VOLUME 23, NUMBER 3   ISSN 0169-2070

ELSEVIER

international journal of forecasting

IIF International Institute of Forecasters

Structured analogies for forecasting

Kesten C. Green [a,*], J. Scott Armstrong [b,1]

[a] *Business and Economic Forecasting Unit, Monash University, c/o PO Box 10800, Wellington, New Zealand*
[b] *The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, United States*

Abstract

People often use analogies when forecasting, but in an unstructured manner. We propose a structured judgmental procedure whereby experts list analogies, rate their similarity to the target, and match outcomes with possible target outcomes. An administrator would then derive a forecast from the information. When predicting decisions made in eight conflict situations, unaided experts' forecasts were little better than chance, at 32% accurate. In contrast, 46% of structured-analogies forecasts were accurate. Among experts who were able to think of two or more analogies and who had direct experience with their closest analogy, 60% of forecasts were accurate. Collaboration did not help.
© 2007 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

Keywords: Availability; Case-based reasoning; Comparison; Decision; Method

It seems natural to use analogies when making decisions or forecasts, as by definition they contain information about how people have behaved in similar situations in the past. One behavioral scientist asserted that "…we may explain human behavior by assuming that decisions are made by analogy with previous cases…" (Kokinov, 2003, p. 168).

The use of analogies is not a recent phenomenon; for example, analogies were commonly used for economic and business forecasting in the 1930s, and their use was described in text books of the time (Goldfarb, Stekler, & David, 2005).

More recently, the use of analogies has become a popular solution to the problem of predicting the cost of software development projects. In a field study of 598 organizations, 61% of those who reported forecasting the cost of software projects kept data on previous projects and predicted the cost of new projects by analogy (Heemstra, 1992). A Google search using the term "software cost estimation" in February 2006 yielded about 58,200 sites. One business has been collecting data on software projects since the mid-1970s for the purpose of helping others make predictions (Myers, 1989).

We expected that analogies would be useful in forecasting decisions in conflict situations because analogies provide useful information for situations that are quite difficult to forecast. This is a common belief. Khong (1992) concluded that most of the decisions made early in the Vietnam War were based

---
* Corresponding author. Tel.: +64 4 976 3245; fax: +64 4 976 3250.
 E-mail addresses: kesten@kestencgreen.com (K.C. Green),
Armstrong@wharton.upenn.edu (J.S. Armstrong).
[1] Tel.: +1 610 622 6480; fax: +1 215 898 2534.

on forecasts derived from analogies. Breuning (2003) found that one-third of the testimony at the Senate hearing on proposals for the first U.S. program for development aid was based on analogies. Believing that analogical information is useful, conflict management researchers have compiled databases. For example, MIT professor Lincoln P. Bloomfield has assembled a historical database of post-World War II conflicts (web.mit.edu/cascon) in order to help policy analysts and others identify appropriate analogies.

Kahneman and Lovallo (1993) reported an anecdote that illustrates how inducing an expert to use analogies in a structured way 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 draft 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. The project took eight years to complete.

1. Hypothesis

We agree that information about analogies should be useful for forecasting. In some situations, such as when a real estate agent recommends a selling price for your house or a car salesman sets the price for a second-hand Honda, the informal use of analogies is likely to provide useful forecasts. We suspect, however, that in many situations people will choose inferior analogies if they do not use a structured approach. As suggested by the availability heuristic (Tversky & Kahneman, 1973), people will tend to choose analogies that are easy for them to recall. Furthermore, analogies that are easy to recall are likely to be those that confirm people's beliefs. In other words, the use of analogies is subject to biases. For example, when the U.S. Environmental Protec-

tion Agency approved a new oil refinery in Eastport, Maine, decision makers relied on the analogy of Milford Haven in the U.K. (Stewart & Leschine, 1986). The EPA decision makers considered that Milford Haven was the most comparable site and looked no further, 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 metric tons of crude oil (Canada Centre for Remote Sensing, 1996).

Neustadt and May (1986) described how the inappropriate selection and inadequate analysis of analogies led U.S. government decision makers to make poor forecasts of the decisions of other governments' leaders. Drawing on their litany of poor decisions by political leaders, they described a structured approach to analyzing current and historical information. They suggested this should lead to a more effective use of experts' knowledge, and hence to improve prediction. For example, they suggested examining the similarities and differences between analogies and the target situation.

