12 Training for sociocultural forecasting: Current status and science and technology gaps[1]

Winston R. Sieck, Global Cognition

1. Introduction

Scientific disciplines across the board have struggled with the challenge of anticipating the future. From early predictions of the positions of planets to modern meteorological forecasting systems, humans have witnessed ongoing improvements in the ability to foresee events. Forecasting within the sociocultural sphere is no exception. Although many scientists are highly skeptical about the ultimate prospects of forecasting phenomena in the area of human social cultural behavior with any quantification or precision, others continue to plod along, making small gains along the way. Research programs that make any attempt to improve forecasting are highly susceptible to being labeled as failures, at least in part because they are plagued from the outset by unrealistic expectations of what they can accomplish in the short term. There is, however, a large difference between only being able to make a measured amount of real progress in the short term, and being unable to achieve anything at all. This chapter describes the current state of the science and technology related to training in sociocultural forecasting, and identifies gaps that research should address to develop capabilities for application in operational settings.

The following incident illustrates the intuitive use of native-level cultural knowledge to anticipate a political leader's actions. It provides several interesting points for discussion, not the least of which concerns how training in sociocultural forecasting techniques could lead to the development of a similar depth of understanding among non-native analysts and operators, at least about relatively narrow forecasting problems like this one (Sieck, McHugh, Klein, Wei, & Klinger, 2004, p. 19):

> The order came down to the head of the Bosnian analysis team on Sunday morning. The General needed the team to identify the propaganda themes that the Croats, Bosniacs, and Serbs would use by Monday. The team came in on a Sunday to put together a briefing.
>
> The team was multinational, with an American, a Dutch, a Turk, a Greek, and a German officer all working together. The American was the team leader. They had been studying the issue and background data for about nine weeks. The team had two separate interpretations and analyses of what was happening and what was driving events. The critical difference had to do with the Serbs, and the possibility that Milosevich would switch from Stalinist communism to pursue a religious angle in the upcoming months.

---

[1] Approved for Public Release; Distribution Unlimited. Case Number 14-2487

> Most of the team members looked at the situation as purely a political process, Milosevich was promising the people who were his natural constituency a better day. They predicted that he would continue to promise a greater Serbia, with the Serbs dominating the political processes and the economic base of whatever was left of Yugoslavia. At dinner, the Greek and the Turkish officers got frustrated with the conversations of the others and finally said, "you don't understand." And the Turkish officer said, "look the problem is you're thinking secular, and they're not." At this point, the reaction of the others was silence. It was one of those conversation killers. The Greek and Turk argued that the political message was no longer powerful enough to rally the Serbs, because there was no passion in it. They said that the only thing passionate in this was the religious separation. The Dutchman said that it might work, though he didn't believe it. He just didn't believe that a Marxist could get away with taking that line, or that he would have any credibility if he tried.
>
> After dinner the leader was approached by the Greek and Turk again. The team leader listened to them because it struck him as very unusual that people from traditional enemy countries came to the same conclusion and worked together to get the word to him. Also, he recognized that they had a very good feel for the culture, much better than the rest of the team. The rest had the distant political-analytical perspective, but they had a much closer, more personal, visceral perspective. The leader included the religious angle for the Serbs. In the end, Milosevich did take the religious angle. He made his first connection to religion in this conflict in May of 1992, several weeks after their prediction.

This case of sociocultural forecasting by a multinational intelligence team illustrates several aspects relevant to sociocultural forecasting:

- Like most forecast problems, it involves the prediction of some change from the status quo.
- The prediction made is categorical rather than probabilistic; the possibility of a religious theme would either be included or not, with no probability attached.
- Several team members seem extremely confident in their predictions, perhaps even overconfident.
- The Greek, Turkish, and Dutch members of the team all had substantial sociocultural knowledge, but that knowledge was of different kinds. The Greek and Turkish members were apparently better able to adopt the "native perspective" on the situation, appreciating what the Serbian population would and would not support.

This last point—about the potential value of understanding how people from a culture of interest think and make decisions before attempting to forecast their future behavior—raises especially interesting questions. This chapter addresses how sociocultural forecasters could be trained to make accurate predictions of culturally different others without the "visceral" understanding of the culture that only natives, and perhaps a few long-term expatriates, would share. It describes selected current science and technology in this area, as well as recommended future efforts to address this question.

The next section provides an overview of literature on probabilistic forecasting, with an emphasis on training analysts how to perform probabilistic forecasting. It then outlines special problems associated with forecasting in the sociocultural domain, along with their training implications. These include the nature and range of problems being forecast, from short-term political actions and events to long-term changes in population cultural values, beliefs, and behaviors. This section also discusses difficulties in representing emergent processes for forecasting, such as those that underlie sociocultural systems, and describes cultural resilience to aid in explaining why cultural change frequently does not happen or often has only temporary effects.

The third section presents an overview of literature and capabilities that support sociocultural forecasting and training. This includes work on concepts for training in cultural sensemaking and their relation to sociocultural forecasting, as well as training in the use of cognitive-cultural models for forecasting sociocultural behavior. In addition, it covers training implications of the Defense Advanced Research Projects Agency (DARPA) and HSCB-funded Worldwide Integrated Crisis Early Warning System (W-ICEWS) program, as well as forecasting training efforts within the Intelligence Advanced Research Projects Activity's (IARPA) Aggregative Contingent Estimation (ACE) program.

The fourth section summarizes science and technology gaps in this area that research has yet to address. The chapter concludes with a description of recommended next steps to move training capabilities toward operational usage.

2. Training Probabilistic Forecasting

Several research efforts have attempted to determine how best to train people to make better probabilistic forecasts. In general, they seek to train analysts in the process by which people generate forecasts, and to determine whether specific types of feedback lead to improved probabilistic forecasts. Such procedures could then be incorporated into forecasting tools and implemented within formalized training and education programs for specialists.

