fails to predict the behaviour of participants in 'ultimatum' game experiments. The experiments involved comparing the game outcomes with the predictions of the participants.

Three of the articles mentioned the predictive accuracy of game theory for real conflicts. Ghemawat and McGahan (1998) used historical market data to examine price competition between electricity generating companies. The authors found that game-theory models had greater explanatory power than 'nonstrategic analysis'. They suggested that game theory is most likely to provide accurate forecasts of conflicts when there is concentrated competition (few players), mutual familiarity, and repeated interaction. Gruca, Kumar, and Sudharshan (1992) proposed a game-theory model of existing players' responses to a new entrant to a market. They compared the predictions from their game-theory model with empirical evidence on actual behaviour in conflicts of this type and with the predictions of an alternative model. They found that their model provided forecasts that were more consistent with the empirical evidence for more combinations of competitive situation than did the alternative model, but did not provide enumeration. Keser and Gardner (1999) found that the Nash equilibrium failed to predict the behaviour of participants in a common-pool resource game experiment. They suggested that ''policies based on that equilibrium's predictions are suspect''. The participants were students experienced in game theory, and the researchers expected the design of the experiment to favour Nash equilibrium decisions.

Finally, Sandholm (1998) compared the predictions of three classes of 'evolutionary game theory' models of social convention on the basis of reasonableness.

Although the 11 articles did not provide formal evidence on the accuracy of forecasts by game theorists relative to reasonable non-game-theoretic alternatives, they did support the contention that game theory is considered to be a forecasting method by some authors.

A search of the Internet using the Google search engine and the same terms I used for the

SSCI search on July 17, 2001 ('game theory' and 'forecasting') produced too many hits to investigate. Using the three phrases 'comparative', 'forecasting accuracy', and 'game theory' yielded 17 hits, including one duplicate. Eleven of the hits were lists of courses or papers—the three search terms were dispersed among unrelated course summaries and abstracts of papers. Four of the hits were of single documents, rather than lists. None of the four offered any evidence on the forecasting accuracy of game theory for real conflicts. One URL returned an 'access denied' message.

Bennett (1995, p. 27) listed four missing dimensions of classic game-theory models: differing perceptions, dynamics, combinatorial complexity, and linked issues. Two extensions of game theory that set out to address its shortcomings have been 'hypergame analysis' (Bennett & Huxham, 1982), and 'drama theory' (Howard, 1994a,b). While hypergames are intended to take players' divergent perceptions into account by describing and analysing a set of subjective but linked games, drama theory moves further away from the sparseness of game theory by attempting to incorporate emotion into the analysis of conflict situations. On the face of it, these developments should deliver greater realism—and perhaps also greater predictive accuracy—than classic game theory. Nevertheless, to test hypotheses about behaviour generated by drama theory, Bennett and McQuade (1996), for example, resorted to role-playing to provide the behavioural benchmark.

Role-playing is the third and last of the three forecasting methods considered. It is ''a technique whereby people play roles and enact a situation in a realistic manner. Role-playing can be used to predict what will happen if various strategies are employed. It is especially relevant when trying to forecast decisions made by two parties who are in conflict'' (Armstrong, 2001d, p. 807). Role-playing is used to simulate rather than analyse conflict situations. Armstrong (2001a) found that role-playing provided more accurate forecasts than chance and unaided judgement.

Role-playing and game theory depend on contrasting assumptions about modelling conflict situations. Those who adopt a game-theory approach must assume that they can radically reduce the complexity of a conflict without predictive validity being lost. The role-play approach, on the other hand, incorporates complexity and emotion into a simulation.

Armstrong (2001c) has suggested that the accuracy of different approaches to predicting decisions made in conflicts is related to the realism with which they allow forecasters to model the situation: the more realistic the representation, the more accurate the predictions are likely to be (Principle 7.2). This hypothesis implies that the use of game-theory knowledge should result in more accurate forecasts than the use of unaided judgement and that the use of role-playing should produce more accurate forecasts than the use of game-theory knowledge.

3. Methods

3.1. Selection of conflicts

I chose six real conflicts to assess the relative accuracy of forecasts from unaided judgement, game theory, and role-playing. The conflicts each occurred between small numbers of decision-makers with much at stake. They were diverse in the parties that were involved, the level of conflict, and the type of decision that was to be made. Furthermore, the conflicts were all unlikely to be recognised by participants. I neither sought nor deliberately avoided conflict situations that resembled familiar game-theory models. Neutral but informed observers wrote the descriptions of the situations.

Three of the situations ('Artists' Protest', 'Distribution Channel', and '55% Pay Plan')

were first used in research by Armstrong (1987). Artists' Protest was a conflict between artists and government over financial support—Armstrong based his description of it on a report in *The Wall Street Journal* (Newman, 1982). Distribution Channel was a proposal by Philco Corporation for commercial cooperation requiring decision makers to trade off conflicting interests—Armstrong took a description of it from a book of marketing case histories (Berg, 1970). The 55% Pay Plan was a conflict between National Football League owners and players over revenue shares—Armstrong based his description of it on two reports in *Sports Illustrated* published prior to the start of negotiations (Boyle, 1982; Kirshenbaum, 1982). A fourth situation, 'Panalba Drug Policy', was first used in experiments conducted by Armstrong (1977) on the influence of role on managers' behaviour. Panalba was a conflict between the board members of a pharmaceutical company and the consumers of one of the company's drugs. Unlike the other conflicts in this research, one party to the conflict (the consumers) was absent except in the abstract. Armstrong wrote: ''The description was based upon the true case of Panalba as reported by Mintz (1969). Information was also taken from Upjohn's Annual Reports. I made up details for this case . . . to make the extreme nature of this case obvious. Attempts were made to obtain further information from the Upjohn Co. to ensure the facts were accurately presented, but they refused to answer'' (p. 196).

