Structured Analogies for Forecasting

Kesten C. Green,* Victoria Management School, Victoria University of Wellington

Wellington, New Zealand

email: kesten@paradise.net.nz

Phone +64-4-976-3243

J. Scott Armstrong, The Wharton School, University of Pennsylvania

Philadelphia, PA 19104

email: Armstrong@wharton.upenn.edu

Phone 610-622-6480

Fax 215-898-2534

February 8, 2004

* Corresponding author. Address correspondence to email address or to home address: 17 Governor Rd, Northland, Wellington, New Zealand.

Structured Analogies for Forecasting

When people forecast, they often use analogies but in an unstructured manner. We propose a judgmental procedure that uses information from analogous situations in a structured way. This procedure involves asking experts to list as many analogies as they can, rate how similar the analogies are to the target situation, and match the outcomes of the analogies with possible outcomes of the target. An administrator would then derive a forecast from the experts' information. We compared structured analogies with unaided judgments for predicting the decisions made in eight conflict situations. These were difficult forecasting problems; the 32% accuracy of the unaided experts was only slightly better than chance. In contrast, 46% of structured analogies forecasts were accurate. Among experts who were independently able to think of two or more analogies, 56% of forecasts were accurate. Combining helped, with modal structured-analogies forecasts accurate for five of the eight conflicts (63%).

Key words: accuracy, analogies, collaboration, combining, conflict, expert, forecasting, judgment.

It seems natural to use analogies when making decisions or forecasts. For example, Breuning (2003) found that one-third of testimony at the Senate hearing on proposals for the first U.S. program for development aid was based on analogies. Khong (1992) concluded that most of the decisions made early in the Vietnam War were based on forecasts derived from analogies. Believing that policy analysts and others would benefit from access to information about analogies, MIT professor Lincoln P. Bloomfield developed a historical database of post-World War II conflicts (web.mit.edu/cascon).

Can analogies help with forecasting? We expect that real situations from the past contain useful information for forecasters. For example, to predict whether a water-fluoridation proposal will succeed in a particular community in New Zealand, one could analyze the many analogous cases in the U.S. In this vein, Pei (2003) analyzed 16 analogies of nation building after military intervention and their outcomes in order to forecast the likelihood of postwar success in Iraq.

When their use is unstructured, we believe analogies are of little value for improving forecast accuracy. Without a structured search, forecasters and decision-makers are unlikely to identify a large number of analogies. Instead, they are likely to find or interpret analogies to support their prior beliefs. Neustadt and May (1986) described how misuses of analogies led U.S. government decision makers to make poor forecasts of the decisions of other governments' leaders. In another example, the Environmental Protection Agency relied on the analogy of Milford Haven in the U.K. when it approved a new oil refinery in Eastport, Maine (Stewart and Leschine, 1986). The EPA believed Milford Haven was the most comparable site, but Stewart and Leschine observed that Milford Haven had not been in operation long enough to provide evidence that it was safe. They were right. The supertanker Sea Empress ran aground near Milford Haven on 15 February, 1996, spilling 70,000 tonnes of crude oil (Canada Centre for Remote Sensing, 1996).

From their litany of poor decisions by political leaders, Neustadt and May (1986) concluded that a structured approach to using analogies should lead to improved prediction. Armstrong (1985, Chapter 6) summarizes evidence that structured methods of judgmental forecasting are more accurate than unstructured ones. When forecasting from analogies, a structured approach can enable better use of knowledge about analogous situations by (1) encouraging experts to consider a wide range of analogies and (2) effectively processing information from the analogies.

Kahneman and Lovallo (1993) report an anecdote that illustrates how structure can affect predictions. Kahneman had worked with a small team of academics to design a new judgmental decision making curriculum for Israeli high schools. He asked each team member to predict the number of months it would take them to prepare a proposal for the Ministry of Education. Predictions ranged from 18 to 30 months. Kahneman then turned to a member of the team who had considerable experience developing new curricula and asked him to think of analogous projects. After some consideration, the man stated that, among the many analogous situations he could recall, about 40% of the teams eventually gave up. Of those that completed the task, he said, none did so in less than seven years. Furthermore, he thought that the present team was probably below average in terms of resources and potential. In the event, the project took eight years to complete.