Probabilistic forecasting has the advantage of enabling forecasters to express uncertainty in their opinions using precise language that permits application of various quantitative metrics to assess the accuracy of sets of forecasts (Lehner, Michelson, & Adelman, 2010; Tetlock, 2005). These accuracy measures allow formal evaluation of forecasting systems, and also constitute a primary basis for generating different kinds of feedback to use in training forecasters.

Perhaps the most common measure of the accuracy of probabilistic forecasts is the mean probability score or "Brier score" (Brier, 1950; Yates, 1994). To obtain the Brier score over a set of forecasts, the probability score for each forecast is first calculated as $(f–d)^2$, where f is the reported likelihood that the target event will occur (e.g., the forecaster's stated probability that Iran and the United States will commence official nuclear program talks before 1 October 2013), and d is the actual outcome (coded as 0 if the target event does not occur and 1 if the target event does occur). Probability scores are calculated for each of many forecasts, and these are averaged to obtain the Brier score. As the equation shows, the Brier score is a measure of error in probability forecasts, so the lower the Brier score, the better the forecasts. Researchers have identified various means for decomposing the Brier score into more fine-grained components that offer greater potential for detecting certain kinds of biases, and for formulating training applications based on feedback from accuracy measures (e.g., Yates, 1982).

The two most commonly discussed subcomponents of probability forecast accuracy are calibration and discrimination (see Figure 1). Calibration measures the extent to which a forecaster states probabilities that match various base rates, or the percentages of times that the target event actually occurs. Often, percentages of event occurrence are plotted for each probability judgment to produce a calibration curve (Lehner, Michelson, & Adelman, 2010). A forecaster might exhibit poor calibration in several ways, primarily over-prediction and overconfidence. Over-prediction means that the average of the forecaster's reported probabilities exceeds the actual proportion of times that the target events occurred. For example, suppose an analyst reports probability forecasts for 50 events such as the U.S.-Iran event given above. If the average of the analyst's probabilities is 70%, and yet only 25% of the events actually occur, then the analysts' forecasts are exhibiting over-prediction. Overconfidence typically refers to situations in which the analysts' forecasts are too extreme. To illustrate overconfidence, imagine a two-step procedure in which an analyst first states a categorical prediction as to whether or not each event will occur. The analyst then reports the probability that his or her categorical prediction is correct. If, over a set of 50 problems, the forecaster is correct in 58% of his or her categorical predictions, yet reports an average probability correct of 90%, then extreme overconfidence is evident. The two-step procedure aids in illustrating overconfidence, although the measure and the elicitation method are independent. Overconfidence can be computed regardless of how the probability forecasts are reported.

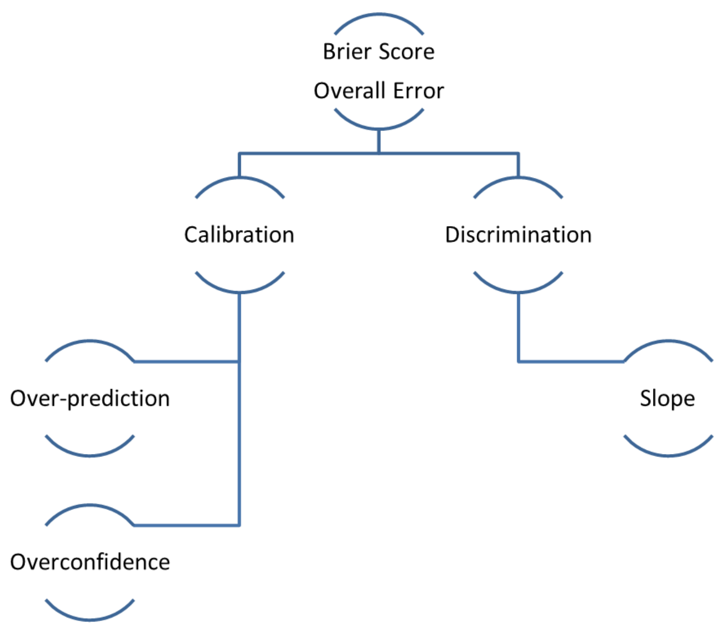

Figure 1. Subcomponents of probability accuracy.

The second general subcomponent of accuracy, discrimination, refers to a forecaster's ability to distinguish between situations when the target event will occur and when it will not occur. A simple measure of discrimination is the "slope" statistic proposed by Yates (1982). Slope is the mean of the probability forecasts for which the events occurred, minus the average of probability forecasts for which the events did not occur. The higher the slope, the better a person's forecasts distinguish between events that occur from those that do not.

It is important to recognize that a forecaster could do well on one of these aspects of probability accuracy but poorly on the other. That is, calibration and discrimination are distinct components of overall forecast accuracy. This has led researchers to examine whether the calibration and discrimination aspects of forecasting accuracy require related or distinct cognitive skills, for example, by attempting to determine whether training procedures that affect one of the measures also produce effects on the second measure (Stone & Opel, 2000).

3. Training to Improve Calibration and Discrimination

Most training in probability forecasting has emphasized improving calibration, presumably because it is generally viewed as simpler to achieve. Perhaps the most intensive training intervention for probability judgment using calibration feedback was conducted by Lichtenstein and Fischhoff (1980). In their study, trainees completed elevent training sessions involving 200 judgment problems. At the completion of each training session, the trainees received feedback on their probability accuracy that included graphs to show their calibration, as well as detailed performance statistics such as the Brier score and numerical measures of calibration. The researchers explained and discussed the feedback with the trainees for five to ten minutes. Following this rather intensive intervention, trainees showed a clear improvement in calibration.