Prior research using these four situations provided information on the accuracy of role-playing and judgemental forecasts. These data were collected by Armstrong and colleagues and were summarised in Armstrong (2001a). In all, 147 judgemental forecasts made by 230 participants, and 119 role-play forecasts from 653 participants were obtained from this source.

I wrote the description of the fifth situation, 'Zenith Investment' for this research project. I based my description on a Grenada Television documentary about a British Steel investment decision that was complicated by conflicting interests among the decision makers (Graef, 1976). I disguised the situation, but did not alter it in any substantive way. I sought advice from a professor who had used the documentary in his teaching for many years. He had also worked at British Steel for one of the managers who played a major part in the investment decision portrayed in the documentary. The professor considered the description to be an accurate representation of the situation.

I also wrote the description of the sixth situation, 'Nurses Dispute'. The situation was a dispute over pay between nursing staff and management. The nurses went on strike angry that they were being offered a much lower pay increase than management had already given to intensive care nurses and junior doctors. A mediator was appointed by a government agency. The principal negotiators for the two parties co-operated with the research and, after they had reached an agreement, some members of the nurses' negotiation team role-played the dispute in a debriefing session using the material I had prepared for this research. Asked how the description could be improved, the principal negotiator did not have any suggestions and considered the material a fair and accurate representation. I used two versions of the situation description that were the same in all respects other than the names of the people and organisations involved. In the first version I used actual names. In the second, I changed the names of the people and organisations after one potential participant said he was unhappy playing the role of a real person and withdrew from his session.

The descriptions of the six situations are available on the Internet at www.kestencgreen.com. I used the same situation descriptions, role descriptions, and exhaustive list of decisions from which the participants were asked to

choose for all three forecasting methods: unaided judgement, role-playing, and game theory. I discuss an exception to the uniform treatment below.

For any forecasting method to be useful, it must be more accurate than chance. I use the term 'chance' to refer to the probability of selecting, at random, the decision that was actually made, from the exhaustive list of potential decisions provided to participants.

3.2. Unaided judgement

The participants using unaided judgement read the description of a conflict and then selected the decision they thought most likely from a list. I recruited these judges on the basis of convenience. They had no special knowledge of the situations nor of the class of problem I asked them to consider.

Except for predictions for the Zenith and Nurses situations, I took most of the predictions using unaided judgement that are used in this research from studies reported by Armstrong (2001a). Role descriptions were not provided to all the judges who took part in the research reported by Armstrong, but Armstrong and Hutcherson (1989) have shown that providing role descriptions to judges has no effect on their forecasting accuracy. In some cases, the judges were paired off to discuss the problem before making their predictions. In other cases, they acted alone.

I provided all participants in the new research with a full set of information for a single situation (including roles for both parties) and a questionnaire. I told them that the material included a description of a conflict that had occurred in the past. I asked them to read the material and, without conferring with others or referring to other sources of information, use their judgement to predict the decision that had actually been made, and then complete and return the questionnaire.

Volunteers from one third-year university class provided 10 Zenith predictions. I asked participants to leave the lecture theatre with the material I had given them and to return the completed questionnaires no more than one hour after having received their instructions.

Volunteers from two third-year classes provided 20 predictions. I used 10 minutes of class time to distribute material on five situations (Nurses was not included) and to brief the students. I instructed them to return their completed questionnaires to their lecturer, or to fold them (so that 'Freepost' information and address were displayed) and post them.

I had offered five marketing students, one information-technology student, one environmental-planning consultant, and one medical doctor $NZ25 (about $US10) to participate in a role-playing session, but I did not need them for that purpose. Instead, they provided eight predictions for the Zenith situation. I asked them to adjourn to a room away from the role-players with the material I had given them and to return completed questionnaires to me no more than one hour after getting their instructions. In a situation similar to the one just described, students who were not needed for role-playing provided predictions for the Nurses situation instead.

3.3. Role-playing

I gave each role-play participant a single role description to read and told participants to adopt the role they had been given for the duration of their simulation. When they had adopted their roles, I asked the role-players to read a description of the conflict. I then told them to form groups, with each group comprising one role-player for each role. For example, in the Nurses situation there were five roles: two management roles, two union roles, and one mediator role. Once the role-players were in their groups, I told them to role-play the conflict from the time

specified in the description until they arrived at a decision or ran out of time. When they finished, each role-player recorded the decision the group had made or, if they had run out of time, the decision he or she thought would have been made had the group been free to continue.

With the exception of six groups of Nurses Dispute role-players, the role-players were, like the judges, recruited on the basis of convenience rather than because they resembled the characters they were to play or were familiar with the roles they were to play or were knowledgeable about the situation. Most of the role-players were university students. I conducted role-play simulations for the Zenith and Nurses situations as part of this research, and took role-play results for the other four situations from Armstrong (2001a).

I brought the participants for the Zenith and Nurses simulations together in lecture theatres or similar settings for their briefings. I gave each participant an information sheet, which provided basic information on the research project and on role-playing, and a consent form. I told them it was very important that they take the role-playing seriously. Additional space, such as a second lecture theatre, meeting rooms, or lobby areas, was available for role-playing. I encouraged the role-players to make good use of the available space for holding both formal meetings and private discussions. At the end of the briefings, I told them 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 the session. This material included the questionnaire that the role-players were to fill in at the end of their role-plays. I drew their attention 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 role-play decision. I also gave role-players printed self-adhesive badges showing their characters' names and positions. I told them to meet in their groups, wearing their name badges, and to introduce themselves to each other while in character.

I told the Zenith Investment role-players that the chairman of Zenith would set a time and place for a meeting and that they should prepare for the meeting in a manner consistent with their printed briefing material. In 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 told Nurses Dispute role-players that, after introductions, the two parties to the dispute should agree to a time and place for their first meeting and that subsequent meetings could be held at the discretion of the parties. I told participants who had been given the mediator roles, one from each group at the session, to meet at a designated place. I told them to discuss mediation and its application to the dispute among themselves until the parties in their own groups called upon them, or 30 minutes had elapsed. If the contesting parties failed to agree within 30 minutes, they 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.

