

FORECASTER'S FORUM

Weather Forecasting by Humans—Heuristics and Decision Making

CHARLES A. DOSWELL III

CIMMS, University of Oklahoma, Norman, Oklahoma

31 December 2003 and 18 June 2004

ABSTRACT

The decision-making literature contains considerable information about how humans approach tasks involving uncertainty using heuristics. Although there is some reason to believe that weather forecasters are not identical in all respects to the typical subjects used in judgment and decision-making studies, there also is evidence that weather forecasters are not so different that the existing understanding of human cognition as it relates to making decisions is entirely inapplicable to weather forecasters. Accordingly, some aspects of cognition and decision making are reviewed and considered in terms of how they apply to human weather forecasters, including biases introduced by heuristics. Considerable insight into human forecasting could be gained by applying available studies of the cognitive psychology of decision making. What few studies exist that have used weather forecasters as subjects suggest that further work might well be productive in terms of helping to guide the improvement of weather forecasts by humans. It is concluded that a multidisciplinary approach, involving disciplines outside of meteorology, needs to be developed and supported if there is to be a future role for humans in forecasting the weather.

1. Introduction

Weather forecasting as practiced by humans is an example of having to make judgments in the presence of uncertainty. For its human practitioners, forecasting the weather becomes a task for which the details can be uniquely personal, although most human forecasters surely use approaches based on the science of meteorology in common to deal with the challenges of the task. In what follows, it is assumed that no instance of the weather is ever exactly identical to another—the weather, in detail, never repeats itself. This assumption is likely to be unprovable, but in the absence of a counterexample, it seems plausible.

Deterministic forecasts typically include no explicit statement about the inevitable uncertainty (Lorenz 1963) associated with them. They can take the form of binary (yes/no) products that are valid within a certain area or at a point, during some forecast valid time. Deterministic forecasts also can be graphical, essentially specifying forecast variables in space and time. Numerical models generally produce fields of forecast variables on a spatial grid at specific times. Postprocessing of model output [e.g., model output statistics (MOS)] can be used to convert deterministic model output to objective probabilistic forecasts and to make forecasts for events that are not explicitly described by the model variables.

On the other hand, probabilistic forecasts account for the inherent uncertainty associated with forecasts by formulating the forecasts in terms of probabilities. This can take the form of the probability of some discrete event occurring within the space–time volume for which the forecast is valid, or can involve the probability of exceeding some threshold during the valid time of the forecast.

For categorical, or *dichotomous* (binary or yes/no) forecasts and events, the correspondence between the forecasts and the events over some collection of forecasts takes the form of the well-known 2×2 contingency table (Table 1). In the case where the forecasts are *polychotomous* with M (>2) forecast classes, which is generally the case for probabilistic forecasting,¹ the table becomes $M \times 2$ when the events are dichotomous and $M \times K$ when the events are polychotomous, with K (>2) event classes.

There are numerous sources for detailed explanations of the methods of forecast validation in either categorical or probabilistic terms (e.g., Panofsky and Brier 1958; Murphy and Winkler 1987; Wilks 1995; Joliffe and Stephenson 2003). It is perhaps less well recognized that the observations themselves are uncertain and so it is also possible to express the observations in a prob-

Corresponding author address: Dr. Charles A. Doswell III, CIMMS, 100 E. Boyd St., Room 1110, University of Oklahoma, Norman, OK 73019.
E-mail: cdoswell@hoh.gcn.ou.edu

¹ Categorical forecasting can be thought of as a limiting case of probabilistic forecasting, where the only forecast categories are 100% (yes) or 0% (no).

TABLE 1. Standard 2×2 contingency table for verifying dichotomous forecasts with dichotomous events.

Forecast/ observed	Yes	No	Sum
Yes	x	y	$x + y$
No	z	w	$z + w$
Sum	$x + z$	$y + w$	N

abilistic format (e.g., Wikle and Anderson 2003); this topic is outside the scope of this essay, however.

Objective forecasting—either categorical or probabilistic—is presently being studied extensively, and a wide range of methodologies is being explored in research and in operational forecast verification: numerical weather prediction (NWP), MOS, ensemble methods, “perfect prog” schemes, expert systems (often based on neural network programs), “fuzzy logic” schemes, and so on. Objective methods are generally described as a *guidance* to human forecasters, although it can be imagined that a plausible goal of objective forecast method developers is to equal or even exceed the accuracy levels of human forecasters. If such a goal can be achieved over much of the range of human forecasting activities, the implications for the future of weather forecasting by humans would be manifestly evident. Descriptions of many objective forecasting methods can be found in Wilks (1995).