We investigated the structured use of analogous situations for forecasting. In this paper, we propose a method (structured analogies). Second, we discuss our hypotheses concerning the accuracy of structured-analogies forecasts and present evidence that our principal hypothesis is not obvious. Third, we discuss our search for prior evidence on the accuracy of forecasts based on information from analogies. Fourth, we describe the procedures we used to test our method. Fifth, we present and analyse our findings. Finally, we discuss the limitations of structured analogies, make suggestions for further research, and present our conclusions.

Procedure for forecasting with structured analogies

Forecasting with structured analogies involves four steps: (1) describe the target situation, (2) identify and describe analogies, (3) rate similarity, and (4) derive forecasts.

(1) Describe the target situation

Prepare an accurate, comprehensive, and brief description. To do so, the administrator should seek advice from unbiased experts or from experts with opposing biases. When feasible, include a list of possible outcomes for the target situation to make coding easier.

(2) Identify and describe analogies

Recruit experts who are likely to know about situations that are similar to the target situation. The proper number of experts depends upon how much knowledge they have about analogous situations, the variability in responses among experts, and the importance of obtaining accurate forecasts. Drawing upon the research on the desirable number of forecasts to combine, we suggest enlisting the help of at least five experts (Armstrong, 2001).

Ask the experts to describe as many analogies as they can without considering the extent of the similarity to the target situation. Then, ask them to match their analogies' outcomes with target outcomes.

(3) Rate similarity

Ask the experts to list similarities and differences between their analogies and the target situation, and then to rate the similarity of each analogy to the target.

(4) Derive forecasts

To promote logical consistency and replicability, use predetermined rules to derive a forecast from experts' analogies. Many rules are reasonable to use. For anecdotal support for the use of rules for deriving forecasts from analogies, consider the following. In a story in the *Online Journal*, Speke and Reuter (2003) reported that four military historians predicted that U.S. and British forces would fail to take Baghdad. The four experts cited analogous situations in support of their predictions. Surprisingly, most of their analogies concerned the taking of besieged cities. The experts argued that the successful sieges were unusual or were different from the Baghdad situation.

Hypotheses

We examined predictive validity using conflicts. Prior research has shown that the method currently used for making predictions for conflicts, unaided judgment, produces inaccurate forecasts (see, for example, Green 2002). We hypothesized that the forecasts of experts who used structured analogies to make predictions about conflicts would be more accurate than those of experts who did not.

Is our primary hypothesis obvious? To answer this question, we obtained opinions from four groups of people, attendees at four of our talks on forecasting for conflict situations: (1) twenty-seven academics and students at Lancaster University, (2) five Royal New Zealand Police College educators, (3) eighteen Harvard Business School alumni, and (4) twelve conflict management practitioners in New Zealand. We asked these individuals to assess the accuracy of forecasting methods when used to predict decisions made in five of the conflicts in Table 1: Artists Protest, Distribution Channel, 55% Pay Plan, Nurses Dispute, and Zenith Investment. We described the forecasting methods and told the respondents that each conflict had between three and six decision options, and a forecaster who selected options at random could expect 28% of forecasts to be accurate. A copy of this questionnaire is available at conflictforecasting.com.

Table 1

Conflict Situations

Artists protest: Members of a rich nation's artists' union occupied a major gallery and demanded generous financial support from their government. What will be the final resolution of the artists' sit-in?

Distribution channel: An appliance manufacturer proposed to a supermarket chain a novel arrangement for retailing its wares. Will the management of the supermarket chain agree to the plan?

55% Pay plan: Professional sports players demanded a 55 percent share of gross revenues and threatened to go on strike if the owners didn't concede. Will there be a strike and, if so, how long will it last?

Nurses dispute: Angry nurses increased their pay demand and threatened more strike action after specialist nurses and junior doctors received big increases. What will the outcome of their negotiations be?

Personal grievance: An employee demanded a meeting with a mediator when her job was downgraded after a her new manager re-evaluated it. What will the outcome of the meeting with the mediator be?

Telco takeover: An acquisitive telecommunications provider, after rejecting another seller's offer to sell a mobile business, made a hostile bid for the whole corporation. How will the standoff between the companies be resolved?

Water dispute: Troops from neighboring nations moved to their common border, and the downstream nation threatened to bomb the upstream nation's new dam. Will the upstream neighbour agree to release additional water and, if not, how will the downstream nation's government respond?

Zenith investment: Under political pressure, a large manufacturer evaluated an investment in expensive new technology. How many new manufacturing plants will it decide to commission?