No efforts were made to increase the realism of the simulations beyond what has been described here. There were no theatrical or technological devices, nor were any confederate role-players used.

The procedures I adopted in this research were similar to those Armstrong and Hutcherson adopted for Artists' Protest, Distribution Channel, and 55% Pay Plan (Armstrong, 1987; Armstrong & Hutcherson, 1989). Although similar in other respects, the role-plays of Panalba Drug Policy (Armstrong, 1977) involved people belonging to a single party (the members of the Upjohn Board) who had a

common purpose—neither representatives of consumers nor of the Food and Drug Administration were present.

A total of 170 university students role-played the Zenith situation in 17 groups. I conducted these role-plays in three sessions with different participants. Five groups role-played the situation in an organisational behaviour class and four in a conflict-of-laws class. I recruited the eight groups of role-players in the third session from among students attending mathematics and computer science lectures. I offered the students who attended the third session $NZ25 to take part. Whereas participants in earlier sessions had been assigned randomly to roles, in the third session I asked 'natural leaders' and 'quantitative analysis experts' to come to the front of the theatre. I allocated 'natural leaders' to the Chairman roles and 'quantitative experts' to the roles of either Finance Director or Chief Planner. I assigned the remaining participants randomly to the seven other roles.

I told all participants that their taking part would help with research on decision making in conflict situations and was likely to be both enjoyable and (in the case of the first two sessions) relevant to their studies. I told participants that the situation they were to role-play had occurred in the past and that it involved a group of senior managers making an important investment decision.

A total of 110 participants role-played the Nurses situation in 22 groups. I conducted three sessions. In the first session, 10 students with work experience who were enrolled in courses on dispute resolution role-played in two groups. In the second session, 90 students recruited with an offer of $NZ25 cash role-played in 18 groups. In the third session, 10 participants selected for experience relevant to the situation (union negotiators, managers, management negotiators, and professional mediators) role-played in two groups.

I asked participants in the first and second Nurses sessions whether they would tend to identify more with union or management in a pay dispute. I then allocated them according to their preferences, with the more equivocal participants given the mediator role. I allocated participants in the third session to roles on the basis of my assessment of the compatibility of their experience with the demands of the roles.

3.4. Prediction by game theorists

The initial sample of 558 game theorists was composed of the members of the Game Theory Society, recipients of the International Society of Dynamic Games E-Letter,[3] and a small number of prominent game-theory experts who were not members of the other two groups. I expected that a large initial sample would be necessary because the task was demanding and I was offering no extrinsic rewards.

I sent everyone in the sample of experts an e-mail message and material on five situations, excluding the Nurses Dispute (Appendix A). I personalised the messages when this was possible. Although the Game Theory Society membership list contained only e-mail addresses, names were included in most of the e-mail addresses. I conducted Internet searches when they were not and ended up with names for all but 57 e-mail addresses in the sample.

I made recipients aware of the purpose of the research with the e-mail message's subject line: 'Using Game Theory to predict the outcomes of conflicts'. Moreover, in the first paragraph I wrote ''I am engaged on a research project which investigates the accuracy of different methods for predicting the outcomes of conflicts''.

I attached MS-Word files containing the material on the situations to the messages. I varied the order of the five situation files across

[3] http://www.hut.fi/Units/SAL/isdg/issue23.txt.

the sample to reduce the risk that respondents who predicted decisions for only some of the situations would do so for similar subsets. In addition, I hoped that varying the order of the situations in the e-mail messages would counter any effect that might have arisen had the experts changed their processes of deriving predictions in some systematic way as they worked through the situations.

The material on each of the situations included descriptions of all of the roles, a description of the situation, and a questionnaire. In the questionnaires, I asked respondents to select the most likely from a list of possible decisions. I also asked them how long they took to derive the prediction and how many years' experience they had working with game theory. I repeated the question on experience for each situation in case the recipient of the e-mail message asked other people to respond to one or more of the problems. In anticipation of some respondents being willing to participate in the research but unwilling to make a prediction for particular situations, I included a question with each situation asking those who did not make predictions to give a reason for not doing so. I also asked them to identify any situations they thought they recognised. I asked respondents to complete and return their questionnaires electronically or to print them out and return the completed forms by fax or post.

After 13 and 14 days, I sent individualised e-mail reminders to the 413 addressees in the original sample who had not responded. The reminder included a copy of the original letter and the attached files in the same order as in the original e-mail message.

I sent individual replies to all those who responded, irrespective of the nature of their responses.

A year after the first appeal to game theorists, I sent material on the Nurses situation to those who had provided predictions in response to the first appeal. I followed this with up to two reminders one week apart to those who had not already responded.

4. Findings

4.1. Unaided-judgement replication and extension

4.1.1. Participation

Participants provided one forecast each for Artists' Protest (eight forecasts), Distribution Channel (five), Panalba Drug Policy (four), Zenith Investment (21), and Nurses Dispute (22).

Students from the two classes who undertook the task in their own time reported taking between 2 and 120 minutes to derive their individual predictions. The mean time was 30 minutes and the median was 25 minutes. I gave the participants making predictions for Zenith and Nurses an hour to do so. Zenith judges took an average of 15 minutes to derive a prediction. I did not ask the Nurses judges to record the time they actually spent on the task.

One Nurses Dispute respondent claimed to know more about the situation than was contained in the material provided but gave no indication of knowing anything about the actual agreement, and so I included the response in the analysis. None of the other respondents admitted to recognising any of the situations they had been given.

Fewer than 10 percent of the participants I asked to undertake the task in their own time returned completed questionnaires. I received responses for four of the five situations provided—none was received for the 55% Pay Plan.