Research in many areas of judgmental decision making and forecasting has shown that structured judgmental processes make more effective use of the information that people possess. This occurs, for example, when people are asked explicitly to decompose a problem (MacGregor, 2001). More generally, Armstrong (1985, Chapter 6) summarized the evidence that structured methods of judgmental forecasting are more accurate than unstructured ones. A structured approach to forecasting with analogies, then, might encourage experts to consider more information from the analogies, and to process it in a more effective way. In contrast, we suspect that experts using their unaided judgment often make forecasts, then search for analogies to support them.

We propose that analogies will only improve accuracy when an objective process is used for their identification and analysis. In order to test our principal hypothesis, we examined the predictive validity of a structured use of analogies for forecasting decisions in conflicts. This is a difficult forecasting task; prior research has shown that the method currently used, unaided judgment, produces inaccurate forecasts (see, for example, Green & Armstrong, 2007). We

hypothesized that forecasts derived from an expert's structured analysis of analogies would be more accurate than forecasts by experts who used their unaided judgment.

2. Prior evidence

We searched for evidence on methods for forecasting with analogies. Schrodt (2002) searched for empirical evidence on the accuracy of forecasts of decisions in conflicts in the foreign-policy arena. He found no evidence on the accuracy of forecasts based on analogies relative to that of forecasts based on any other method.

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.

Shepperd and Schofield (1997) compared forecasts of software development costs from analogies with forecasts from models estimated using stepwise regression. The completion costs of historical cases most similar to the target were averaged to provide analogies forecasts. The analogies forecasts were more accurate for all nine data sets they used on the basis of mean absolute percentage errors (MAPEs). The software that the authors used to derive the analogies forecasts is available at http://dec.bournemouth.ac.uk/ESERG/ANGEL/.

Using similar procedures to Shepperd and Schofield (1997), Angelis and Stamelos (2000) found that analogies forecasts were somewhat more accurate for one data set, but were markedly less accurate for a second. The authors suggested that where there are sufficient data and strong relationships, regression models are likely to outperform analogy methods.

Comparability analysis is a procedure developed by the US Air Force for forecasting by analyzing analogous data. In a study on attendance at a small-town boutique movie theatre, Klein (1998) compared the accuracy of forecasts of attendance at 35 movies from comparability analysis with both the accuracy of the theatre manager's forecasts and the accuracy of the median forecasts of 17 locals. The correlations between the forecasts and actual attendance were, respectively, 0.45, 0.31, and 0.17.

We conducted a further search for evidence by using the *Social Sciences Citation Index* for the period from 1978 to August 24, 2004 using the terms "analogies" and "forecasting," and then "analogies" and "prediction." We also searched the Internet in August 2004 using Google and the terms "comparative", "forecasting," "prediction," "accuracy," and "analogies". We conducted similar searches on JSTOR. In November 2001, we sent e-mail appeals to 278 members of the International Institute of Forecasters list server and to 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 inaccurate and overly optimistic. Predictions by participants who had been asked to think of analogous situations were less biased, especially when they described how the analogies related to the assignment. Unrealistic optimism was reduced substantially.

While modest, prior research shows that the use of analogies can provide some improvement in accuracy relative to the accuracy of forecasts from other methods. Little, however, has been done to identify how the use of analogies might be most effectively structured and under what conditions their use is most beneficial.

3. Procedure for forecasting with structured analogies

Experts often have useful information about analogies, but they process it in ways that are subject to biases. This is especially likely for emotionally charged topics. Thus, we expected that a structured process could substantially improve the use of experts' information, and thereby improve the accuracy of forecasts.

Our structured approach to using analogies for forecasting requires experts to identify analogies and their outcomes, and to assess the similarity of each of the analogies to the target in a structured way. The procedure involves five steps: First, the administrator (1) describes the target situation, and (2) selects experts; then the experts each (3) identify and describe analogies, and (4) rate similarity; and finally, the administrator (5) derives forecasts.

3.1. Describe the target situation

The administrator prepares an accurate, comprehensive, and brief description. To do so, the administrator should seek advice either 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.

3.2. Select experts

The administrator recruits experts who are likely to know about situations that are similar to the target situation. The administrator should decide how many experts to recruit based on 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).

3.3. Identify and describe analogies

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

3.4. 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. We suggest providing a scale against which the experts can rate the similarity of their analogies.

