ELSEVIER

International Journal of Forecasting 18 (2002) 383–387

international journal of forecasting

www.elsevier.com/locate/ijforecast

Game theory, game theorists, university students, role-playing and forecasting ability

George Wright

Graduate School of Business, University of Strathclyde, 199 Cathedral Street, Glasgow G4 0QU, UK

Abstract

This comment argues that game theory is limited in its ability to analyse and forecast the outcomes of one-off, real-world, conflict situations. It also argues that we have no reason to expect that game theorists would be any better at such forecasts than university students because the conflicts studied by Green did not, on the face of it, have ecological validity for his game theorist participants. Nevertheless, students and game theorists have their own, individual experiences of real-life conflicts and their resolutions. My thesis is that when individuals are enmeshed in role-play simulations of conflicts the relevance of this prior learning becomes salient and can be utilised for the accurate forecasting of the outcomes of the conflicts.

Keywords: Game theory; Conflicts; Real-world situations

Can game theory aid in forecasting the outcomes of conflict situations?

In the feature paper, Green shows that game theorists' predictions were more accurate than the unaided judgment of university students. But, when university students were first asked to role-play the participants in six heterogeneous conflict situations their group-based resolutions of the conflict—or the group-based forecasts of the outcomes of these conflicts—were more accurate than those of the game theorists. Green concludes that role-playing 'will be more accurate than other methods for forecasting decisions in conflicts because it provides a more realistic representation.'

E-mail address: wright@gsb.strath.ac.uk (G. Wright).

Armstrong, in commentary on Green, argues that Green's study was well-executed and its design was, in fact, advantageous to the game theorist participants.

Goodwin makes the point that, whilst the game theorists may have made informal use of game theory concepts in their predictions, the characteristics of the real-world conflict situations used by Green may be more or less amenable to game theoretic modelling and subsequent accurate prediction.

Developing this theme, Erev, Roth, Slonim, and Barron argue that game theory is useful because there are relatively well-defined sets of situations in which previous empirical and experimental research has documented the high predictive value of specific game-theoretic

0169-2070/02/$ – see front matter
PII: S0169-2070(02)00028-6

models. They summarise two examples: the aggregated-level outcome of entry job markets, and repeated laboratory-based games—where high predictive power is achieved by replacing rationality-based predictions with empirically-based modelling of adaptive learning by the game participants. Erev et al. also demonstrate how the predictive value of such conceptualisations can, in itself, be forecasted and go on to suggest that the value of the role-playing method could also be improved by explicit computation of predictive value.

Bolton challenges Green's method of testing game theory and argues that role-playing is dependent on game theory—in that knowledge of game theory is necessary in order to make an initial design of the role-plays. Bolton argues that game theory, especially when combined with the results from studies in experimental economics, can be a useful tool for prediction.

In his commentary, Shefrin documents, from his classroom experience, that formal game theory is unable to predict the behaviour of students in classroom-based role-plays undertaken as part of his game theory course. Developing this theme, Shefrin discusses the design of the Spectrum Auctions and the subsequent bids that were obtained. He shows that the base, game-theoretic, design of the auctions evoked non-rational bids from corporate bidders. He argues that insights from behavioural decision research, such as loss aversion and the winner's curse, can help explain bidder behaviour.

Clearly, game theory is limited in its ability to analyse and forecast the outcomes of one-off, real-world, conflict situations such as the ones Green studied. The forecasting of: (i) generalised market behaviour, and (ii) the outcomes of context-free, laboratory gaming seems the best game-theory-based forecasting that is currently possible. Note that the laboratory gaming forecasts are not pure, game theoretic forecasts since they include behavioural adjustments to formal models and necessitate repeated plays—such that the game outcomes which evoke the adjustments are part of the modelling process.

1. Expertise and quality of judgment

Green's paper implicitly suggests that the expert judgment of game-theorists should produce more accurate forecasts than non-expert judgment. But is there strong underpinning evidence to believe that this will be so?