4.1.2. Predictions

The unaided-judgement predictions new to this research were (in an unweighted average across the situations) correct for 31 percent of

predictions. This rate is little different from chance (27 percent) for the five situations for which I received responses. New predictions for situations that were also reported by Armstrong (2001a) (Artists, Distribution, and Panalba) were correct for 19 percent of predictions. This finding is consistent with Armstrong (2001a), who reported 13 percent correct for the same situations.

4.2. Role-playing extension

4.2.1. Participation

For the Zenith Investment situation, roughly half the students in the two classes that were not offered payment stayed to take part in role-plays. After their briefing, the organisational behaviour class had little more than 30 minutes remaining for role-playing and completing questionnaires. The conflict-of-laws class had a full hour available, and all took this long, while the groups of paid recruits took as long as an hour for their role-playing alone. Some of the role-players given only half an hour complained of not having enough time, as did some who were given an hour. Nevertheless, all 17 groups made decisions. One of the paid recruits recognised the situation but did not recall the decision that was made or discuss her knowledge with her group.

For the Nurses Dispute situation, participants in the first and second sessions typically took between 45 minutes and one hour for their role-playing. The two groups of role-players in the third session took one-and-a-half hours for their role-playing and neither group had come to an agreement when I asked them to stop.

4.2.2. Predictions

For the Zenith Investment situation, the role-play decisions matched the actual decision for 10 of the 17 groups (59 percent). This was better than chance, which is 33 percent for Zenith. The finding was similar to the unweighted average for the role-plays of the four situations reported by Armstrong (2001a), which was 60 percent. Chance for these four situations is, on average, somewhat lower at 26 percent, however. The forecasting accuracy of the eight groups with 'natural leaders' and 'quantitative analysis experts' allocated to roles was similar to that of the other nine groups—there were five correct forecasts from each treatment.

For the Nurses Dispute situation, role-play agreements matched the actual agreement between management and union in 18 of the 22 groups (82 percent). This is better than chance. Although neither of the groups in the two-group (third) session reached an agreement in the course of their role-plays, both groups chose the actual decision when asked what decision they thought would have occurred had their role-playing continued.

4.3. Game theorists

4.3.1. Participation

I received responses of various types from 269 addresses (48 percent of the initial sample). Of these, 78 were invalid-address messages, 18 were automatic rejection messages (typically on leave messages), and six were statements that the addressee was not a game-theory expert. The balance of 167 responses consisted of messages from 95 game theorists who stated that they did not wish to participate, 51 game theorists who promised to respond or who wanted more information but did not respond after receiving it, and 21 game theorists who each responded with a completed questionnaire for at least one situation.

4.3.1.1. Reasons for refusing to participate

Of the 95 people who responded but refused to participate, 90 provided reasons (Appendix B). Most (72) stated they were too busy with other commitments to spend time helping with

this research. Eight maintained that their game-theory speciality was not applicable to the problems provided. Ten wrote either that it was not appropriate to apply game theory to the problems given (six) or that there was insufficient information for them to derive predictions (four).

4.3.1.2. Selective prediction

Nine of the 21 experts who returned completed questionnaires did not provide predictions for all of the original five situations. I obtained 85 usable predictions in all from the 21 experts out of a potential total of 105 predictions. It seems reasonable to assume that they made predictions for the situations they considered themselves most capable of predicting accurately. One respondent recognised the Artists' Protest situation and I excluded his response from the analysis.

4.3.1.3. Further participation

When I appealed again to the 21 game theorists who had responded to the original appeal, 14 provided usable predictions for the Nurses Dispute situation and one sent an unusable response.

4.3.1.4. Probabilistic forecasts

While the questionnaires instructed respondents to 'check one—✓' or something similar, the e-mail message stated ''you may assign probabilities to possible outcomes, rather than picking a single outcome, if you consider this to be appropriate''. One game theorist who responded for a single situation (Artists) provided probabilities rather than choosing a single decision. He assigned a zero probability to the actual decision.

4.3.2. Predictions

Overall, the game-theory experts' predictions matched the actual decision for 37 percent of their predictions. This rate is better than chance (an unweighted average of 27 percent over the six situations) and better than unaided judgement (28 percent).

5. Discussion

5.1. Relative predictive accuracy

In Table 1, I combined the results from this research (chiefly game-theory experts' predictions for six situations, and unaided judgement

Table 1
Accuracy of unaided judgement, game theorist, and role-play predictions. Percent correct predictions (number of predictions)

	Chance	Unaided judgement[a]	Game theorist	Role-play[b]
Artists' Protest	17	5 (39)	6 (18)	**29** (14)
Distribution Channel	33	5 (42)	31 (13)	**75** (12)
55% Pay Plan	25	27 (15)	29 (17)	**60** (10)
Zenith Investment	33	29 (21)	22 (18)	**59** (17)
Panalba Drug Policy	20	34 (68)	**84** (19)	76 (83)
Nurses Dispute	33	68 (22)	50 (14)	**82** (22)
Totals (unweighted[c])	27	28 (207)	37 (99)	**64** (158)

[a] Results reported in Armstrong (2001a), except Zenith Investment, Nurses Dispute, and 17 predictions for other situations from this research: Artists (one correct/$n = 8$); Distribution (1/5); Panalba (1/4).

[b] Results reported in Armstrong (2001a) except Zenith Investment and Nurses Dispute.

[c] Percentage figures in this row are unweighted averages of the percentage of correct responses reported for each situation.

and role-play predictions for Zenith Investment and Nurses Dispute) with the unaided judgement and role-play predictions for four situations from Armstrong (2001a).