3.5. Derive forecasts

To promote logical consistency and replicability, the administrator should decide on the rules for deriving a forecast from experts' analogies. Many rules are reasonable to use. For example, one could select the analogy that the expert rated as most similar to the target, and adopt the outcome implied by that analogy as the forecast.

Our structured analogies procedure is based on the assumption that while unaided experts can provide useful information, they are not good at processing complex information. For that reason, we did not rely on the experts to make forecasts, but instead used a rule. On the other hand, perhaps experts' understanding of their own analogies might enable them to forecast more accurately than we could by using rules. To test this aspect of our procedure, we asked our experts to predict the decision made in the target situation after they had described and rated their analogies.

Does it help if experts collaborate and discuss their analogies with others? Collaboration could either 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 all experts were asked to report the number of people with whom they discussed the forecasting problem.

4. Procedures used for the study

4.1. Preparing materials

We compiled descriptions of conflicts, including brief descriptions of the roles of the parties involved in the conflicts. The conflict descriptions were all accounts of real situations. We abstracted all but one (Personal Grievance) from mass media reports or experts' accounts. The lead author developed the Personal Grievance description 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, he gathered information from published sources (Langdon, 2000a,b, c; Radio New Zealand, 2000a,b,c) and by interviewing representatives of the two disputant parties. When we considered it to be 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, and their responses were eliminated.

In all, we used eight conflict situations in our research. We provided between three and six possible outcome options for each of them (Table 1). Our descriptions were short, running to no more than two

pages. The full descriptions are provided at conflictforecasting.com. The materials, the identity of the disguised conflicts, and descriptions of actual outcomes are available to researchers on request.

4.2. Selecting experts

To select 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 that 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

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? (6 options)

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? (3 options)

55% Pay plan: Professional sports players demanded a 55% 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? (4 options)

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? (3 options)

Personal grievance: An employee demanded a meeting with a mediator when her job was downgraded after her new manager re-evaluated it. What will be the outcome of the meeting? (4 options)

Telco takeover: An acquisitive telecommunications provider, after rejecting a seller's mobile business offer, made a hostile bid for the corporation. How will the standoff between the companies be resolved? (4 options)

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 neighbor agree to release additional water, and if not, how will the downstream nation's government respond? (3 options)

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? (3 options)

am engaged in a research project on the accuracy of different methods of predicting the outcomes of conflicts..." (Appendix A). We only sent 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. Most importantly, we counted on people to recognize when they had expertise on a topic, and we asked them about their experience.

We sent as many as three reminders. Details of the lists and participation are provided at conflictforecasting.com.

4.3. Using the methods

In our e-mail appeal, we gave experts instructions on how to participate (Appendix A). For structured-analogies participants, our one-page questionnaires asked the experts to (1) describe each analogous situation along with the outcome; (2) describe their source of knowledge about it; (3) list the similarities and differences compared to the target conflict; and (4) 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, a completed structured-analogies treatment questionnaire for one of the conflicts, Telco Takeover, is provided as Appendix B.

Questionnaires for unaided-judgment participants first asked 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.

We varied the order in which we attached the conflict documents to our e-mail appeals. To test our hypotheses, we sought responses for each of the following treatments with our appeals:

1. unaided judgment (no instructions on how to forecast) without collaboration,
2. unaided judgment with collaboration,
3. structured analogies without collaboration,
4. 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-analogies 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 the 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, a random assignment to treatments was unnecessary. The form of collaboration was at the discretion of the participants.

4.4. Coding responses

We obtained two groups of unaided-judgment forecasts from experts. One was from the unaided-judgment treatment (62 forecasts), and the other was from experts who were asked to use structured analogies, but could think of no analogies (44 forecasts). We analyzed the results separately for each group; the forecasts were similar, with the latter group 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 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 the predictive validity should increase with the relative similarity. Where there was a tie, we selected the outcome that had the most support from the expert's analysis of the analogies. (Details on the rules for determining support are provided at conflictfor-ecasting.com.) Given our uncertainties about the best procedure, we subsequently analyzed other mechanical schemes.

We asked a convenience sample of five people, who were told the actual outcomes of the conflicts, to rate the outcome options we provided to the research participants. The raters were told that an option that matched the actual outcome of a conflict should be given a rating of 10. Forecasts were counted as accurate if the outcome option chosen by our rule was the option that had been given the highest median rating by our raters. 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."