The median of our respondents' expectations was that 50% of forecasts from both unaided judgment and structured analogies would be accurate. While their stated expectations suggested that the respondents believed forecasts would be similarly accurate no matter which method was employed, inspection of their rankings showed a difference. Sixty-two percent of the respondents expected forecasts based on structured-analogies to be more accurate than those based on unaided-judgment.

We compared these expectations with practice, using results from a survey of practitioners conducted by Armstrong, Brodie and McIntyre (1987). When asked how their firms predicted the actions of competitors, 85% of the 59 practitioners said that they used the opinions of experts who knew about the situation. We are not aware of any organizations that have used structured analogies by this or any other name. In effect, then, common practice seems to rely on unaided judgments by experts.

While we supposed that using rules to derive forecasts from experts' analyses of analogies would help accuracy, one might argue against the use of mechanical schemes. Because they understand their own analogies, experts might be able to make forecasts that were more accurate than those made using rules. To test this possibility, after the experts had described their analogies, we asked them to predict the conflict outcome.

Does it help if experts collaborate and discuss analogies with others? Collaboration could help experts to produce more analogies and flesh out the details, or it could hinder them by suppressing their creativity and search. Both positions are reasonable, so we had no prior hypothesis on collaboration. We asked some experts to collaborate with others and to report the number of people they discussed the forecasting problem with.

Prior evidence

Schrodt (2002) searched for empirical evidence on the accuracy of forecasts of decisions in foreign-policy conflicts. His search encompassed forecasts based on analogies, but he found no evidence on the accuracy of such forecasts relative to any other method. Earlier, Tetlock (1999) had found that forecasts by 20 experts of the outcomes of such conflicts were no more accurate than could be expected from chance.

In a marketing study, McIntyre, Achabal and Miller (1993) tested a procedure called case-based reasoning, which is a way to structure analogies, for forecasting sales during sales promotions. When tested on two products, the forecasts were no more accurate than those of an expert buyer.

We conducted a further search for evidence by using the *Social Science Citation Index* (SSCI) for the period 1978 to January 2004 using the terms "analogies" and "forecasting," and then "analogies" and "prediction." We searched the Internet on January 17, 2004 using Google™ and the terms "comparative", "forecasting," "prediction," "accuracy," and "analogies". We conducted similar searches on JSOR. In November 2001, we sent e-mail appeals to the 278 members of the International Institute of Forecasters list server and to the 579 members of the Judgment and Decision Making mailing list. We also contacted key researchers. The only relevant study we uncovered was Buehler, Griffin, and Ross's (1994). They asked 123 participants to estimate how long it would take to complete a computer assignment. Their predictions made using unaided judgment, were overly optimistic. Participants who had been asked to think of analogous situations were more accurate, especially when they described how the analogies related to the assignment. While no more reliable than those using unaided judgment, participants who related analogies to their assignment were less biased, being twice as likely to complete on time.

In sum, prior to the research we describe here, little evidence was available on the accuracy of forecasts based on the use of analogies relative to the accuracy of forecasts made using other methods. Furthermore, no prior evidence exists on the use of structured analogies.

Procedures used for the study

Recruiting experts

To recruit experts, we sent e-mail messages to ten public list servers, two organizations' e-mail lists, the faculty of a university political science department, and a convenience sample of 15 experts. We chose lists that were likely to include high proportions of experts on conflicts or on judgmental forecasting. We took additional steps to ensure people were suitably qualified for these tasks. In our appeals, which were personalized when possible, the lead author wrote "I am writing to you because you are an expert…" and "I am engaged in a research project on the accuracy of different methods for predicting the outcomes of conflicts…". We sent only descriptions of conflicts that were likely to be relevant to the particular recipients. For example, we did not send a situation dealing with a

proposed new marketing channel to experts in employment relationship disputes or political science. We varied the order in which we attached the conflict documents to the e-mail messages. We sent as many as three reminders. Details of the lists and participation are provided at conflictforecasting.com. To test our hypotheses, we sought responses for each of the following treatments:

- unaided judgment (no instructions on how to forecast) without collaboration,
- unaided judgement with collaboration,
- structured analogies without collaboration,
- structured analogies with collaboration.

For our first appeal, we sent equal numbers of each treatment to members of the International Association of Conflict Management mailing list. The structured-analogy and collaboration treatments were more onerous for participants than unaided judgment, so we obtained relatively few responses for those treatments. As a consequence, in most of our subsequent appeals we sought responses for structured analogies with collaboration. Finally, we sought responses for combinations of conflict and treatment for which we needed more forecasts. because we were seeking participants for their expertise, rather than as part of a representative sample of some larger group, random assignment to treatments was unnecessary. The form of collaboration was at the discretion of the participants.