Role-play forecasts were more accurate than predictions by game theorists, which were, in turn, more accurate than unaided judgement predictions. These findings are consistent with Schelling's (1961) observation that role play provides more realistic representations of conflict than do theoretical models, and with Armstrong's (2001b) suggestion that forecasting accuracy for conflicts is positively related to the realism of the forecasting method. The difference in accuracy between prediction by game theorists and prediction using unaided judgement is substantial and is statistically significant at 95 percent ($\chi^2 = 4.3$, degrees of freedom 1, $P = 0.04$).[4] The error was about 12 percent lower across the six situations, thus supporting the claim that game theory can improve the forecasting of conflict situations. The difference in accuracy between prediction by game theorists and prediction by role-play decisions is large and statistically significant ($\chi^2 = 24.6$, degrees of freedom 1, $P < 0.001$). I applied the statistical tests to the Table 1 data in the form of two separate two-by-two contingency tables (two accuracy levels by two methods).

Some participants provided more than one forecast. These participants may have been able to apply what they learned to subsequent forecasts. If this were the case, the game-theory experts were most likely to benefit because they provided an average of five forecasts each and were free to revise their forecasts. Role-players, on the other hand, generally played a role in a single situation, although some were asked for their opinion on the decision that was made in an unrelated situation (Armstrong, 1987; Armstrong & Hutcherson, 1989). Unlike the role-players, the game theorists were not restricted in the time they had to derive their forecasts nor in the resources they could exploit, although my written instructions asked them to avoid conferring with others.

Motivation may have influenced forecasting accuracy. Game theorists had the most reason to be motivated. I told them that I planned to use their forecasts in a comparison of forecasting methods, and that I would publish their names in the resulting paper (Appendix A). The participants playing roles or using unaided judgement, on the other hand, knew they would remain anonymous.

The superiority of game theorists' forecasts over forecasts based on unaided judgement may be due to the game theorists' greater familiarity with conflicts. This possibility does not affect the conclusion that those who wish to predict the decisions made in conflicts would do better to ask game theorists than to rely on judges with no expertise. The relative accuracy of people who are experts on conflicts and who use unaided judgement to predict decisions made in conflicts is a matter for further research.

5.2. Factors that might have disadvantaged game theorists' predictions

Perhaps the game theorists were not really experts. I put respondents through a number of 'filters' before asking them for their forecasts—filters that would reasonably be expected to exclude those who were not experts. These filters were: (a) the composition of the original sample (mostly Game Theory Society members); (b) the initial e-mail message, in which I stated ''I am writing to you because you are a game theory expert''; and (c) the potential respondents' own assessments of their ability to successfully apply a game-theory approach to the problems.

I asked respondents about their experience

[4]Unless otherwise stated, all χ^2 tests are of two-by-two contingency tables with continuity correction (Siegel & Castellan, 1988, Eq. (6.3), p. 116).

with game theory. Their median experience was six and a quarter years. Most of the respondents had enough experience to merit game-theory-expert status. More experience as a game theorist did not, however, lead to more accurate predictions (Table 2).

Perhaps the respondents did not spend enough time on the tasks to demonstrate the strength of a game-theory approach to predicting decisions made by parties in conflict. The respondents were presumably busy people, and I did not offer to pay them for their work. Nevertheless, the respondents responded voluntarily and were, it seems likely, motivated to perform well having taken on the tasks. If, after starting on the task, potential respondents began to doubt their ability to predict the decisions in the time they had available, they could have abandoned the task without risking embarrassment. It seems plausible that the withdrawal of respondents who doubted their own abilities would tend to bias responses towards those from the more capable. The link between confidence and accuracy is, however, tenuous (Arkes, 2001).

The game theorists reported spending nearly 40 minutes, on average, deriving their predictions for each situation, including 10 minutes reading the material. This figure is in keeping with the up to 45 minutes that Armstrong (1977) allowed role-players in his research. One game theorist respondent, who stated apologetically that he had spent 'only' about 20 minutes on each of the situations, went on to write that ''spending more time would have not changed my answers, except if I had communicated with others''. In response to follow-up e-mail messages, nine of the 12 respondents stated that they did not believe that more time would have changed their predictions. Little or no gain in accuracy was associated with spending more time on the forecasting task ($\chi^2 = 0.08$, degrees of freedom 2, $P = 0.96$)[5] (Table 3). Further, only for the Panalba situation was there a monotonic increase in accuracy across the three time categories.

Perhaps the respondents were unable to demonstrate the strength of a game-theory approach to prediction because they lacked adequate information about the situations. Four experts cited inadequate information about the situations as a reason for not participating in the research. Yet these four comprised only four percent of those who provided a reason for not participating.

I designed the research to compare the forecasting accuracy of the methods by providing the *same* material to the practitioners of all three methods. I assumed that the material provided was a fair representation of the type of information that could realistically be assembled about an unfolding conflict. Perhaps the experts could have produced better forecasts had they been provided with *different* information. Perhaps, given equivalent resources, game-theory experts would have collected different types of

Table 2
Game theorists' predictions: the effect of experience on accuracy

Experience	n[a]	% Correct
Fewer than 5 years	(31)	39
Between 5 and 10 years	(35)	40
More than 10 years	(33)	33

[a] Number of predictions.

Table 3
Game theorists' predictions: the effect of time spent on accuracy

Time spent on forecast	n[a]	% Correct
Up to 25 minutes	(36)	36
26 to 40 minutes	(43)	37
More than 40 minutes	(20)	40

[a] Number of predictions.

[5] Test of independence in a two-by-*n* classification (Fisher, 1973, p. 87).

information about the situations and would, as a consequence, have produced better forecasts. Further research may address these issues.

Perhaps the game theorists would have been more accurate in forecasting different types of conflicts. The game-theory experts provided the most accurate forecasts for the Panalba Drug Policy conflict, while their accuracy for the other conflicts was poor. In sympathy with Shubik (1975), Gintis (2000) offered a possible explanation for this dichotomy in the game theorists' forecasting accuracy: the rational actor model from neoclassical economics. He wrote ''. . . when faced with market conditions—anonymous, nonstrategic interactions—people behave like self-interested, outcome-oriented actors . . . In other settings, especially in the area of strategic interactions, people behave quite differently'' (p. 240). The Panalba conflict is an 'anonymous, nonstrategic interaction', and the decision makers in the conflict did behave as 'self-interested, outcome-oriented actors'. The five other conflicts, on the other hand, involved face-to-face strategic interaction and ongoing relationships. This distinction suggests that managers would be sensible to seek forecasts of conflicts from game theorists only if those in conflict are likely to behave as 'rational actors'.