Researchers have delineated several types of feedback potentially relevant to forecast training (Benson & Önkal, 1992). Giving trainees the results of their probability accuracy measures, as in the example above, represents just one kind of feedback, which is sometimes referred to as "performance feedback" (Benson & Önkal, 1992). Several researchers have suggested that performance feedback is a particularly effective method for improving calibration under suitable conditions (e.g., by reducing overconfidence). However, performance feedback has not proven particularly successful at increasing discrimination, since it contains no substantive information that would help trainees to determine whether or not a particular event will occur.

In order to disentangle the issues, researchers have also delineated another type of feedback potentially relevant to forecast training: environmental feedback (Benson & Önkal, 1992). Whereas performance feedback gives trainees information about the accuracy of their forecasts, including components such as measures of overconfidence, over-prediction, or discrimination, environmental feedback provides information about the event to be predicted. For our purposes, environmental feedback refers to any domain-specific information relevant to the forecasting task, such as associations between cues and outcomes. Various forms of cultural information relevant to a leader's decisions, for example, would be especially relevant to predictions in the sociocultural realm. In contrast to performance feedback, environmental feedback appears to improve discrimination but not calibration.

In one study designed to test the relative benefits of distinct kinds of feedback, participants were asked to provide forecasts and given performance feedback. One group of participants received feedback on their calibration, while another group received feedback on their discrimination scores. These both reflect performance feedback, as they relay information as to how well the trainee forecast in the problem domain. As in earlier work, performance feedback regarding calibration improved participants' calibration scores but had no impact on discrimination (Benson & Önkal, 1992).

Subsequently, Stone and Opel (2000) explicitly contrasted the impact of environmental and performance feedback on the accuracy of probability judgments. They randomly assigned participants to performance feedback, environmental feedback, or no feedback (control) conditions. Their study procedure included three parts: pre-training probability judgments, training, and post-training probability judgments. The performance feedback group received individualized feedback regarding their calibration performance. The environmental feedback group received information in the form of a 30-minute lecture on a historical subject designed to increase their substantive knowledge about the topic of their predictions. All participants received a handout that described five to ten important characteristics relevant to the topic at hand. They then practiced using the new knowledge to re-analyze 20 of their probability judgment problems from the pre-training session. The control (no feedback) group completed an unrelated task during the training period.

Stone and Opel confirmed that performance feedback led to improvements in calibration, but not discrimination, and that environmental feedback led to improvements in discrimination, but not calibration. Importantly, the researchers also found that substantive training in the form of providing additional environmental cues and information can actually degrade calibration. In particular, after participants received environmental feedback, their overconfidence increased (Stone & Opel, 2000). The latter finding suggests that merely increasing people's knowledge about a particular area or domain, for instance by encouraging analysts to digest volumes of information about the culture of a region of interest, can lead to an increase in forecast confidence not matched by commensurate increases in accuracy.

Within the domain of sociocultural forecasting, this implies that knowledge of the culture or area is insufficient to accurately anticipate future events. Instead, training programs must also incorporate forecast performance feedback using measures of calibration. This represents an important issue from a practical standpoint, as it indicates that training methods that attempt to influence one component of accuracy (calibration or discrimination) may not have any impact on the other component. Table 1 summarizes these relationships.

Table 1. *Types of Probability Feedback and Their Effects on Performance*

Type of Feedback	Description	Effects on Accuracy
Performance	Information about components of forecast accuracy (calibration, discrimination)	• Improves calibration • No effect on discrimination
Environmental	Information about the event to be predicted, such as cultural information	• Can improve discrimination if information is diagnostic • Can harm calibration by increasing overconfidence

Stone and Opel concluded that calibration and discrimination reflect distinct cognitive skills, which they called calibration expertise and substantive expertise. Substantive expertise refers to domain-specific knowledge in a certain area, such as sociocultural expertise in a given region. Calibration expertise reflects the ability to render probability forecasts that exhibit neither over- or under-prediction nor under or overconfidence. Furthermore, distinct types of feedback and training are required to improve these different cognitive skills and thus promote overall expertise in probabilistic sociocultural forecasting.

4. Forecasting in the Sociocultural Domain

This section describes issues associated with forecasting in the sociocultural domain, as well as their training implications. For illustration, this section employs an example regarding crowd behavior in the Middle East and the Arab Spring.

One issue concerns the scope of problems being forecast, which can range from immediate reactions to medium-term political actions and events to long-term change in population cultural values, beliefs, and behaviors. First, analysts might use sociocultural knowledge and tools to predict local, immediate, transitory outcomes, such as newsworthy events of the day. For instance, Sieck, Simpkins, and Rasmussen (2013) investigated crowd member responses to security force actions in the Middle East. The cultural, behavioral, and situational factors they examined should aid in anticipating whether an Arab crowd would turn violent or remain peaceful in a time span of a few moments to a few hours. Training based on those models would enable U.S. personnel to make sense of, anticipate, and hence more effectively manage crowds in places such as Iraq (Sieck, Smith, Grome, & Rababy, 2010). Such training would also likely have aided an analyst to track events as they unfolded during the Arab Spring to forecast the likelihood of violent eruptions on the same time scale.

At a second level, sociocultural sensemaking might be used to predict outcomes in the middle term, such as changes in political leadership. An example at this level would be an analyst forecast of the likelihood that Hosni Mubarak would not remain as Egypt's leader as a result of the Arab Spring protests.