Preparing materials

We provided participants with an e-mail message giving (1) instructions on how to participate, (2) descriptions of conflicts, (3) brief descriptions of the roles of the parties involved in the conflict, and (4) questionnaires.

The conflict descriptions were accounts of real situations. We abstracted all but one (Personal Grievance) from mass media reports or experts' accounts. The first author developed the Personal Grievance from information collected in interviews and from exchanges of e-mail messages with the parties involved in the dispute. In the case of Nurses Dispute, the first author gathered information in addition to that which had been published by interviewing representatives of the two disputant parties. When we considered it necessary, we disguised the conflicts that had already occurred to reduce the chance that our participants would know the outcomes. As a precaution, we asked our

experts whether they recognized the situations. In eight cases, experts correctly identified a conflict. These responses were eliminated.

In all, we examined eight conflict situations (Table 1). Our descriptions were short, running to no more than two pages. The descriptions are provided at conflictforecasting.com. The materials, identity of the disguised conflicts, and descriptions of actual outcomes are available to researchers on request.

Our one-page questionnaire for structured-analogies participants asked the experts to, first, describe each analogous situation; second, describe their source of knowledge about it; third, list similarities and differences compared to the target conflict; and fourth, provide an overall similarity rating (where 0 = no similarity… 5 = similar…10 = high similarity). Finally, we asked the experts to select (from a list of possible outcomes that we prepared for each target conflict) the outcome closest to the outcome of their analogy. To illustrate, we append a completed structured-analogies treatment questionnaire for one of the conflicts, Telco Takeover. We began our questionnaires for other participants by asking them to select the outcome they thought would occur. We gave them the same lists of possible outcomes that we gave to the structured-analogies participants.

Coding responses

We obtained two groups of unaided-judgment forecasts from experts. One was from the unaided-judgment treatment (62 forecasts), and the other from experts who were asked to use structured analogies but could think of no analogies (44 forecasts). We analyzed results separately for each group and the forecasts were similar; the latter group's being somewhat more accurate. We combined the two groups under the title "unaided judgment" for our analyses, reasoning that neither of these groups used structured analyses.

For each conflict, we derived a structured-analogies forecast from each expert's analogy information, where the information was available. It is trivial to derive a forecast from analogies information when an expert provides a single analogy. On the other hand, many mechanical schemes could be used to derive a forecast when an expert provides information on more than one analogy. To obtain a forecast, we selected the target conflict outcome implied by the analogy given the highest similarity rating by the expert. Our reasoning was that predictive validity should increase with relative similarity. Where there was a tie, we selected the outcome that had the most support

from the expert's analysis of analogies. (Details on the rules are provided at conflictforecasting.com). Given our uncertainties about the best procedure, we subsequently analyzed other mechanical schemes.

We coded forecasts as accurate if the outcome option chosen by our rule was substantially the same as, and most similar to, the actual outcome of the conflict. Outcome options were unconditional statements of decisions and did not specify timing, for example, "Expander's takeover succeeded at, or close to, their August 14 offer price of $43-per-share."

Results

In all, 66 experts provided forecasts for between one and eight conflicts (Table 2). As we hypothesized, forecasts from structured analogies were more accurate than unaided-judgment forecasts. They were more accurate for seven of the eight conflicts. Averaging the accuracy figures across the conflicts, structured-analogies forecasts were 46% accurate compared to 32% for the unaided experts' forecasts (using unweighted averages) The difference is statistically significant ($P = 0.04$), using a one-tailed permutation test for paired replicates (Siegel and Castellan, 1988). We used this test to compare the differences in the percentage of correct forecasts between the two methods for each conflict (e.g., for Artists Protest, the difference between structured analogies and unaided judgment was 17%). Viewed another way, by having experts provide analogies rather than forecasts, we reduced the average forecast error by 21% (where forecast error is the percentage of forecasts that were wrong).