The issue of how *human* forecasters use the information at their disposal to make forecasts, however, has not been studied comprehensively, and there is comparatively little such work in the meteorological literature (for a recent informal publication on the subject, see Hahn et al. 2002, also available online at <http://www.wdtb.noaa.gov/resources/projects/CTA/Final123102rev030108.doc>). Therefore, the forecasting community has only limited information about how humans forecast the weather. Nevertheless, there are extensive works dealing with the issue of judgment under uncertainty (e.g., Kahneman et al. 1982; Hammond 1996) in general, with a few studies relating specifically to weather forecasting (e.g., Stewart et al. 1992, 1997). In these works, a distinction is drawn between objective information, often derived by statistical methods,² and *heuristics*. The latter is a general term for the subjective methods used by humans for making judgments in the presence of uncertainty. In considering what is currently known about heuristics from the decision-making literature, it is possible to make some inferences about how weather forecasting heuristics work. These inferences form the substance of this essay, although the applicability of results from judgment and decision-making studies in general to the specific task of weather forecasting is not

² It is worth noting that so-called objective methods usually involve many *subjective* choices when they are actually implemented; the notion of objectivity as used in this context is associated with the property that objective methods always produce the same output for any given input.

generally known. In a technological environment where automation of many tasks formerly done by humans is widespread, the issue of the extent to which forecasting can be taken over by objective, automated systems is currently of considerable concern, especially to human forecasters.

2. Sources of weather forecast uncertainty

The sources of uncertainty in weather forecasting can be divided into two distinct groupings. As in the exploration of ensemble forecasting (e.g., Buizza and Palmer 1998; Buizza et al. 1999; Stensrud et al. 1999), these can be referred to as *initial condition* uncertainty and *model* uncertainty. For most forms of forecasting, including both objective methods and those practiced by humans, both of these sources of uncertainty are present, although which source dominates can vary from one meteorological situation to another.

a. Initial condition uncertainty

A weather forecast can be thought of as the combination of a *diagnosis* of ongoing processes with a *prognosis* about the evolution of those processes. Doswell (1986a) has described this in terms of the truncated Taylor series approximation:

$$\Phi(t_o + \delta t) = \Phi(t_o) + \left. \frac{\partial \Phi}{\partial t} \right|_o \delta t, \quad (1)$$

where $\Phi = \Phi(\mathbf{X}; t)$ is some state vector that is a function of its position in space (denoted by the vector \mathbf{X}) and time t . Formally, this is comparable to the methods used in NWP models, but is likely to be a poor characterization of the human, subjective weather forecasting process. Metaphorically, however, it can still be appropriate in the context of human forecasting in the sense that a forecaster begins with as accurate an understanding of ongoing processes as possible (a *diagnosis*), and then combines this knowledge with some method for estimating the time evolution of the situation (*prognosis*, as described in Doswell and Maddox 1986). Initial condition/diagnosis uncertainty is associated with the first term on the rhs of (1). The limitations on our observations mean that the state vector at the initial time $t = t_o$ is given by

$$\Phi(\mathbf{X}; t_o) = \tilde{\Phi} + \Phi', \quad (2)$$

where $\tilde{\Phi}$ is the unknown true value and Φ' is the observational error.

Observations are uncertain because the instruments used to obtain them are imperfect (instrument error), and because their distribution in time and space is inadequate (sampling error). Sampling error is usually the most important source of meteorological initial condition uncertainty, leading to an imperfect diagnosis—ongoing physical processes that are inadequately re-

solved cannot be quantitatively accounted for in a diagnosis that contains sampling deficiencies. Although qualitative observations of small-scale processes may, at times, be possible with remote sensing systems like radar and spacecraftborne multispectral imaging, these remote sensing systems have a limited capability to describe those processes in terms of the standard meteorological variables (i.e., those used in the governing equations).³ Sampling error that produces uncertainty in the initial conditions is also a major source of uncertainty with numerical prediction models, naturally. Instrument error imposes some additional uncertainty—its importance is greatest when sampling error is small, because sampling error generally dominates observational uncertainty in meteorology.