Finally, a third level concerns distant, long-term outcomes, such as changes in deep-rooted cultural values. A forecast problem at this level might address the likelihood that the Arab Spring and changes in political institutions within Egypt would reduce the cultural value of "power-distance" over the next decade. Power-distance measures the degree to which members of the culture accept and expect unequal distribution of power in the society (Hofstede, 2001). People in high power-distance cultures tend to be comfortable with the idea that their leaders (at the national, organizational, or family level) hold the lion's share of power. Middle Eastern countries tend to rank very high on Hofstede's power-distance index (Hofstede, 2001). For instance, if the cultural value placed on power-distance remains constant within Egypt, new governments and leadership would probably assume just as much control as their predecessors, since this is expected at a deep level by their constituents.

To date, most research and application of sociocultural forecasting has apparently focused primarily on the middle level. However, outcomes at this level may well not mean what we think they mean in terms of longer term, strategic-level implications. Sociocultural forecasting and persistent sensing of fundamental cultural values that represent long-term changes to societies constitute an important area for further applied research and development.

Consideration of long-term, fundamental cultural values leads naturally to a second issue in sociocultural forecasting: the need to understand why cultural change frequently does not happen or often has only temporary effects. An important factor associated with a culture's resistance to change is the looseness or tightness with which its members hold on to cultural norms: the standards by which members of the culture live. These norms represent the shared expectations and rules that guide the behavior of people within social groups, and as such are learned from and reinforced by parents, friends, teachers, and others as children grow up in a society.

Gelfand and a large international team investigated cross-cultural differences in the extent to which cultural norms matter to the members of the society (Gelfand et al., 2011). Some societies exhibit "cultural tightness," insisting on strong conformity to their cultural norms in all areas. Other, "culturally loose" societies tolerate far more deviance from the norms. Gelfand and colleagues theorized that tightness and looseness are reflected at different levels within a culture that mutually support one another. Specifically, Gelfand et al. (2011) described evidence related to four levels:

1. Ecological and Historical Threats. Hostile neighbors, disease, and dense populations increase the need for coordinated and disciplined action from the population. Factors such as these tighten cultural norms. As the threats diminish, adherence to norms becomes looser.

2. Sociopolitical Institutions. Culturally tight nations tend to have more autocratic governments, restricted media, stronger suppression of dissent, and more severe punishments for crime.
3. Everyday Social Situations. All kinds of interactions with fellow members of the culture are more formal in nations with tight cultural norms. These include situations at home, the workplace, school, places of worship, parks, and others. Loose cultures provide more room for individual discretion in such situations; a wider range of behavior qualifies as "appropriate."
4. Psychological Adaptations. People's minds become attuned to the different requirements of living in places with tight or loose cultural norms. Individual psychology then further supports the level of cultural tightness or looseness. People living in tight cultures become more focused on avoiding mistakes, are more cautious in their own behavior, and more closely monitor themselves and others for norm violations.

A sociocultural forecaster might take many cultural values and factors into account when generating a prediction. Cultural tightness or looseness of norms represents one of a small set of potential factors that specifically address tolerance for deviance and hence willingness to change at the cultural level. It is therefore a potentially useful factor to consider across many forecast problems. In addition, the Gelfand team's analysis of the issue in terms of the four mutually supporting levels provides an excellent illustration of the general factors involved in maintaining any cultural system in its current state.

5. Training in Sociocultural Forecasting

Several programs have addressed topics related to training in sociocultural sensemaking and forecasting, including the Office of the Secretary of Defense's Human Social Culture Behavior (HSCB) Modeling Program, IARPA's ACE program, and DARPA's W-ICEWS program. This section discusses selected research and development topics from each of these programs.

5.1. HSCB modeling program

Cognitive-Cultural Modeling

The HSCB program has sponsored the use of cognitive-cultural models that can aid in forecasting sociocultural behavior. Cognitive-cultural models are graphical representations of a culture's shared values, conceptions, and causal beliefs that influence decisions by members of interest in that culture (Sieck, 2010; Sieck, Rasmussen, & Smart, 2010). These models help cultural outsiders to assume the viewpoint of cultural insiders. Cognitive-cultural models also aid in identifying cultural elements that should receive high priority in training, and in anticipating the behavior of the members of the culture. In their most advanced form, cognitive-cultural models also represent quantitative information about the prevalence of the ideas included.

Researchers developed the models via Cognitive-Cultural Analysis (CCA). This process was originally termed "Cultural Network Analysis," but the name was changed to highlight the cognitive emphasis of the models, i.e., that the models seek to reveal how members of the culture think and make decisions. The CCA approach adopts the theory of "cultural epidemiology," which implies that ideas can be studied using some of the same techniques that epidemiologists use to study diseases (Berger & Heath, 2005; Sperber, 1985).

CCA ensures that the cognitive aspects of a culture can be incorporated into practical use, such as forecasting and training applications, by identifying key decisions or judgments of interest as a first step. Once the key decisions have been determined, cultural analysts construct models which include the cultural ideas that directly influence those decisions.

CCA encompasses several techniques needed to build cognitive-cultural model diagrams. The primary representation format for CCA is the influence diagram, used very successfully for some time to map knowledge in decision analysis (Howard, 1989). Research has developed techniques to (a) elicit concepts, causal beliefs, and values from people in interviews or survey instruments, (b) extract the ideas from interview transcripts or other text material (e.g., text harvested from the World-Wide Web), (c) analyze the degree of commonality in ideas between cultural groups, (d) align and assemble consensus groups of ideas into maps, and (e) relate them to demographic variables. Although CCA seeks to represent the idea networks in a common, scientific format, it nonetheless maintains the content of cultural knowledge as expressed by members of the cultural group (Sieck, 2010). This is exactly the information that analysts need to acquire the visceral understanding described in the illustrative case at the beginning of this chapter.

The resulting cognitive-cultural models include estimates of the prevalence of such cultural ideas. Capturing the proportions of people who actually maintain the relevant beliefs provides a full description of the current cognitive state of a culture. With respect to prediction, prevalence information yields relevant status quo base-rates to which forecasters can anchor and then adjust to ensure forecast realism. In this way, cognitive-cultural models that include quantitative estimates of idea prevalence give forecasters environmental feedback in a form that has the potential to improve both aspects of probability accuracy: discrimination and calibration.