Table 2

Accuracy of structured-analogies

and unaided-judgment forecasts by experts

Percent correct forecasts (number of forecasts)

	Chance	**Unaided judgment**		**Structured analogies**	
Telco Takeover	25	0	(8)	**8**	(12)
Artists Protest	17	10	(20)	**27**	(11)
55% Pay Plan	25	18	(11)	**57**	(14)
Personal Grievance	25	31	(13)	**36**	(14)
Zenith Investment	33	36	(14)	**38**	(8)
Distribution Channel	33	38	(17)	**50**	(12)
Water Dispute	33	50	(8)	**92**	(12)
Nurses Dispute	33	**73**	(15)	57	(14)
Averages (unweighted)	28	32	(106)	**46**	(97)

Effect of number of analogies

Tetlock (2004) found that foreign-policy experts who invoked several analogies for a single situation made predictions that were more accurate than those who invoked only one. Consistent with this, we found that forecasts from experts who could think of two or more (plural) analogies were more accurate than those with only a single analogy in six of the eight conflicts. Accuracy went from an average of 38% for those able to describe only one analogy, to 56% for those who described two or more.

Since the decisions in some conflicts were more difficult to forecast than those in others, we next examined the reduction in error versus chance. Basing forecasts on a single analogy reduced error by an average of 15% compared to those based on chance. In contrast, forecasts based on plural analogies reduced error by 39% (Table 3). The difference in error between single-analogy forecasts and plural-analogy forecasts is statistically significant at $P = 0.02$ (one-tailed permutation test for paired replicates). The error could be reduced by 42% versus chance by accepting only forecasts from experts who described three or more analogies. Thus the usefulness of an individual expert was related to the number of analogies she described.

Table 3

Accuracy of forecasts by number of analogies

Percent error reduction vs chance (number of forecasts)

	None[1]		**One only**		**Two or more**	
Telco Takeover	-33	(8)	-33	(5)	**-14**	(7)
55% Pay Plan	-33	(2)	26	(9)	**73**	(5)
Distribution Channel	-19	(5)	0	(6)	**50**	(6)
Artists Protest	-3	(7)	-3	(7)	**40**	(4)
Personal Grievance	20	(5)	0	(8)	**33**	(6)
Water Dispute	25	(8)	**100**	(4)	81	(8)
Zenith Investment	**25**	(6)	-12	(4)	**25**	(4)
Nurses Dispute	**100**	(3)	40	(10)	25	(4)
Error reduction (unweighted)	10	(44)	15	(53)	**39**	(44)
Percent correct forecasts (unweighted)	34		38		**56**	

1 Forecasts from experts we asked to use the structured analogies method, who were unable to think of analogies. We classified these forecasts as unaided judgment forecasts in all our other analyses.

Discussion

Value of experts' experience

We tested whether experts with more experience were more accurate than those with less. We used two measures to test this hypothesis: (1) we asked our experts how many years experience they had as "a conflict management specialist," and (2) we asked them to rate their experience (on a scale from 0 to 10) with situations similar to the target conflict.

Experts with five or more years experience as conflict management specialists were *less accurate* (average across conflicts of 21% error reduction compared to chance) than those with less experience (26%). Furthermore, experts who gave high ratings to their experience with similar conflicts were *less accurate* (16% error reduction) than those who gave themselves lower ratings (31%).

Value of similarity ratings across experts

We made comparisons across experts to see whether forecasts based on analogies that the experts rated as highly similar (eight or higher on the zero-to-ten scale) proved more accurate than those based on lower-rated analogies. Forecasts based on analogies the experts rated as highly similar were no more accurate (an average of 42% accuracy across conflicts for the 39 forecasts) than the other forecasts (46% for the 58 forecasts). Thus, in our study of eight conflicts, comparisons of similarities *across* experts were of little value. This result did not apply to when we selected the analogy the individual expert thought to be the most similar for the conflict under consideration.

Effect of experts' familiarity with their analogies

We expected that when experts had first-hand knowledge of analogous situations, their information would be more useful because they would be likely to know more about the situations than those who knew about them only from others' accounts. To examine our assumption, we coded the experts' notes to show whether the top-rated analogies were based on direct or indirect experience. Our efforts produced fifty forecasts based on analogies from direct experience. These forecasts were substantially more accurate than the 106 forecasts by experts who used their unaided judgement (49% vs 32% when averaged across the eight conflicts with $P = 0.02$ based on a one-tailed permutation test for paired replicates). The forecasts were more accurate for seven out of eight conflicts. Viewed another way, the forecasts based on analogies from the direct experience of experts reduced the average error across

conflicts by 31% (compared to chance) while forecasts that were not based on direct experience provided only a 13% error reduction.