5. Results

As Tetlock (1999) demonstrated, it is difficult for experts to forecast decisions made in conflict situations. He found that forecasts by 20 experts of the outcomes of foreign-policy conflicts were no more accurate than could be expected from chance. Our results were similar. Our 66 unaided experts were correct for 32% of predictions in an unweighted average across the eight conflicts (Table 2).

As we hypothesized, the forecasts from structured analogies were more accurate. They were more accurate for seven of the eight conflicts. Averaging the accuracy figures across the conflicts, the structured-analogies forecasts were 46% accurate, compared to 28% for chance. Viewed another way, structured analogies reduced the average forecast error by 21% compared to unaided-judgment forecasts (where

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)

Bold figures denote the most accurate forecasts for each conflict or overall.

forecast error is the percentage of forecasts that were wrong).[2]

5.1. Value of experts' experience

We tested whether structured-analogies forecasts were more accurate when they came from experts with more experience than when they came from those with less. We used two measures: (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.

Structured-analogies forecasts from experts with five or more years' experience as conflict management specialists were *less accurate* (on average across conflicts), with 21% error reduction compared to chance, than those with less experience (26% error reduction). Furthermore, where experts gave high ratings to their experience with similar conflicts, their forecasts were less accurate (16% error reduction) than where they gave themselves lower ratings (31%). Our findings suggest that conventional measures of experience are not useful for selecting experts for forecasting using structured analogies. It seems unreasonable to suppose that experience harms forecast accuracy, but this is something that needs further study.[3]

5.2. Effect of number of analogies

We found that forecasts based on data from experts who could think of two or more (plural) analogies were more accurate than those based on data from experts who only recalled a single analogy, for six of the eight conflicts. The accuracy averaged 38% for forecasts derived from single-analogy data, but 56% for those derived from plural-analogy data.

---

[2] We calculate the average error reduction figures as $\{(100-F_C)-(100-F_X)\}/(100-F_C)*100$, where F_C is the unweighted average percentage accuracy across conflicts of the comparison forecasting method (or chance), and F_X is the corresponding figure for the forecasting method of interest.

[3] The initial results from an extension currently being undertaken within the intelligence community found that the forecasts of middle-ranked reserve officers and trainees were less accurate than could have been achieved by choosing a decision at random from the alternatives. We will continue to conduct research on the effect of experience on forecast accuracy, as we find our results on the topic baffling.

Table 3
Accuracy of forecasts by number of analogies

Percent error reduction versus chance (number of forecasts)					
	None [a]		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)
Average error reduction (unweighted)	10	(44)	15	(53)	**39** (44)
Average % correct (unweighted)	34		38		**56**

Bold figures denote the most accurate forecasts for each conflict or overall.

[a] These are 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.

All else being equal, conflicts with more outcome options are more difficult to forecast than those with fewer options. To control for this, we examined the reduction in error versus chance. Forecasts based on the recall of a single analogy reduced the error by an average of 15% compared to chance, while forecasts derived from plural analogies reduced the error by 39% (Table 3). The error was reduced by 42% versus chance by accepting only the 31 responses that included three or more analogies (not shown in Table 3). Thus, the usefulness of an individual expert was related to the number of analogies he described.

5.3. Effect of experts' familiarity with their analogies

We expected that the information experts provided would be more useful, the more closely involved they had been in the analogous situations they identified. For example, someone who was an adult during the Vietnam War is likely to know more about that situation than someone born since, and someone who fought in the war is likely to know more again. To examine this, we identified forecasts that had been based on analogies from either experts' own experiences (45) or that of close others (5 forecasts based on

the experiences of, for example, a wife or brother-in-law). In an unweighted average across the eight conflicts, these direct-experience forecasts were more accurate (49%) than the 45 forecasts based on analogies from third-party accounts (37%). Viewed another way, the forecasts based on analogies from experiences close to experts reduced the average error across conflicts by 31% (compared to chance), while forecasts that were based on indirect experience provided only a 13% error reduction.

5.4. Familiarity and plural analogies

The ideal situation when forecasting with structured analogies is to find experts who can think of many analogies with which they have had direct experience. When our experts were able to think of two or moreanalogies, and they had direct experience of the analogy that was most similar to the target, the structured-analogies forecasts were 60% accurate (23 forecasts). In the other cases, the 72 forecasts were 39% accurate.