CCA is ideally suited to simulations using Bayesian modeling, although the application is somewhat different from Bayesian modeling applications that rely on expert inputs. In standard expert applications of Bayesian belief networks (BNs), researchers elicit structural and probability inputs from experts to determine the representation of a physical system, with the aim of accurately and quantitatively predicting key physical outcomes based on the experts' understanding of influences within the system. By contrast, in a cognitive-cultural modeling application the aim is to capture cultural knowledge in order to accurately anticipate the perceptions and decisions of members of cultural groups. The key principles governing application of BNs to cultural modeling are that BNs focus on the population, rather than on the individual psychological level of analysis; nodes represent concepts held in common by some percentage of the population; edges represent causal beliefs, also distributed across members of the population; and probabilities denote the prevalence of ideas in the population, not the strength of belief.

Projects sponsored by the HSCB Program have used CCA to develop several quantitative cognitive-cultural models of Afghan decision making based on interviews and a survey study of 400 Afghans from several different provinces (Sieck, Javidan, Osland, & Rasmussen, 2012). CCA and one of the

Afghan cognitive-cultural models were also used in the development of a prototype tool for simulating cultural behavior, known as the Cultural Belief Network Simulation Tool (CulBN) (Sieck, Simpkins, & Rasmussen, 2011). CulBN provides a user interface that enables the creation, visualization, and manipulation of input values in a cognitive-cultural model. The interface was combined with a standard BN simulation engine that could generate forecasts contingent on specific changes within the model (Sieck, Simpkins, & Rasmussen, 2011). The system allowed forecasters to interact with the model, for instance by entering hypothesized changes in certain elements and then visualizing the predicted consequences in terms of quantitative adjustments to cultural prevalence values. Some of the primary anticipated benefits of CulBN in supporting the training of forecasters include that it provides a coherent approach to constraining and making sense of a vast array of concepts, causal beliefs, and values, as they are relevant to particular decisions and behaviors of interest to the forecaster.

One of the challenges for novice sociocultural forecasters is managing the overload of information potentially relevant to a forecast. CulBN's structure may help reduce the amount and increase the relevance of information that analysts must take into consideration, which should support both forecast calibration and discrimination. In addition, experienced analysts can usefully share cognitive-cultural models to provide environmental feedback that can aid novice analysts in gaining expertise on specific topics relevant to their area of study. These possibilities should be tested in applied research settings.

Training in Cultural Sensemaking

Training in cultural sensemaking provides learners with cultural knowledge relevant to explaining culture-specific behavior, as well as metacognitive strategies for coping with unexpected behaviors and consequently acquiring new knowledge (Rasmussen & Sieck, 2012; Rasmussen, Sieck, & Osland, 2010). Such training may use cognitive-cultural models to provide direct and specific input to support novice thinking. Rasmussen et al. (2010) outlined a theoretical framework for cultural sensemaking that connects metacognitive skills to region-specific knowledge. They also described a novel approach to instructional analysis and design, specifically developed to identify learning objectives and content for training in cultural sensemaking. Demonstration of the method led to the development of a training booklet and materials that some military-cultural instructors have incorporated into their courses (Rasmussen & Sieck, 2010).

Within this theoretical framework, cultural sensemaking refers to the processes by which people make sense of and explain culturally different behaviors (Osland & Bird, 2000). In such cases, it is natural that peoples' initial perspectives are driven by expectations based on normal behavior learned within their own culture (Archer, 1986). An initial challenge within a cultural sensemaking situation or training simulation is recognizing when the models one would normally use no longer apply. Next, individuals seek the information they need in order to develop culture-appropriate understanding of the current situation. This provides a basis for projecting likely subsequent actions and, in the present context, for developing informed probabilistic forecasts.

An important metacognitive skill to support cultural sensemaking is the ability to build the knowledge required to explain and predict behavior. Further, this overall skill is embedded within a framework of related metacognitive skills that allow individuals to obtain, apply, test, and refine their cultural knowledge. These metacognitive skills are culture general in the sense they support attainment of culture-specific knowledge within any culture (Rasmussen & Sieck, 2012).

Instructional analysis for training in cultural sensemaking begins with Cognitive Task Analysis methods to identify culturally relevant situations for inclusion in a scenario-based training program (Chipman, Schraagen, & Shalin, 2000). The specific nature of the resulting scenarios would depend on the region of interest, as well as on the trainee's job (e.g., small unit leader vs. sociocultural analyst). Curriculum developers then use CCA to characterize both native decision making within these challenging situations and learner expectations regarding the native decisions. Specific knowledge-level learning objectives then result from comparing the learner models and native models to identify the gaps in learners' understanding and their misconceptions regarding how natives are likely to decide.

For example, in the application to Afghan decision making, the native cognitive-cultural model provided the target concepts for training U.S. Marines to understand Afghan behavior. New Marines with no prior deployments made open-ended assessments that were used to generate their expected models of the Afghan decisions and motivations. The program developers used a coding scheme to perform a quantitative assessment of the accuracy of these target learners' understanding of the cultural model and identify critical belief-value relationships that they either failed to perceive or had misunderstood. The knowledge-level (or cognitive-level) learning objectives for the training program focused on closing the gaps and remedying the misconceptions revealed through this assessment.