Bennett's (1995) 'missing dimensions' (mentioned earlier) may also help managers to distinguish between conflicts for which game theory can provide accurate forecasts and those for which this is unlikely. For example, these dimensions are weak or absent in the Panalba conflict. Whether managers would be able to use Bennett's 'missing dimensions' or Gintis's 'market conditions' to make the distinction between conflicts that are tractable under game theory is a matter for further research. Ironically, even if such a distinction could be made in advance, Gintis (2000) suggested that game theory may not add value, because, in situations in which it is likely to help make accurate predictions, ''standard neoclassical models do well *without* the intellectual baggage of game theory'' (p. 240).

While Gintis declared that predicting the decisions made in individual real conflicts is ''not what game theory is good at'',[6] other researchers seem to think that game theory can be used to provide useful forecasts for some types of conflict for which 'rational actor' assumptions would seem rather brave. The use of methods based on game theory to forecast civil conflicts has been mentioned (Brams & Togman, 2000; Organski, 2000), as has their use for predicting incumbents' responses to a new competitor (Gruca et al., 1992), for common-pool resource problems (Keser & Gardner, 1999), and for conflicts involving concentrated competition, mutual familiarity, and repeated interaction (Ghemawat & McGahan, 1998).

My research provides some evidence of game theorists' forecasting accuracy for some of these types of conflicts. While none of the six conflicts I used involved sectarian violence, it could be argued that the Artists' Protest was a civil conflict. The game-theory experts, on average, did no better than chance in forecasting the decision made in this conflict. Any pay dispute (for example, the 55% Pay Plan and the Nurses Dispute) meets the conditions listed by Ghemawat and McGahan, if they are taken literally, as does the Zenith Investment conflict. The game-theory experts averaged 34 percent correct for these three conflicts. This compares with 41 percent for unaided judgement, 67 percent for role-playing, and 30 percent for chance. None of the six conflicts involved a new competitor or a common pool resource. Game theorists' predictions may be more accurate for conflicts of these types than they were for the

[6]Personal communication from Herbert Gintis, 15 June 2001.

six I used in this research. Further research may answer this question.

5.3. Did the game theorists use game theory?

For each of their predictions, I asked game theorists, ''Broadly, what approach did you use to derive your prediction?''. The answers to this question were diverse. Five university students, enrolled in an Honours-level economics course that included game theory, rated the responses. Each of these student raters had completed between two and five courses with game-theory content. I told the raters nothing about the research participants or the situations. Specifically, they did not know that the participants were game-theory experts. They each rated up to 97 of 98 responses both for evidence of game-theory knowledge (implicit or explicit) and for the extent to which the responses implied that knowledge of game theory had been applied to making a prediction. The raters derived their ratings independently using a zero ('none') to 10 ('high') scale. The rating questionnaire, which includes the game theorists' responses, is available on the Internet at www.kestencgreen.com.

I calculated Cronbach's Alpha scores to test interrater reliability for the students' knowledge and application ratings. Alphas were 0.69 (n = 73) and 0.42 (n = 55), respectively. I averaged the five raters' ratings to provide two mean ratings for each game-theorist response. The overall average rating for knowledge was 2.6 and for application 2.2. Some game-theorist responses were rated highly: the maximums of the five-rater means were 8.0 and 8.2, respectively. Ratings were mostly quite low, however.

While I asked the raters to assess the responses without giving them contextual information, the game theorists knew the context when they wrote their responses. It is reasonable to assume that they felt no obligation to demonstrate their knowledge of game theory or to describe its application in their (typically brief) responses—after all, I had told them that I selected them because of their game-theory expertise. For these reasons, I examine relative rather than absolute ratings in the following analysis.

The ratings provide modest support for the contention that a more structured application of game theory would have increased the accuracy of game theorists' forecasts. Of the 33 predictions that were associated with a mean knowledge rating of 3.0 or higher, 48 percent were correct. This compares with 33 percent correct for the 64 predictions associated with lower ratings. The difference between the proportions is not significant, however ($\chi^2 = 1.7$, degrees of freedom 1, $P = 0.20$). The proportion correct among the 19 predictions with application ratings of 3.0 or higher was 42 percent compared to 37 percent among the 78 with lower ratings. Again, the difference is not statistically significant ($\chi^2 = 0.02$, degrees of freedom 1, $P = 0.89$). There was no meaningful relationship between the situations and the mean ratings for the situations ($F(5,92) = 0.9$; $P = 0.47$ for both ratings).

Ratings aside, what is important for the purpose of this research is that the respondents were aware that I was assessing game-theory forecasts and that I expected them to apply their game-theory knowledge and skills to the forecasting problems (Appendix A). The responses of the six who considered it inappropriate to apply game theory to the problems lends further support to my contention that respondents knew what was expected of them. If some respondents failed to apply game-theory knowledge to the problems (and some responses suggested this), this implies that these respondents considered that knowledge not useful for, or applicable to, these problems, or that the cost of applying it was too high.