5.5. Mechanical schemes to derive forecasts

We wondered whether experts who had used the structured-analogies process then provided forecasts that were more accurate than unaided experts. They did. Their predictions were on average 42% accurate (94 forecasts), compared to 32% for the unaided-judgment forecasts. As we anticipated, however, a structured mechanical process was more effective for deriving forecasts from the experts' analogies information than the experts' own judgments. As we have seen, structured-analogies forecasts were 46% accurate. Why the difference when experts derived their own forecasts? Analogies are only useful if they are used. In 22 cases, experts made forecasts that were inconsistent with the outcomes of their own analogies; of these, 25% were accurate. When the mechanical rule was used to derive forecasts from these experts' analogies, 45% were accurate.

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 in order to determine whether we could improve the accuracy further if we derived combined forecasts from all of the 210 analogies with similarity ratings and implied decisions. For example, if an expert provided information on three analogies, for the purpose of testing our four combining alternatives, we effectively derived three forecasts, instead of the one we would have derived using the structured-analogies method.

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 for the first alternative, for each conflict, we based the forecast on the option with the highest similarity rating (39% accurate). For the third alternative, each expert's analogies were allocated to decision options in proportion to the option's share of the sum of the expert's similarity ratings. The option allocated the most analogies weighted in this way was our forecast for the conflict (40% accurate). The fourth alternative was like the third, except that we weighted each expert's analogies by the average similarity rating for the option as a proportion of his total average similarity ratings (39% accurate). To sum up, all of these alternatives provided forecasts that were less accurate than those derived by applying the mechanical scheme that we had specified prior to testing the accuracy of structured analogies.

5.6. Effect of collaboration

While we had no directional hypothesis about collaboration, we analyzed the data to see whether collaboration among experts was useful. When experts using structured analogies collaborated with others, their median working time was 45 min compared to 30 min 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 using 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.

6. 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. While this may be difficult to know in advance of receiving their structured-analogies analysis, one can gauge people's expertise from that analysis—i.e., how many analogies did they provide and did they have direct experience with those situations? Such an assessment of expertise can be made before knowing whether the forecasts derived from their analogies are accurate.

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.

7. Further research

Our conclusions are based on a sample of only eight situations, and this is the first published study on the use of structured analogies. One should be wary of little-replicated studies, as the findings may turn out to be of limited applicability (Armstrong, 2006). Our results seem extreme to us, and so we would like to see replications and extensions of the research to identify conditions under which structured analogies fail, and the conditions under which the method is most effective. Research using additional situations would also help to better assess how to improve the procedures.

More research needs to be done to develop the operational procedures for the method. 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 the analogies' similarity to a target help administrators derive forecasts that were even more accurate? 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 potential target outcomes?

It seems plausible that the Delphi technique could be used to improve the assessments of analogies' similarities to a target, potentially increasing accuracy further at a low cost. Rowe and Wright (2001) provide evidence on the value of Delphi, and software for the implementation of Delphi is provided at forecasting-principles.com.

We have examined conflict situations because of their importance, and the difficulty of obtaining useful forecasts. Structured analogies might also improve forecasting for situations other than conflicts. We expect that it would be most useful where situations are complex and where there are plural analogies.

Research is needed on how to encourage the 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. To help practitioners and researchers use of the structured analogies method, we have posted a self-learning program on the Education Page at forecastingprinciples.com.

8. Conclusions

It is difficult to forecast what decisions will be made in conflict situations. On average, unaided experts were correct for only 32% of their predictions. This was little better than chance at 28%.

For our structured-analogies method, the two key criteria for identifying an expert were the number of analogies generated, and the presence of direct knowledge about those analogies. When experts produced two or more analogies from experience, the forecasts from structured analogies were correct for 60% of the predictions. Given the importance of forecasts in conflict situations and in other arenas, such an improvement could have considerable benefits.

Acknowledgements

Appendix A. E-mail message appeal and instructions: Structured analogies/collaboration treatment.

Subject: Using analogies to predict the outcomes of conflicts

Dear Dr _____

I am writing to you because you are an expert on _____. I am engaged on a research project on the accuracy of different methods for predicting decisions made in conflicts. At this stage, I'm investigating the formal use of "analogies" for forecasting. That is, forecasting on the basis of the outcomes of similar conflicts that are known to the forecaster.

What I would like you to do is to read the attached descriptions of some real (but disguised) conflict situations and to predict the outcome of each conflict. If you can't read the attachments, please let me know and I'll send the material in your preferred format if I'm able.