b. Model uncertainty

Model/prognosis uncertainty is associated with the second term on the rhs of (1), by which the initial state is extrapolated forward in time. Because our understanding of the atmosphere is imperfect, it is natural to believe that even if we somehow had obtained perfect observations of the initial state, our predicted evolution of the initial state would fail to be perfect as a consequence of an imperfect prognostic model. In any real forecasting system, objective or subjective, the prognosis (temporal extrapolation) of the model state by humans is dependent in a complex way on the initial condition uncertainty. The diagnosis of ongoing meteorological processes is just as critical for human forecasters as the accurate specification of the initial conditions is for a numerical prediction model. An inappropriate subjective forecasting model is a possible consequence of a poor diagnosis, whereas in objective forecasting, the model is typically fixed, in terms of some set of equations used for prognosis.

As discussed in Doswell (1986a), simple *linear* extrapolation in time is a low-order forecast model—its limitations with respect to a nonlinear atmosphere are obvious, although for short projection times, linear extrapolation forecasts might achieve some useful level of accuracy. The time period during which linear extrapolation is acceptable depends on (a) the accuracy threshold chosen, (b) the meteorological process being extrapolated, and (c) the specific example under consideration. Some processes, such as fronts, can persist and move more or less uniformly for days. Other processes, such as thunderstorms, evolve rapidly on time scales of an order of 10 min or less. Moreover, each instance of

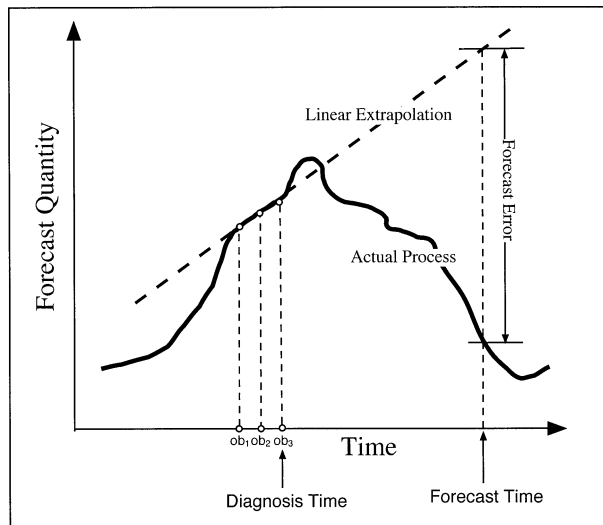


FIG. 1. Schematic illustration of some actual process (bold solid line) as measured by some forecast quantity as a function of time (solid line). A series of observations (Ob_1 , Ob_2 , and Ob_3) provide a basis for linear extrapolation (dashed line) from the diagnosis time (Ob_3) to some forecast time F incurring some error at the forecast time. Given some limit on what is considered acceptable, at some point after the diagnosis time (Ob_3) the forecast exceeds the threshold of acceptability. That time is called the *valid linear extrapolation time* in Doswell (1986a).

a front or thunderstorm is never identical in all aspects to any other such instance. Some fronts (or thunderstorms) evolve more rapidly than others. Hence, the period of valid linear extrapolation (Fig. 1) varies from time to time, event to event, and case to case.

Nonlinear extrapolation can produce forecasts that exceed the chosen accuracy bounds more rapidly than forecasts based on simple linear extrapolation—a *bad* nonlinear model (perhaps poorly chosen by the forecaster based on a difficult diagnosis) can be *worse* than linear extrapolation, especially when the initial diagnosis is wrong (Fig. 2). What is obviously desired is a *good* nonlinear forecast model, such that any initial condition uncertainty does not grow rapidly with extrapolation time and allows a reasonably accurate prognosis. Such models are, unfortunately, hard to come by.

c. Uncertainty about the use of information

Given modern observation and analysis technology, human forecasters have compared the flood of new information in forecast offices to trying to drink from a fire hose; the torrent of data and products derived from the data threatens to overwhelm them. Stewart et al. (1992) have shown that simply having *more* information available does not necessarily result in forecast improvement. Forecasters have become concerned about knowing what are the most pertinent and helpful products to examine and use, and just precisely *how* to use those products when forecasting is subject to operational

³ Humans are limited in their capacity to absorb data in a finite time, so they are subject to a different sort of sampling error than objective forecasting systems (like NWP); they too can only consider a subset of the available data. However, humans also can use their capacity for pattern recognition to incorporate *nonquantitative* observations (e.g., graphical images), whereas most objective diagnosis systems are as yet unable to make direct use of patterns, per se.