5.4. Alternative assessment of forecast accuracy

Despite the relative *in*accuracy of game-theory experts' forecasts, a manager who obtained several forecasts for a single conflict might still have been led to expect the decision that actually occurred, or at least have been warned that it might occur. This might be the case if the actual decision made in the conflict was the most popular choice of the game theorists or if the most popular choice of the game theorists was *similar* to the actual decision. Nevertheless, it seems reasonable to assume that such a set of forecasts would be less likely to lead a manager to expect the actual decision than a set in which an absolute majority were accurate. A manager is most likely to commit to a course of action when provided with an unambiguous forecast and to reap the benefits of this commitment if the forecast is accurate. I propose the following accuracy-score regime to allow comparisons between the forecasting methods based on the likelihood that a set of forecasts would lead managers to confidently expect the actual decision:

Accuracy of forecast set	Accuracy score
An absolute majority choose the actual decision	2
Actual decision is the most popular choice (≤ 50 percent)	1
Either popular choice or absolute majority choice is *similar* to actual decision	0.5
No clear choice	0
Either popular choice or absolute majority choice is *dissimilar* to actual decision	−0.5
Popular choice is substantially different from the actual decision	−1
Absolute majority choice is substantially different from the actual decision	−2

For both the Panalba and Nurses conflicts, the game theorists' absolute majority choice was the actual decision (Table 1). The accuracy score is therefore 2 for both Panalba and Nurses.

In the case of the Artists' Protest, the most popular choice by game theorists[7] was option C (42 percent). Option C was a compromise between the status quo and the actual decision, which was the government conceding to the artists' demands. This forecast could have alerted the parties to the likelihood that the actual decision would favour the artists. Score: 0.5.

In the case of the Distribution Channel conflict, most game theorists (54 percent) forecast a short-term trial for the scheme, but the actual decision was to make a long-term commitment. The forecast would not have been likely to lead the proposer of the scheme to expect a long-term commitment. Score: −0.5.

In the case of the 55% Pay Plan, the great majority of game theorists (88 percent) forecast a strike by the players, but their forecasts were evenly divided between the three strike options—short, medium, and long. Such a forecast would have led the parties to expect a strike but not necessarily a long strike, which is what occurred. Score: 0.

In the case of the Zenith Investment conflict, the majority of game theorists (56 percent) forecast that no new steel production plants would be purchased. The Board chose to purchase two new plants. The absolute majority game-theorist forecast was completely wrong in this instance. Score: −2.

As Table 4 shows, the total accuracy score for the game-theorist forecasts for the six conflicts is 2.0 out of a possible total of 12. Making a similar calculation for the role-play forecasts is straightforward: for all conflicts except the Artists' Protest, an absolute majority of forecasts was correct and therefore the role-play forecasts score two points for each of these. The popular role-play decision Artists' Protest was option B, which was similar to the actual

[7] $n = 17$, as one game theorist provided a probabilistic forecast.

Table 4
Accuracy of forecasts: an alternative assessment. Accuracy Scores: −2 to +2

	Predictions by game theorists	Role-playing
Artists' Protest	0.5	0.5
Distribution Channel	−0.5	2
55% Pay Plan	0	2
Zenith Investment	−2	2
Panalba Drug Policy	2	2
Nurses Dispute	2	2
Totals (unweighted)	2.0	10.5

decision.[8] In fact, 13 of the 14 role-play decisions either matched the actual decision (option A) or were similar to it (options B and C). This implies a score of 0.5 for Artists' Protest and a total accuracy score for role-playing of 10.5.

5.5. Cost of forecasts

Managers deciding between forecasting methods are often influenced by the relative costs of methods as well as their accuracy. When seeking to forecast decisions in a conflict, asking a neutral observer to describe the situation is likely to be a sensible starting point. A week (40 h) may be enough time for an experienced person to assemble a concise and accurate description of a situation. The actual time, however, will depend upon the importance and complexity of the problem and on any deadlines inherent in the situation. The time taken to compile a situation description is the major part of the fixed cost of forecasting.

The cost per forecast varies across the methods. Students were used for unaided judgement and role-play forecasts in this research. Students' time is likely to cost, at most, one-sixth as much as that of experts. Although most of the forecasts reported in this paper took somewhat less than an hour to produce, in practice it would be difficult to recruit students or experts for less than one hour at a time for a commercial forecasting task. On this basis, the marginal cost of a game-theory forecast is six times more than the cost of an unaided-judgement forecast by a student. The number of role-players needed for a role-play forecast will vary depending on the situation and the judgement of the forecaster. A two-party conflict can be role-played using two students to represent each party; four students in total. The cost of a role-play forecast in this case would be two-thirds the cost of a game-theory forecast. The role-plays of the six situations used in this research involved an average of six role-players per simulation. With six role-players per simulation, the marginal cost is similar to that of the game-theory forecasts: the equivalent of one hour of expert time. Thus 10 independent role-play or game-theory forecasts for a single conflict may cost the equivalent of 50 h of expert time. Ten unaided-judgement forecasts by domain experts would have a similar cost, but the cost of using students as judges would be nearer the equivalent of 42 h of expert time.

6. Conclusions

The primary purpose of my research was to assess the relative worth of unaided judgement, game theory, and role-playing for predicting decisions made in real-world conflicts involving few players and high stakes. To do this, I combined new findings on the accuracy of game-theory experts' predictions, role-play predictions, and predictions based on unaided judgement with role-play and unaided-judgement findings summarised in Armstrong (2001a). I then compared the accuracy of the forecasts from the different methods. The results support the view implied by Schelling (1961) and stated directly by Armstrong (2001a) that

[8] Personal communication from J. Scott Armstrong, 11 January 2002.

role-playing will provide more accurate forecasts than other methods for forecasting decisions in conflicts because it provides more realistic representations.

The predictions of game-theory experts did offer an improvement over the traditional approach: their forecasts were more accurate, on average, than unaided judgement. Role-play predictions were better than chance and unaided judgement for *all* situations, and better than game-theory experts' predictions for all but one situation. Game theorists' forecasts, on the other hand, varied more widely in their accuracy than did role-play forecasts. Their forecasts were less accurate than the forecasts of people using unaided judgement for two of six situations and were no better than chance for four of the six situations. Game theorists are experts on conflicts, whereas the other research participants were not. Further research would be necessary to determine whether game-theory expertise confers any advantage over the unaided judgement of experts on conflicts who are *not* familiar with game theory. The cost of the forecasting methods is similar.