The cognitive learning objectives in the example above are culture- and job-specific, as is appropriate for trainees who must focus on a specific region for a year or more. Training in cultural sensemaking also includes culture-general learning or metacognitive strategies that build sensemaking competence. For example, in the Afghan-based demonstration project, developers identified learning objectives to improve information-seeking strategies by comparing the questions that the target learners asked to better understand the situation to the questions that expert cultural sensemakers tend to ask in order to create deep understanding (Sieck, Smith, & Rasmussen, 2008). Experts ask cultural sensemaking questions to obtain deeper insight into the belief-value relationships relevant to explaining and anticipating behavior. Generally, such questions take the form of why, why not, how, what if, etc. (Graesser, Baggett, & Williams, 1996).

The framework for training in cultural sensemaking describes a process through which culture-specific learning can contribute to culture-general competence. The underlying concept is to provide learners with baseline cognitive-level knowledge of the factors that influence the decision making of culturally different others in specific situations, along with the metacognitive strategies that enable the learners to progress further by building on that initial, informed understanding. This should enable trainees to learn more efficiently from the complex, real-life cultural situations they will encounter on the job, and further expand their storehouse of experiences. Applied research has tested the principles of this training approach, but future work is needed to develop a full training system based on the approach and test its efficacy at supporting sociocultural forecasting.

5.2. DoD's W-ICEWS program

DoD's W-ICEWS program focused primarily on developing and testing computational models to anticipate and understand instability and violent political conflict (Kettler & Hoffman, 2012; O'Brien, 2010). The developers envisioned a system that could provide military commanders with predictions as to which countries would most likely experience domestic and international crises, with forecasts ranging from the short to the long term. Teams in the W-ICEWS program developed competing models to predict historical cases of instability using data from news and other country background information.

DARPA designed W-ICEWS to enable the comparison and evaluation of several different forecasting approaches. A team led by Lockheed Martin-Advanced Technology Laboratories (LM-ATL) developed the system that generated the most accurate forecasts, making correct predictions about 80 percent of the time in Phase 1, and improving to around 95 percent accuracy in Phases 2 and 3 (Kettler & Hoffman, 2012). LM-ATL used Bayesian methods to produce forecasts by integrating a few distinct kinds of modeling systems, such as agent-based models, logistic regression models, and geospatial network models. The models took a range of factors into account, including sociocultural information such as ethnic-political identities, social similarity profiles, authority structures, trade ties, flow of people, and geographic organization. Interestingly, researchers found that the system predicted few cases of civil war at probability levels above 50%, and some of the models apparently never generated probabilities greater than 30% (O'Brien, 2010).

To determine the possible training implications of this research and development effort, consider that models of this type would likely serve as inputs to an analyst or planner responsible for informing a command team about sociocultural events of interest. In such cases, the analyst should have a conceptual understanding of the models and the process by which they arrive at their predictions. Understanding the concepts underlying model aggregation, and especially the reasons why aggregated, hybrid models often outperform others would also be extremely useful. This level of understanding is essential to ensure that operators trust and incorporate the forecasting system predictions into their briefings and recommendations (cf. Wedgwood, Ruvinsky, & Siedlecki, 2012).

Beyond this, operators need to understand the range of predictions produced by the systems: specifically, how to interpret low probabilities and consider base-rate information (overall proportions of times that events occur). A sizeable body of research shows that superior forecast performance results from giving substantial weight to base-rates, and that computational models using formalized estimation make optimal use of the base-rate information. Models that produce only slight adjustments from base-rates do so because the specific factors they rely on are only weakly informative as to the events being predicted. This occurs because few organizations currently have access to sociocultural information that discriminates much beyond base-rate predictions. Analysts and operators must learn to appreciate that the specific sociocultural information they have on hand is most likely not as discriminating as it intuitively appears.

5.3. IARPA's ACE Program

IARPA's ACE program on contingent forecasting of international actor reactions to possible U.S. and third-party actions has also undertaken efforts to support training in applied forecasting. According to IARPA (2013), the project uses crowdsourcing techniques to "dramatically enhance the accuracy, precision, and timeliness of forecasts for a broad range of event types, through the development of advanced techniques that elicit, weight, and combine the judgments of many intelligence analysts." An example forecast problem is: "Will Iran and the United States commence official nuclear program talks before 1 April 2013?" The resulting forecasts span a global domain, from the outcome of presidential elections in Taiwan to the potential of a downgrade of Greek sovereign debt, which presents interesting challenges in terms of bringing relevant regional and cultural knowledge to bear. IARPA evaluates ACE predictions using the widely accepted Brier score, or average sum of squared differences between probability forecasts and the actual outcomes, as described previously (Brier, 1950).

With respect to training, one project site employs a "Forecasting Ace University" that provides detailed information on specific aspects of forecasting, along with a set of forecast training modules (Warnaar et al., 2012). The training topics include the scoring rules used by the site, a description of high-stakes forecasting, calibration of forecasts, the use of base-rates to alleviate over-prediction, and ways of evaluating the credibility of information sources to promote better discrimination. The system delivers computer-based training (CBT) modules on several topics over the web (i.e., e-learning). Each module includes multimedia presentations and interactive exercises along with explanation feedback to aid concept acquisition and application. Site users also receive information and feedback relevant to specific forecasts, such as information about others' forecasts and rationales (environmental feedback), trends in crowd forecasts over time, and information about their own forecast accuracy (performance feedback).

Another team has also sought to determine how best to train and support forecasters, and has reported demonstrated gains due to training (Ungar, Mellers, Satopää, Tetlock, & Baron, 2012). This team initially expressed some skepticism about the possible benefits of allowing individual forecasters to share information (a form of environmental feedback) or to receive training in the use of probabilities, including performance types of feedback. For example, they noted that permitting experts to share information about their predictions and reasoning had the same potential of leading to "group-think" as of leading to better models of the phenomenon. On the probability training side, they noted that conceptual knowledge of probability and statistics does not disable forecasters' natural intuitions and use of flawed heuristics when making predictions. They also pointed out that systematic cognitive biases, such as overconfidence, might be more readily corrected using mathematical transformations during the aggregation of forecasts across members of the "crowd."