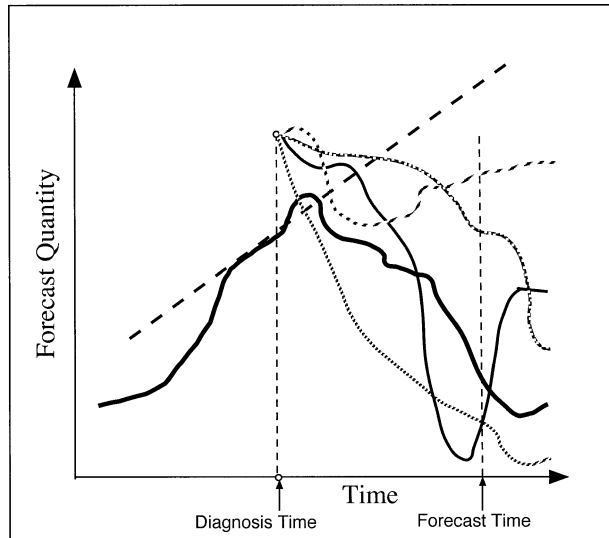


FIG. 2. Schematic illustration, using the same hypothetical process as in Fig. 1, that a faulty diagnosis at the initial time often leads to bad prognoses, with four possible extrapolations of varying quality based on the initial diagnosis.

forecasting deadlines. Education and training are critical factors that have been less than successfully developed in most operational forecasting (e.g., see Doswell et al. 1981; Doswell 1986b), but whenever there is no known objective path to the optimal use of information, human weather forecasters must rely on their experience and intuition derived therefrom (see section 5a, below), that is, heuristics.

d. The duality of error

The notion of the 2×2 contingency table represents the correspondence between forecasts and observations, but it carries within it some inherent uncertainty of its own, regarding what Hammond (1996) refers to as “policy formation.” Perfect forecasting would produce only correct dichotomous forecasts, and so all of the forecasts would be on the principal diagonal of Table 1; the off-diagonal cells would be empty. Consider Fig. 3, which is similar to a so-called Taylor–Russell diagram (Hammond 1996; Stewart 2000). The figure depicts the verification results for forecasts of “severe” events when a threshold is used to convert forecaster-perceived probability (Murphy 1985) into a dichotomous (or categorical) forecast of a severe weather event. The area of the ellipse shown is a measure of the overall forecasting accuracy of the forecasting system (which might be for an individual forecaster or a group of forecasters). Uncertainty increases the spread of the observed magnitudes for any given forecast probability—absolutely perfect forecasts would mean only two forecast probabilities, zero and unity. No severe events would occur with zero forecast probability, while all of the observed severe events would occur with a forecast probability

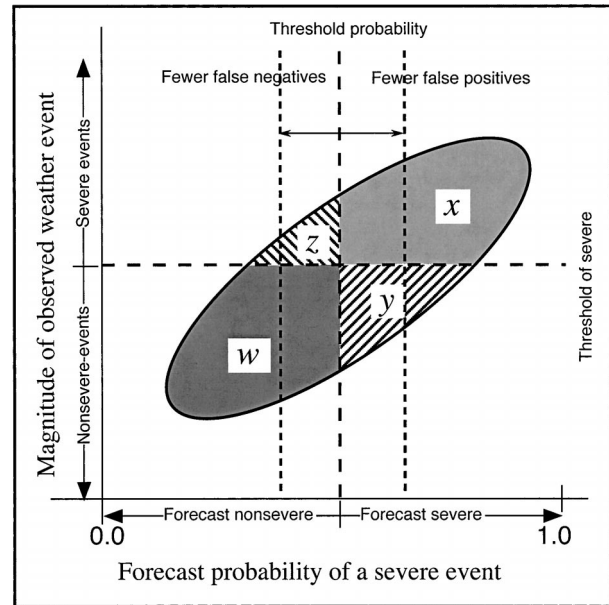


FIG. 3. A graphical depiction of the contents of Table 1, using presentation similar to a Taylor–Russell diagram, showing the consequences of irreducible uncertainty associated with any forecasting system.

of unity. Because severe weather is generally a rare event, the diagram implicitly is dominated by a large number of points corresponding to correct forecasts of nonsevere events, most of which are easy forecasts, as discussed in Doswell et al. (1990). Note that the threshold for defining a severe event (the horizontal line on Fig. 3) can be rather arbitrary (Doswell 1985), whereas the probability threshold for deciding to issue a categorical severe forecast (the vertical line on Fig. 3) is a matter of policy. Knowledge of meteorology affects neither of these.