Acknowledgements

I am grateful for the cooperation of the game theory experts (listed in Appendix C) who contributed their judgements and their helpful comments on this research. Without their help, my research would not have been possible. I also thank J. Scott Armstrong, Urs Daellenbach, John Davies, six anonymous reviewers who provided helpful advice and criticism, David Harte and Shirley Pledger who provided advice on statistical methods, and Mary Haight who made many useful editorial suggestions. This paper is, in part, based on research funded by the Public Good Science Fund administered by the Foundation for Research Science and Technology (FRST contract: Vic 903). The contract is administered by Raymond Harbridge and Pat Walsh.

Appendix A. E-mail messages to game theorists

First e-mail message

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, PO Box 5530, Wellington, New Zealand. Tel.: + 64-4-499-2040; fax: + 64-4-499-2080. e-mail: kesten.green@vux.ac.nz

Second e-mail message

Subject: Help with research on game theory.

Dear Dr Y

About three weeks ago I sent you an email asking for your help with research I am conducting on game theory. I really need more responses, and I wonder whether you might consider responding.

The material I originally sent is repeated below. If you are unable to read the attachments, please let me know and I will send you .txt files instead.

Regards
Kesten Green
etc.

Appendix B. Game theorist responses

Not appropriate to apply game theory to the problems provided (six responses)

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''.

Insufficient information to derive a prediction (four responses)

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''.

Unresolved responses (51)

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 C. Game theorist respondents

The following respondents completed the set tasks (above the line) or provided useful comments on the research topic (below the line):

Manel Baucells	Holger Meinhardt	Maurice Salles
Emilio Calvo	Claudio Mezzetti	Giorgos Stamatopoulos
Gary Charness	Hannu Nurmi	Tristan Tomala
Bereket Kebede	Andre Rossi de Oliveira	Yelena Yanovskaya
Somdeb Lahiri	Ronald Peeters	Shmuel Zamir
Massimiliano Landi	Alex Possajennikov	José Zarzuelo
Andy McLennan	Eleuterio Prado	Anthony Ziegelmeyer
Peter Bennett	Vito Fragnelli	Harold Houba
Pierre Bernhard	Herbert Gintis	Marc Kilgour
Steven Brams		

References

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, pp. 495–515.

Armstrong, J. S. (1977). Social irresponsibility in management. *Journal of Business Research*, *5*, 185–213.

Armstrong, J. S. (1987). Forecasting methods for conflict situation. In Wright, G., & Ayton, P. (Eds.), *Judgemental forecasting*. Chichester: Wiley, pp. 157–176.

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, pp. 15–30.

Armstrong, J. S. (2001b). Selecting forecasting methods. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic, pp. 365–386.

Armstrong, J. S. (2001c). Standards and practices for forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic, pp. 679–732.

Armstrong, J. S. (2001d). The forecasting dictionary. In Armstron, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic, pp. 761–824.

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., & Hutcherson, P. D. (1989). Predicting the outcome of marketing negotiations. *International Journal of Research in Marketing*, *6*, 227–239.

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.

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. (Ed.), (1970). *Mismarketing: case histories of marketing misfires*. New York: Doubleday, pp. 87–131.

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, pp. 325–342.

Bullock, A., & Trombley, S. (Eds.), (1999). *The new fontana dictionary of modern thought*, 3rd ed.. London: Harper Collins.

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.

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.

Fisher, R. A. (1973). *Statistical methods for research workers*, 14th ed.. New York: Hafner.

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.

Gintis, H. (2000). *Game theory evolving: a problem-centred introduction to modelling strategic behavior*. Princeton, NJ: Princeton University Press.

Graef, R. (1976). Decision Steel. Grenada Colour Productions.

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.

Henderson, H. (1998). Viewing 'the new economy' from diverse forecasting perspectives. *Futures*, *30*(4), 267–275.

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.

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.

Kirshenbaum, J. (1982). Right destination, wrong track. *Sports Illustrated*, 1 February, 7.

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 USA*, *97*(7), 3777–3781.

Mintz, M. (1969). F.D.A. and Panalba: a conflict of commercial and therapeutic goals. *Science*, *165*, 815–881.

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.

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, pp. 343–359.

Parmigiani, S., Ferrari, P. F., & Palanza, P. (1998). An evolutionary approach to behavioral pharmacology: using drugs to understand proximate and ultimate mechanisms of different forms of aggression in mice. *Neuroscience and Biobehavioral Review*, *23*(2), 143–153.

Rapoport, A., & Orwent, 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.

Sandholm, W. H. (1998). History-independent prediction in evolutionary game theory. *Rational Society*, *10*(3), 303–326.

Schelling, T. C. (1961). Experimental games and bargaining theory. *World Politics*, *XIV*(1), 47–68.

Schwenk, C. R. (1995). Strategic decision-making. *Journal of Management*, *21*(3), 471–493.

Shubik, M. (1975). *Games for society, business and war*. Amsterdam: Elsevier.

Siegel, S., & Castellan, Jr. N. J. (1988). *Nonparametric statistics for the behavioral sciences*, 2nd ed.. Singapore: McGraw-Hill.

Sonnegard, J. (1996). Determination of first movers in sequential bargaining games: an experimental study. *Journal of Economic Psychology*, *17*(3), 359–386.

Statman, M., & Tyebjee, T. T. (1985). Optimistic capital budgeting forecasts: an experiment. *Financial Management*, *Autumn*, 27–33.

Suleiman, R. (1996). Expectations and fairness in a modified Ultimatum game. *Journal of Economic Psychology*, *17*(5), 531–554.

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.

Biography: Kesten GREEN is a Research Associate of the Industrial Relations Centre at the School of Business and Public Management at Victoria University of Wellington and Managing Director of Decision Research Ltd, an independent research company. Kesten was also a co-founder of the New Zealand economic forecasting and consulting firm Infometrics Ltd.