Bolger and Wright (1994) have reasoned that expertise, to be of practical use, should be measurable in improved performance over forecasts or diagnosed given by those people or systems thought of as 'inexpert'. This is, of course, the focus of Green's studies. In their review of the expertise literature within the social and decision science literature, Bolger and Wright documented that 6 of the 20 extant studies had showed 'good' performance by experts in the domains of weather forecasting, the number of tricks to be made in bridge playing, odds forecasting in horse racing, interest rate prediction, and research and development (R&D) outcome prediction. Of the other 14 studies, 9 showed poor expert performance while the remaining 5 showed equivocal performance. Bolger and Wright interpreted this pattern of performance in terms of the 'ecological validity' and 'learnability' of the tasks that were posed to the experts. *Ecological validity* is the degree to which the experts were required to make judgments inside or outside the domain of their professional experience, express their judgments in unfamiliar metrics, or both. *Learnability* is the degree to which good judgment can be learned in the task domain—if objective data and models, or reliable and usable feedback are unavailable, then it may not be possible for a judge in that domain to improve his or her

performance significantly with experience. In such cases, Bolger and Wright argued, the performance of novices and experts is likely to be equivalent. Bolger and Wright concluded that expert performance will be largely a function of the interaction between the dimensions of ecological validity and learnability—if both are high then good performance will be manifest, but if one or both are low then performance will be poor. For example, weather forecasters in the United States and other countries are routinely required to give confidence estimates attached to their forecasts. Forecasts for the next day's weather are succeeded by timely, unconfounded feedback—since very few human interventions can confound the prediction–outcome relationship. It follows that experimental studies of weather forecasters' confidence in their weather predictions will tend to be ecologically valid and that experimental tasks will, most likely, utilize potential weather events where the participating forecasters have experienced prior conditions of learnability (see Murphy & Brown, 1985, for a review). Bolger and Wright, however, also drew attention to subtle factors, other than these task variables, contributing to the quality of *observed* expert performance. These factors include:

1. Task difficulty—for example, an expert pet show judge would not have problems in discriminating between cats and dogs but might make mistakes distinguishing between different breeds.
2. Measurement type—Consider a poorly manufactured 1-m measure that gives highly reliable but invalid measurements of length; clearly, consensus (reliability) among experts does not necessarily mean that the consensus is valid. Reliability is a necessary, but not a sufficient, condition for valid judgments.
3. Power of tests—Here the sensitivity of statistical tests that are used to detect expert versus novice differences are attenuated by the small sample sizes that are typical of empirical studies of (scarce) expert participants.

As the above discussion shows, the label of 'expert' may not indicate that an individual, or group of individuals, is able to make more valid judgments than non-experts—a variety of task and measurement factors have been shown to enhance or degrade the relative validity of expert versus non-expert judgment.

Rowe and Wright (2001b) reviewed the eight extant studies of expert versus lay judgments of risk and analysed them in terms of Bolger and Wright's model. They concluded that it is still unknown whether expert judgment of risk is more veridical than lay judgment because the veracity of risk judgment was not evaluated in seven of the studies. In the single study in which it was evaluated, methodological problems confound the results. Rowe and Wright concluded that, overall, there is, as yet, no reason to doubt Bolger and Wright's (1994) main conclusion: that if the learnability of expert risk judgment is low (i.e., the prediction/feedback loop is weak or nonexistent), then expert judgment of risk is unlikely to be more veridical than that of non-experts. Differences in expert/non-expert risk judgments, if they exist, do not imply that one set of judgments has greater validity than another. 'Believable' risk assessments produced by those recognised as 'experts' are not, necessarily, valid risk assessments.

In terms of Green's study, we have no reason to expect that the game theorists would be any better at forecasting the outcome of conflict situations than the students. This is because, as the commentaries show, Green's practical, real-world conflicts do *not*, on the face of it, have ecological validity for his game theorist participants. Conflict situations of the type used in

Green's studies have not been the focus of game-theory-based research. As such, no differences in expert versus lay forecasting performance can be expected.

2. Why, then, did role-play by non-experts do so well?

Somehow, role-playing brought out the best from the university students—in that their simulations of conflict situations resulted in resolutions of the conflicts that were close to the real-life resolutions. Green's students could not be expected to have strong technical or domain knowledge about each, or all, of the conflict situations. And, as such, the results of studies which have shown the utility of such knowledge in forecasting situations (see Webby, O'Connor, & Lawrence, 2001; Sanders & Ritzman, 2001) is, in this instance, not useful in understanding Green's results.

What then is the essence of role-play that '. . . provides a more realistic representation'? (Green, page xx).