Uncertainty carries with it the inevitability of both false positives and false negatives, depending on where the thresholds fall. This relationship constitutes the duality of error: at a given level of forecast accuracy for the system, *false negatives can only be reduced by increasing the false positives, and vice versa*. The concept of the duality of error (Hammond 1996; Stewart 2000) is directly related to signal detection theory (Mason 1982), which has been used in a recent article on tornado warnings by Brooks (2004). To some extent, forecasters might be able to improve their forecasting ability in various ways, thereby reducing the size of their own personal ellipse. However, there is some inherent limit to the accuracy of *any* forecasting system, which could be described as the “state of the science,” but is going to be difficult to define because that ultimate predictability is generally not known quantitatively.

As discussed in Hammond (1996), the aim of *research* is to reduce the area of the ellipse; that is, increase the accuracy of the forecasting system. However, it is the role of a *policy maker* to determine the ratio of

false negatives to false positives. That is, someone must decide what is the optimum ratio of false negatives to false positives, which defines where the threshold probability lies on Fig. 3. Generally in weather forecasting, false negatives are seen as a less desirable outcome than false positives ("false alarms"), because they are associated with the unfavorable notion of an unforecast weather event, perhaps with casualties as a result. False positives have their own costs that are not necessarily trivial, but generally do not usually cause casualties. This asymmetry in the perceived penalties for the two types of forecast errors means that it is common for the threshold to be pushed toward the left on Fig. 3. If a forecaster receives no *explicit* guidance about how to choose this threshold, then each forecaster must make her/his own choice about how to make the decision when issuing dichotomous forecasts in the face of irreducible uncertainty. In addition to producing forecast bias, this leads directly to inconsistent performance among forecasters and forecast offices, which surely decreases the accuracy of the collective forecasting effort.

3. Assessment of uncertainty by forecasters

For objective forecasting methods, such as MOS, it can be relatively straightforward to develop probabilistic forecasts. In weather forecasting by humans, however, the imperfect available data, including output from various objective forecasting systems, are used by the forecasters to make both a diagnosis and a prognosis. Diagnosis uncertainty may not be *formally* incorporated into any uncertainty statement regarding the forecast, but this diagnosis uncertainty nevertheless plays an important role in the prognosis uncertainty (as suggested by Fig. 2).

Humans obviously do not solve the equations of fluid dynamics in their heads to make weather predictions. Moreover, Kahneman and Tversky (1973) have shown that people (including forecasters) do not generally apply the quantitative methods of statistical analysis to make decisions, either. The nearly unlimited amounts of quantitative information available to forecasters, in principle, are usually impractical to use to make decisions under the typical operational conditions of finite time for forecast preparation. Human forecasters apply heuristics to simplify the process of reaching a decision in the face of all of this information complexity. Examples of such rules include (a) *availability*, which refers to how readily information thought to be relevant comes to hand (and mind); (b) *representativeness*, which refers to how the situation under consideration is perceived to fit some model of the situation; and (c) *anchoring and adjustment*, which refers to the process by which quantitative uncertainty assessments begin with some anchoring value and are adjusted to account for the available information.