Each attached file contains a conflict description and a short questionnaire. Please follow these steps for each conflict:

1/ Read the description and
2/ try to think of several analogous situations and
3/ about how similar your analogies are to the conflict.
4/ Fill-in the questionnaire (electronically if you can)
   a) describe your analogies
   b) rate your analogies
   c) make your prediction (either pick an outcome or assign probabilities)
   d) record the total time you spent on all tasks
   e) return the questionnaire.

One of the objectives of this research is to assess the effect of collaboration on forecast accuracy. You have been allocated to the collaboration treatment, so please *do* discuss these forecasting problems with colleagues. Do *not*, however, discuss them with other people who have received this material as I want independent responses from participants.

Although I intend to acknowledge the help of all of the people who assist with this research, my report will not associate any prediction with any individual.

Your prompt response is very important to the successful completion of my project. Please help me to prove the sceptics wrong about the level of cooperation I get!

Appendix B. Example response.

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)			(B)	(C)
(i) description	(ii) source	(iii) similarities & differences	Rate	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   [✔_]
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 [✔__]
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
either this document as an email attachment to…
or this questionnaire (with your initials at right) by fax to…

Your initials: [_XYZ_]

References

Angelis, L., & Stamelos, I. (2000). A simulation tool for efficient analogy based cost estimation. *Empirical Software Engineering*, *5*, 35−68.

Armstrong, J. S. (1985). *Long-range forecasting.* New York: John Wiley. (Full text at forecastingprinciples.com).

Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 417−439). Boston: Kluwer Academic Publishers. (Full text at forecastingprinciples.com).

Armstrong, J. S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting*, *22*, 583−598.

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

Goldfarb, R. S., Stekler, H. O., & David, J. (2005). Methodological issues in forecasting: Insights from the egregious business forecast errors of late 1930. *Journal of Economic Methodology*, *12*, 517−542.

Green, K. C., & Armstrong, J. S. (2007). Value of expertise for forecasting decisions in conflicts. *Interfaces*, *37*, 287−299.

Heemstra, F. J. (1992). Software cost estimation. *Information and Software Technology*, *34*, 627−639.

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.

Klein, G. A. (1998). *Sources of power: How people make decisions?* Cambridge, MA: MIT Press.

Kokinov, B. (2003). Analogy in decision-making, social interaction, and emergent rationality. *Behavioral and Brain Sciences*, *26*, 167−168 (Full text at http://www.nbu.bg/cogs/personal/kokinov/bbskokinov.pdf).

Langdon, C. (2000a, September 20). Nurses vote today on strike. *The dominion* (pp. 3). Edition 2.

Langdon, C. (2000b, September 21). Nurses support call for strike. *The dominion* (pp. 3). Edition 2.

Langdon, C. (2000c, December 6). Nurses' pay boosted, strike off. *The dominion* (pp. 3). Edition 2.

MacGregor, D. G. (2001). Decomposition for judgmental forecasting and estimation. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 107−123). Boston: Kluwer Academic Publishers.

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

Myers, W. (1989). Allow plenty of time for large-scale software. *IEEE Software*, *6*, 92−99.

Neustadt, R. E., & May, E. R. (1986). *Thinking in time: The uses of history for decision makers.* New York: Free Press.

Radio New Zealand Limited (2000a). September 20. Brenda Wilson (Chief Executive, New Zealand Nurses Organisation) interviewed by Geoff Robinson. Morning Report, Transcript: *Newztel News Agency Ltd*.

Radio New Zealand Limited (2000b). September 20. Rae Lamb (Health Correspondent, *Radio New Zealand*) interviewed by Mary Wilson with excerpted material from interviews with Annette King (Minister of Health), Susan Rolls (Emergency Nurse at Wellington Hospital), and Russell Taylor (Wellington Nurses Union Organiser). Checkpoint, Transcript: *Newztel News Agency Ltd*.

Radio New Zealand Limited (2000c). September 22. Margot Mains (Chief Executive Officer, Capital Coast Health) interviewed by Geoff Robinson. Morning Report, Transcript: *Newztel News Agency Ltd*.

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi technique. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 125−144). 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 May 7, 2004, from http://www.ukans.edu/~keds/papers.dir/Schrodt.APSA02.pdf.

Shepperd, M., & Schofield, C. (1997). Estimating software project effort using analogies. *IEEE Transactions on Software Engineering*, *23*, 736−743.

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.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207−232.