Having found that accuracy increased with the number of analogies an expert provided and was higher when experts based these forecasts on analogies from direct experience, we expected that accuracy would be greater still when the two factors coincided. This proved to be correct. When experts based their forecasts on plural analogies, accuracy averaged across the eight conflicts was 60% when the analogies were from direct experience (23 forecasts) and 33% (20 forecasts) otherwise ($P = 0.08$, one-tailed permutation test for paired replicates).

Mechanical schemes to derive forecasts

To test our expectation that mechanically combining analogical information from individual experts was a reasonable approach, we asked our experts to make predictions after they had recalled and analyzed analogies. As we anticipated, a structured mechanical process was more effective than experts' own judgments. In 22 cases, experts made forecasts that were inconsistent with the outcomes of their own analogies. Of these, 25% were accurate. When we used our mechanical rule to derive forecasts from these experts' analogies, 45% were accurate.

Combining across experts

In general, when several forecasts are available, one should combine them mechanically (Armstrong 2001). We adopted the modal forecast for each conflict as our combined forecast. Combined structured-analogies forecasts were correct for five of the eight conflicts (63%), a substantial improvement over the average of 46% accuracy for individual forecasts from the method. As experts' unaided-judgement forecasts were no better than chance, combining did help.

When experts thought of more that one analogy, our mechanical scheme did not use all of the analogical information to make predictions. We tested four alternative approaches that used all of the 210 analogies with similarity ratings and implied decisions in order to determine whether we would improve accuracy further if we used all available analogical information to derive combined forecasts. For our first alternative, we used the outcome implied by the most analogies, and obtained an average accuracy of 40% across all conflicts, compared to 46% for the approach we had adopted. For the second, instead of assuming that the analogies were all of equal value as we did in the first

alternative, for each conflict we chose the option with the highest aggregate similarity rating as the forecast (39% accurate). For the third alternative, we reallocated each expert's total number of analogies for a conflict to decision options in proportion to each option's share of the expert's similarity ratings (40%). For the fourth alternative, we calculated each expert's average similarity rating for each option; we then allocated all of each expert's analogies to each option in proportion to the average similarity rating for the decision as a fraction of the sum of the expert's average similarity ratings (39%). In short, all of these alternatives were less accurate.

The accuracy of experts' unaided-judgment forecasts was not increased by combining. Nor would one have expected improvement, as the accuracy of the individual experts was no better than chance.

Effect of collaboration

While we had no directional hypothesis about collaboration, we analyzed the data to see whether collaboration among experts might be useful. When experts using structured analogies collaborated with others, their median working time was 45 minutes compared to 30 minutes for those who worked alone. (We do not know how much time the collaborators spent on the task, nor do we know the nature of their collaboration.) As it happened, those who collaborated claimed to have had much more experience with conflict-management (median of 14 years versus 5 years) and experience with similar conflicts (a median self-rating of 4.0 out of 10, versus 2.8). Despite the greater investment of resources by more knowledgeable experts, collaboration produced no gain in accuracy: Forecasts from solo experts were on average 44% accurate across conflicts (75 forecasts), compared to 42% for forecasts by collaborating experts (22 forecasts).

Given our findings, we saw no need to distinguish between solo and collaborative forecasts in our analysis. In view of the time savings, we recommend that structured analogies be done by individuals.

Limitations

The structured analogies method is useful only in cases in which experts can think of analogies. This limitation can be overcome in many situations by identifying people with relevant expertise.

Using structured analogies is more costly than using unaided judgment. However, relative to the costs of making bad decisions in many conflict situations, such as selecting strategies to achieve peace in the Middle East or to deal with threatening behavior by the North Korean government, the costs are negligible.

Further research

Research on additional situations would help to better assess the improvements that might be expected, and the conditions under which structured analogies is most effective. Our conclusions are based on a sample of only eight situations.

Much needs to be done to develop the operational procedures for using structured analogies. For example, what is the best way to frame the issues for the experts so that they provide more and better analogies? Would a more structured approach to rating analogies' similarity to a target help administrators derive more accurate forecasts? To what extent might improvements in accuracy be obtained, in the case of well-documented analogies, by checking the facts of the situation and correcting any errors in experts' matching of analogy outcomes with target outcomes.

Like previous studies, our study showed that combining is useful. While we used simple combining, it seems likely that a more structured approach, such as the Delphi technique, could offer further gains at a low cost. Rowe and Wright (2001) provide evidence on the value of Delphi, and software for implementing of Delphi is provided at forecastingprinciples.com. Experts' confidence ratings may be useful for weighting structured-analogies forecasts in a combination (Arkes, 2001).