Whereas accurate assessments can surely result from

the use of heuristics, it is also true that heuristics can, at times, be a source of biases that degrade forecast quality. From a purely formal, analytical viewpoint, judgment under uncertainty using data in the form of multiple fallible indicators can be based on Bayesian probability theory. If the probability of event x is denoted by $p(x)$, whereas the probability of event x , given that some indicator y has been observed, is denoted by $p(x|y)$ and called a *conditional* probability, then Bayes' theorem can be stated in the form

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \quad (3)$$

The simple probabilities are sometimes called *prior* probabilities and the conditional probabilities are sometimes called *posterior* probabilities, indicating that they are probabilities after some condition has been met. Suppose that H represents some hypothesis about the data, and D represents the existing data, while another, competing hypothesis about the same data is denoted by H' . If we substitute in (3) with $y = H$ or H' and $x = D$, and then form the ratio

$$\frac{p(H|D)}{p(H'|D)} = \frac{p(D|H) p(H)}{p(D|H') p(H')}, \quad (4)$$

then this can be expressed more simply as

$$\Omega_1 = L\Omega_o, \quad (5)$$

where Ω_1 is the posterior odds in favor of hypothesis H over H' , whereas Ω_o is the prior odds, and $L = p(D|H)/p(D|H')$ is often referred to as a *likelihood ratio*. When considering the observations, forecasters often must choose between two or more competing hypotheses and, according to Bayesian principles, should do so in a way compatible with (4) or (5). I know of no forecasters in operational forecasting who follow this formal procedure, although it is conceivable that some decision aid system might be developed to make such an analysis for a forecaster.

In this argument, note that $p(H|D)$ does not generally equal $p(D|H)$. This asymmetry creates biases in judgment because much of the case study-based scientific research about meteorological hypotheses presumes that the event has, in fact, occurred (see Doswell et al. 2002 for a rare contrary example), and so provides quantitative information about $p(D|H)$. However, forecasters usually deal with the opposite situation in operational practice: $p(H|D)$. Forecasters typically do not have detailed prior and posterior probability information at hand. Further, it is likely that a complete diagnosis carried out according to Bayesian logical principals would involve a *series* of applications of the theory. That is, each hypothesis would comprise a series of subhypotheses that would consider different aspects of the data separately, because it is unlikely that a human would be able to consider all aspects of the data (including objective forecast guidance) at once. This adds consid-

erable complexity to the analysis and virtually assures the inability of human forecasters to apply objective Bayesian logic to the operational process of making diagnostic judgments.

As a simple example of this in forecasting practice, when an upper-level potential vorticity anomaly moves over a front, it is widely recognized that the probability of low-level cyclogenesis increases. Most forecasters would not be familiar with the quantitative (prior) probability of this, however. The development of that cyclone would, in turn, increase the probability of poleward advection of moisture and the eastward movement of high-lapse-rate air above the surface, possibly creating a convectively unstable environment that would increase the probability of severe thunderstorms. Each step would formally involve an application of Bayes' theorem using whatever prior and posterior probability information is available to the forecaster. As already noted, I know of no human forecaster who follows such a procedure, and this raises the possibility of biased assessments of the probability of a forecast event. Nevertheless, humans accomplish this task using heuristics with some considerable success, in spite of *not* following such a formal procedure (e.g., Murphy and Winkler 1977; Stewart 2000).

It has been demonstrated repeatedly that simple linear models can simulate fairly accurately how experts will make judgments using multiple fallible indicators (e.g., Fischhoff 1982a). These simple linear models actually tend to *outperform* the experts tested in such studies because the models make judgments in an entirely consistent way. Human experts are not always consistent in their judgment, typically to the detriment of their performance. On the other hand, weather forecasting experts seem consistently able to outperform objective methods based on both linear and nonlinear models, but human forecasts still have been shown in some cases to be predictable by a simple linear model (e.g., Lusk et al. 1990). These findings seem to suggest that additional studies aimed specifically at weather forecasters are needed to resolve the extent to which humans are able to outperform simple linear models, and the circumstances under which such models might equal or outperform human forecasters.

It is a principle of probability that the more uncertain the prediction, the more regressive the forecast should be. By this is meant that the statistical notion of *regression toward the mean*—when uncertainty about some forecast situation is at its maximum, the best forecast is climatology (the mean). Among other things, for weather forecasts, it means that the useful range of forecast departures from climatology should decrease with increasing forecast projection time because forecast uncertainty inevitably increases with projection time. At the predictability limit (whatever projection time that might be for a given forecast problem), the useful range of forecasts becomes zero and the forecast should become a constant given by climatology at all forecast

ranges beyond that point. Use of this principle in forecast practice varies, and not accounting for it can result in unnecessarily poor forecast verification.