Despite their initial pessimism, the team experimentally tested the effects in a controlled fashion. Conditions within the experiment included probability training and no training, as well as forecasters who worked independently and those who were assigned to teams of 15 to 20 members who consulted with one another over each forecast (the full design included several other conditions as well). Their primary findings included that the team condition, which enables environmental feedback in the form of sharing region-specific and other information related to each forecast, significantly outperformed the independent forecaster conditions. In addition, the researchers found probability training was beneficial, and indeed augmented the benefits of the team condition, such that the probability training proved most effective for people who made forecasts within team environments.

Another interesting aspect of W-ICEWS and ACE is that both directly embed training as a core component of the forecasting system. Researchers should explore additional development approaches along the lines of integrating forecast systems and forecast training systems.

6. Advancing Training in Sociocultural Forecasting

The efforts surveyed above reveal several science and technology gaps that researchers should address to move this capability area toward operational usage. This particular sub-area of sociocultural sensemaking is perhaps less developed than others, so the following discussion primarily addresses topics that applied research should resolve, such as potential skills, knowledge, and abilities that should be tested as training requirements. However, this section also contains some suggestions as to operational training system requirements.

As noted previously, studies on training for probabilistic forecasting indicate that knowledge of the culture or area does not by itself enable analysts to accurately anticipate future events; performance feedback using probability accuracy measures, such as calibration, is also needed. Applied research with professionals in the sociocultural domain should test this proposition. If it holds, then training systems that improve calibration of sociocultural forecasts could potentially lead to immediate gains in both the forecasting process and forecast accuracy among these experts.

Fundamental cultural values that underlie long-term changes to societies should receive greater attention in applied research and development, as they influence current events and future prospects within areas of interest. Along these lines, applied research should test the potential value of using a small set of sociocultural factors directly related to cultural change as general predictors for problems in the sociocultural sphere. Research should also test the hypothesis that analysts could be readily trained to recognize these factors, and that this would aid in preventing over-prediction and enhancing calibration of sociocultural forecasts. In addition, research should test the hypothesis that training to consider this specialized set of factors can improve understanding of why cultural change frequently does not happen or produces only temporary effects. Similarly, researchers should direct more efforts toward fine-grained analyses to understand the comparative predictive value of specific sociocultural factors typically used in current sociocultural models, and to determine new measures that yield more discriminative information. Uncovering new measures will likely be best served by precise and persistent sensing of regions of interest.

Researchers should examine the potential utility of training analysts to routinely and intuitively represent sociocultural forecast problems at the four distinct levels of analysis (ecological-historical, sociopolitical institutions, everyday social situations, and psychological adaptations) identified by Gelfand et al. (2011). Such research would test the important hypothesis that intuitive understanding of the interactions among these levels, which mutually support the status quo, can improve sociocultural forecasting accuracy, especially in terms of calibration.

Research should seek to verify whether cognitive-cultural models actually provide a framework that supports forecasting by enabling analysts to reference a cultural-insider point of view, and explore several additional hypotheses associated with the potential benefits of such models. For example, one hypothesized benefit is that the form of the model encourages thinking about specific elements of information pertinent to forecasts, and hence aids novice analysts to focus their efforts on acquiring specific cultural information relevant to their forecast problems. Another is that inclusion of quantitative estimates of cultural-idea prevalence could improve both discrimination and calibration. Applied research efforts are needed to test this claim, as well as to examine the training required to enable use of cognitive-cultural models. Such training could allow analysts to comprehend these models, as well as to create them directly from the source materials they are working with, as a way to take a cultural other's perspective.

A fair amount of applied research has sought to validate the principles underlying the approach to training analysts in cultural sensemaking, and has shown promising results. Future work should aim to develop a full CBT system based on the approach and should test its efficacy at supporting sociocultural forecasting. A lightweight e-learning system that relies on standard multimedia presentation capabilities combined with interactive exercises would likely offer a reasonable near-term solution. However, systems that incorporate interactive tutors using artificial intelligence may better identify and adapt to the wide variety of possible background knowledge and attitudes that learners bring to the cultural training environment (Woolf, 2009).

Research programs should explore training that might improve operators' conceptual understanding of various formal sociocultural models (e.g., logistic regression, agent-based models, hybrid models) and the processes by which they arrive at their predictions. Understanding the concepts underlying these models and the nature of the benefits of aggregation across models would help improve operator trust and adoption of the forecasting systems. A related line of applied research would investigate training to understand the current limits of sociocultural prediction.

Existing training efforts provide encouraging support to the idea that probability-oriented training improves specific types of sociocultural forecasts. The findings should be extended to other types of sociocultural forecasting. In addition, some programs have succeeded in integrating forecast training with forecasting support systems; further development efforts should capitalize on this initial success.

As stated at the outset, all areas of science struggle to meet the challenge of predicting events. Research and development efforts that address sociocultural sensemaking, particularly in the areas of modeling and simulation, have made important advances in addressing this challenge in the cognitive science of culture. Future efforts to fill the gaps and test the hypotheses described here will further enhance the capability to train forecasters working in this domain, and can thereby improve predictions in operational settings.

References

Archer, C. M. (1986). Culture bump and beyond. In J. M. Valdes (Ed.), *Culture bound: Bridging the cultural gap in language teaching* (pp. 170-178). Cambridge, UK: Cambridge University Press.

Benson, P. G., & Önkal, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting, 8*(4), 559-573.

Berger, J. A., & Heath, C. (2005). Idea habitats: How the prevalence of environmental cues influences the success of ideas. *Cognitive Science, 29*, 195-221.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1-3.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman & V. L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 3-23). Mahwah, NJ: Lawrence Erlbaum Associates.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., . . . Arnadottir, J. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*(6033), 1100-1104.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10*, S17-S31.

Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Thousand Oaks, CA: Sage.

Howard, R. A. (1989). Knowledge maps. *Management Science, 35*, 903-922.

Intelligence Advanced Research Projects Activity (2013). *Aggregative Contingent Estimation (ACE)*. Retrieved from http://www.iarpa.gov/Programs/ia/ACE/ace.html

Kettler, B., & Hoffman, M. (2012, July). *Lessons learned in instability modeling, forecasting and mitigation from the DARPA Integrated Crisis Early Warning System (ICEWS) program*. Paper presented at the 2nd International Conference on Cross-Cultural Decision Making: Focus 2012, San Francisco, CA.

Lehner, P., Michelson, A., & Adelman, L. (2010). *Measuring the forecast accuracy of intelligence products* (pp. 1-13). MITRE Technical Report (MTR 104625). Retrieved from https://www.mitre.org/sites/default/files/pdf/10_4625.pdf

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance, 26*, 149-171.

O'Brien, S. P. (2010). Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review, 12*(1), 87-104.

Osland, J. S., & Bird, A. (2000). Beyond sophisticated stereotyping: Cultural sensemaking in context. *Academy of Management Executive, 14*(1), 65-79.

Rasmussen, L. J., & Sieck, W. R. (2010). *What happens after the 3rd cup of tea? A cultural sensemaking guide to Afghanistan*. Washington, DC: GPO.

Rasmussen, L. J., & Sieck, W. R. (2012). Seven mental habits of highly effective warrior diplomats: Strategies for developing and practicing cross-cultural expertise in the military. *Military Review, March-April*, 71-80.

Rasmussen, L. J., Sieck, W. R., & Osland, J. (2010). Using cultural models of decision making to develop and assess cultural sensemaking competence. *Advances in Cross-Cultural Decision Making.* Boca Raton, FL: CRC Press.

Sieck, W. R. (2010). Cultural network analysis: Method and application. In D. Schmorrow & D. Nicholson (Eds.), *Advances in Cross-Cultural Decision Making* (pp. 260-269). Boca Raton, FL: CRC Press.

Sieck, W. R., McHugh, A. P., Klein, G., Wei, S., & Klinger, D. W. (2004). *Uncertainty management for teams: The strategy of developing shared understanding in the face of uncertainty.* (Technical Report N00014-04-M-0148). Fairborn, OH: Klein Associates.

Sieck, W. R., Javidan, M., Osland, J., & Rasmussen, L. J. (2012, July). *Characterizing cultural construals of behavior: A methodology and application to Afghan honor and integrity*. Paper presented at the International Association for Cross-Cultural Psychology (IACCP) 21st International Congress, Stellenbosch, South Africa.

Sieck, W. R., Rasmussen, L. J., & Smart, P. (2010). Cultural network analysis: A cognitive approach to cultural modeling. In D. Verma (Ed.), *Network Science for Military Coalition Operations: Information Extraction and Interaction* (pp. 237-255). Hershey, PA: IGI Global.

Sieck, W. R., Simpkins, B., & Rasmussen, L. J. (2011). A cultural belief network simulator. *Social Computing, Behavioral-Cultural Modeling and Prediction*, *6589*, 284-291.

Sieck, W. R., Smith, J., Grome, A. P., Veinott, E. S., & Mueller, S. T. (2013). Violent and peaceful crowd reactions in the Middle East: Cultural experiences and expectations. *Behavioral Sciences of Terrorism and Political Aggression, 5*(1), 20-44.

Sieck, W. R., Smith, J., & Rasmussen, L. J. (2008, December). *Expertise in making sense of cultural surprises.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Sieck, W. R., Smith, J. L., Grome, A. P., & Rababy, D. A. (2010). Expert cultural sensemaking in the management of Middle Eastern crowds. In K. L. Mosier & U. M. Fischer (Eds.), *Informed by Knowledge: Expert Performance in Complex Situations*. Boca Raton, FL: Taylor and Francis.

Sperber, D. (1985). Anthropology and psychology: Towards an epidemiology of representations. *Man, 20*, 73-89.

Stone, E. R., & Opel, R. B. (2000). Training to improve calibration and discrimination: The effects of performance and environmental feedback. *Organizational Behavior and Human Decision Processes, 83*(2), 282-309.

Tetlock, P. E. (2005). *Expert political judgment*. Princeton, NJ: Princeton University Press.

Ungar, L., Mellers, B., Satopää, V., Tetlock, P., & Baron, J. (2012, November). *The good judgment project: A large scale test of different methods of combining expert predictions*. Paper presented at the 2012 AAAI Fall Symposium Series, Arlington, VA.

Warnaar, D. B., Merkle, E. C., Steyvers, M., Wallsten, T. S., Stone, E. R., Budescu, D. V., . . . Argenta, C. F. (2012, November). *The aggregative contingent estimation system: Selecting, rewarding, and training experts in a wisdom of crowds approach to forecasting.* Paper presented at the 2012 AAAI Spring Symposium Series, Arlington, VA.

Wedgwood, J. E., Ruvinsky, A., & Siedlecki, T. (2012, July). *What lies beneath: Forecast transparency to foster understanding and trust in forecast models*. Paper presented at the 2nd International Conference on Cross-Cultural Decision Making: Focus 2012, San Francisco, CA.

Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing E-learning*. Burlington, MA: Morgan Kaufmann.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Decision Processes, 30*, 132-156.

Yates, J. F. (1994). Subjective probability accuracy analysis. In G. W. P. Ayton (Ed.), *Subjective Probability* (pp. 381-410). New York, NY: Wiley.