We have examined conflict situations because of their importance and the difficulty of obtaining useful forecasts. However, structured analogies might improve forecasting in situations other than conflicts.

Research is needed on how to encourage adoption of structured analogies. Currently, people use unaided judgment, a method that is little better than chance, to decide whether to go to war, get a divorce, make a hostile takeover bid, go on strike, or mount a competitive pricing campaign. Better forecasts would aid decision making in such situations.

Conclusions

Structured analogies improved predictions for decisions made in conflict situations. While the single most important criterion for identifying an expert was the number of analogies she generated, another important criterion was whether she had direct knowledge of her own analogies. When an expert produced two or more analogies from her own experience, forecasts based on the information she provided were likely to be accurate. Collaboration with other experts was not helpful.

The improvement in accuracy achieved by structured analogies forecasts was substantial: a 35% reduction in error for forecasts based on plural analogies compared to forecasts made by unaided experts. Given the importance of forecasts in conflict situations and other arenas, such improvement could have considerable benefits.

As with previous studies, combining proved useful, resulting in accurate structured-analogies forecasts for five of the eight conflicts (63%). This is an improvement on the accuracy of individual structured-analogies forecasts. More importantly, it is a substantial improvement on the accuracy of experts' unaided judgements, whether combined or not.

Appendix

Telco Takeover Bid

1) (A) In the table below, please briefly describe
(i) your analogies,
(ii) their source (e.g. your own experience, media reports, history, literature, etc.), and
(iii) the main similarities and differences between your analogies and this situation.
(B) Rate analogies out of 10 (0 = no similarity… 5 = similar… 10 = high similarity).
(C) Enter the responses from question 2 (below) closest to the outcomes of your analogies.

(A) (i) description,	(ii) source,	(iii) similarities & differences	(B) Rate	(C) Q2
a. Bank takeover	**Personal**	**Issue same, industry different**	**8**	**C**
b. Govt Agency merger	**Personal**	**Takeover same, government, but ordered takeover**	**4**	**D**
c. Facility Merger	**Personal/family**	**Combine similar operations**	**3**	**B**
d.				
e.				

2) How was the standoff between Localville and Expander resolved? *(check one ✓, or %)*
a. Expander's takeover bid failed completely [__]
b. Expander purchased Localville's mobile operation only [__]
c. Expander's takeover succeeded at, or close to, their August 14 offer price of $43-per-share [X_]
d. Expander's takeover succeeded at a substantial premium over the August 14 offer price [__]

3) If you have *not* given a prediction, please state your reasons:

4) Roughly, how long did you spend on this task?
{include the time you spent reading the description and instructions} [_1__] hours

5) How likely is it that taking more time would change your forecast?
{ 0 = almost no chance (1/100) … 10 = practically certain (99/100) } [_0_] 0-10

6) Do you recognise the actual conflict described in this file? Yes [__] No [X__]
If so, please identify it: **[_____]**

7) How many people did you discuss this forecasting problem with? [_2___] people

8) Roughly, how many years experience do you have as a conflict management specialist? [20+] years

9) Please rate your experience (out of 10) with conflicts <u>similar to this one</u> [6____] 0-10

When you have completed this questionnaire, please return
<u>either</u> this document as an email attachment to…
<u>or</u> this questionnaire (with your initials at right) by fax to… **Your initials:** [_XYZ_]

Acknowledgements

We thank the experts who participated in the research reported here. They included Roderic Alley, Barry Anderson, Don Baker, Corrine Bendersky, Constant Beugre, Doug Bond, Michelle Brackin, José Ramón Cancelo, Nihan Cini, David Cohen, Ike Damayanti, Serghei Dascalu, Nikolay Dentchev, Ulas Doga Eralp, Miguel Dorado, Erkan Erdil, Jason Flello, Paul Gaskin, Andrew Gawith, Kristian Skrede Gleditsch, Joshua Goldstein, David Grimmond, George Haines, Claudia Hale, Ragnar Ingibergsson, Patrick James, Michael Kanner, John Keltner, Daniel Kennedy, Susan Kennedy, Oliver Koll, Rita Koryan, Talha Köse, Tony Lewis, Zsuzsanna Lonti, Dina Beach Lynch, David Matz, Bill McLauchlan, Kevin Mole, Ben Mollov, Robert Myrtle, W. Bruce Newman, Randall Newnham, Konstantinos Nikolopoulos, Glenn Palmer, Dean G. Pruitt, Perry Sadorsky, Greg Saltzman, Amardeep Sandhu, Marlies Scott-Wenzel, Deborah Shmueli, Mike Smith, Marta Somogyvári, Harris Sondak, Dana Tait, Scott Takacs, Dimitrios Thomakos, William Thompson, Ailsa Turrell, Bryan Wadsworth, James Wall, Daniel Williams, Christine Wright, Becky Zaino. We also thank Lisa Bolton, Nikolay Dentchev, Don Esslemont, Stanley Feder, Paul Goodwin, Clare Harries, Oliver Koll, and Tom Yokum for providing pre-submission peer review. Editorial assistance was provided by Mary Haight and Marian Lee.