Research on the effectiveness of Delphi provides a clue. Rowe and Wright (1996, 2001a) have shown that the provision of feedback of the *rationales/arguments* for fellow panellists' forecasts is the essential cause of improvements in forecasting accuracy over Delphi rounds. Rowe and Wright found that although individual panelists were less inclined to change their forecasts as a result of receiving 'reasons' feedback—compared to statistical feedback of means/means or no feedback at all—when the panelists did change their forecasts they tended to change towards more accurate responses. In short, less accurate panelists deferred to the better-argued logic underpinning the predictions of the more accurate panelists. It seems intuitively reasonable that such an information exchange of the reasons underpinning a particular panellist's forecast may not be as rich as the mixture of information, and insights, generated and exchanged in the course of a simulation exercise of the type that Green conducted. It is noteworthy that the participants in Rowe and Wright's (1996) study were, like Green's role-players, university students rather than experts.

Another approach to preparing for the future is scenario planning (see Van der Heijden, 1996). Here, the analysis of stakeholder groupings, their power to influence on unfolding future, and their interest in that future are a focus of attention. The aim is to understand the actions and aspirations of particular stakeholder groupings within particular scenarios. Such an analysis seems intuitively similar to role-playing conflict situations in that the output of both scenario planning and role-playing processes will be the surfacing of the motivations and self-interested actions of actors. Intuitively, an understanding/enactment/of actor motivations/behaviour seems more fundamental to predicting/resolving the actual outcomes of conflicts than either technical or domain knowledge. Intuitively, it would seem that one's own experiences of the past resolution of conflicts—perhaps as recalled or previously experienced and including personal as well as non-personal conflicts—would be a strong guide to the prediction/resolution of the outcomes of conflicts. In other words, if the resolution of conflicts are, generally, the result of the operation of basic human motivations and value systems then, in Bolger and Wright's terms, the conditions for ecological validity and prior learnability hold for both game theorists and university students.

In short, Green's game theorists had two sorts of experience that could have been brought to bear as a basis for predicting the outcomes of the conflict situations. As I have argued, the game theoretic experience was not transferable to Green's conflicts. However the game theorists also had their own, individual, experiences of real-life conflicts and their resolutions. Of course the university students will have had had

similar experiences—but perhaps fewer of them since they will, as a group, have been younger. My thesis is that only when individuals are enmeshed in role-play simulations will the relevance of this experience become obvious—since Green's conflicts will, initially, have been seen as outside the domain of this experience at a superficial, face-content, level.

From this analysis, several future studies seem important to unpack the components of role-playing and their role in potentially improving the forecasting of the outcomes of conflicts. For example, does the partial role-playing of conflicts to near-resolution enable individual participants to predict the actual outcomes of both: (i) the (continued) role-play, and (ii) the real situation that the role-play was designed to model? Are older, more experienced, people better able to make forecasts of actual outcomes after such partial role-playing? Does simulation of role-playing—for example by asking individuals to reflect on conflict situations from, in turn, the perspectives of each of the participants—enhance that individual's forecasting ability? In short, given Green's demonstration of the power of role-playing, the underpinning component processes now deserve the attention of focussed research.

References

Bolger, F., & Wright, G. (1994). Assessing the quality of expert judgment: issues and analysis. *Decision Support Systems*, *11*, 1–24.

Murphy, A. H., & Brown, B. G. (1985). A comparative evaluation of objective and subjective weather forecasts in the United States. In Wright, G. (Ed.), *Behavioral decision making*. New York: Plenum.

Rowe, G., & Wright, G. (1996). The impact of task characteristics on the performance of structured forecasting techniques. *International Journal of Forecasting*, *12*, 73–90.

Rowe, G., & Wright, G. (2001a). Expert opinions in forecasting: the role of the Delphi technique. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Boston: Kluwer.

Rowe, G., & Wright, G. (2001b). Differences in expert and lay judgments of risk: myth or reality? *Risk Analysis*, *21*, 341–356.

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustments of statistical forecasts. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Boston: Kluwer.

Van der Heijden, K. (1996). *Scenarios: the art of strategic conversation*. Chichester: Wiley.

Webby, R., O'Connor, M., & Lawrence, M. (2001). Judgmental time series forecasting using domain knowledge. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Boston: Kluwer.