References

Arkes, H. R. (2001). Overconfidence in judgmental forecasting, in J. S. Armstrong (ed.) *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (2001). Combining forecasts, in J. S. Armstrong (ed.) *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Armstrong, J. S. (1985). *Long-Range Forecasting*. New York: John Wiley. Full text at hops.wharton.upenn.edu/forecast.

Armstrong, J. S., Brodie, R. J., & McIntyre, S. H. (1987). Forecasting methods for marketing: Review of empirical research. *International Journal of Forecasting, 3*, 355-376. Full text at forecastingprinciples.com.

Breuning, M. (2003). The role of analogies and abstract reasoning in decision-making: Evidence from the debate over Truman's proposal for development assistance. *International Studies Quarterly, 47*, 229-245.

Buehler, R., Griffin, D., Ross, M. (1994). Exploring the 'planning fallacy': Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67*, 366-381.

Canada Centre for Remote Sensing (1996). 'Sea Empress' oil spill monitoring. *RADARSAT Image*, Milford Haven, Wales, United Kingdom, February 22, 1996. Retrieved July 1, 2003, from http://www.ccrs.nrcan.gc.ca/ccrs/data/satsens/radarsat/images/uk/ruk01_e.html.

Green, K. C. (2002). Forecasting decisions in conflict situations: A comparison of game theory, role-playing, and unaided judgement. *International Journal of Forecasting, 18*, 321-344. Full text at forecastingprinciples.com.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science, 39*, 17-31.

Khong, Y. F. (1992). *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965*. Princeton, NJ: Princeton University Press.

McIntyre, S. H., Achabal, D. D., & Miller, C. M. (1993). Applying case-based reasoning to forecasting retail sales. *Journal of Retailing, 69*, 372-398.

Neustadt, R. E., & May, E. R. (1986). *Thinking in Time: The Uses of History for Decision Makers*. New York: Free Press.

Pei, M. (2003). Lessons of the past. *Foreign Policy, 137*, 52-55. Available from http://www.foreignpolicy.com/story/story.php?storyID=13752

Rowe, G. & Wright, G. (2001). Expert opinions in Forecasting: The role of the Delphi technique, in J. S. Armstrong (ed.) *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Schrodt, P. A. (2002). Forecasts and contingencies: from methodology to policy. Paper presented at the American Political Science Association meetings, Boston, 29 August – 1 September. Retrieved December 10, 2002, from www.ku.edu/~keds/pdf.dir/Schrodt.APSA02.pdf.

Siegel, S. & Castellan, N. J. Jr. (1988). *Non-parametric Statistics for the Behavioral Sciences*, 2nd ed. Singapore: McGraw-Hill.

Speke, C., & Reuter, S. (2003). German military historians predict Anglo-American defeat in Iraq. *Online Journal*, March 29. Available from http://morris.wharton.upenn.edu/forecast/Conflicts/PDF%20files/Iraqi%20conflict%20forecast.pdf.

Stewart, T. R., & Leschine, T. M. (1986). Judgment and analysis in oil spill risk assessment. *Risk Analysis,* 6, 305-315.

Tetlock, P. E. (1999). Theory driven reasoning about possible pasts and probable futures: Are we prisoners of our perceptions? *American Journal of Political Science*, *43*, 335-366.

Tetlock, P. E. (2004). Knowing the limits of one's knowledge: Foxes are better calibrated than hedgehogs, in P. E. Tetlock, *Clashing Conceptions of Good Judgment: Cognitive Styles, Biases and Correctives*. New York: Cambridge University Press. In press.