It is my experience that most human forecasters indeed develop a *personal* set of rules (heuristics) for accomplishing the tasks of diagnosis and forecasting with the available information. The development of personal heuristics begins during the education and early operational practice phases of a forecaster's career and, in some cases, proceeds indefinitely. Oskamp (1965) has shown that *confidence* in an expert's own judgment typically increases with experience virtually indefinitely. As time goes on, practitioners of all sorts, including weather forecasters, of course, are increasingly confident in their assessments and predictions. Unfortunately, Oskamp (1965) also finds that the objective *accuracy* of those judgments does not invariably and continuously increase in step with that increased confidence. Rather, accuracy tends to level off fairly early and remain at some roughly constant level once its maximum has been obtained. Further, Oskamp (1965) observes that *confidence in a judgment is not typically well related to the objective accuracy of that judgment*. This finding is particularly disturbing because it provides an empirical basis for my impression that many forecasters tend to resist revision of their thinking at some point in their careers, even in the face of continuing objective evidence that their heuristic approaches are, in fact, in *need* of revision. This problem is well recognized in judgment and decision-making studies, and various explanations have been advanced. Lichtenstein et al. (1982) have said that one partial remedy for overconfidence is for the forecaster to try to imagine ways in which one's judgments could be wrong. They also have asserted that it is typically challenging to assess how difficult or easy a particular forecast task might be; there is uncertainty about how to assess uncertainty.

As noted already, weather forecasters have an admirable record of being able to do that very thing, to assess uncertainty. When precipitation forecasting was mandated in 1965 by the National Weather Service (NWS) to be in probabilistic terms, subjective assessment of precipitation probabilities came to be a routine part of every NWS forecaster's job. The encouraging part of the results of making probability-based forecasts every day is that human weather forecasters have demonstrated a degree of *reliability*⁴ in the task that is considered remarkable to nonmeteorologists (e.g., Fischhoff 1982b, p. 439). Most studies of probability assessment in the judgment and decision-making literature have not used weather forecasters as subjects, and the

⁴ By reliability is meant the degree to which the observed frequencies of forecast events match the forecast probabilities. In a diagram with forecast probabilities on one axis and observed relative frequencies on the other (the so-called reliability diagram), the results for human forecasters generally are found near the 45° line, corresponding to perfect reliability.

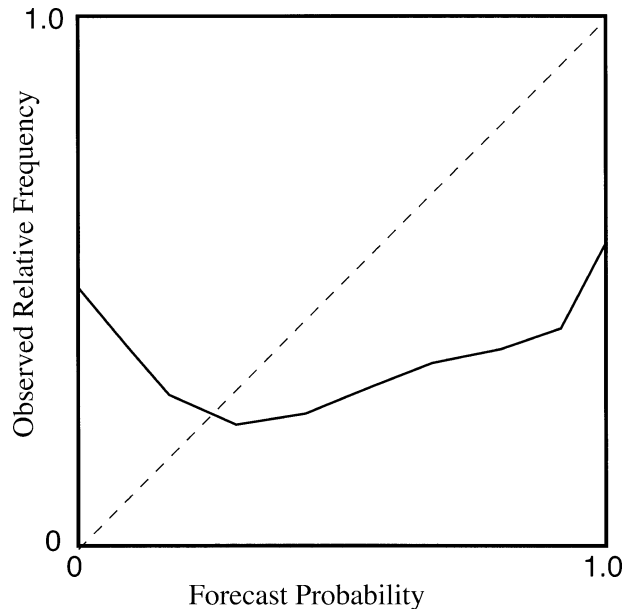


FIG. 4. The flat reliability curve associated with an overconfident forecaster, where the dashed line represents perfectly reliable probability forecasts, such that the observed relative frequency always equals the forecast probability. It can be seen that the observed frequency is larger than it should be for low-probability forecasts and is smaller than it should be for high-probability forecasts—a characteristic of overconfidence.

results of those studies typically show reliabilities vastly inferior to those of weather forecasters. Reasons for this disparity are not clear.

Exhibiting reliability alone does not guarantee that the forecasts are in fact accurate, even though accurate forecasts would necessarily be reliable (see Stewart 2001). The correspondence between forecasts and observations involves much more than reliability (see Murphy 1991b, 1993); nevertheless, *calibration* of forecasts to achieve reliability is an important first step to accurate forecasts. Forecasters new to the task of making probability assessments typically exhibit overconfidence in their ability to forecast and tend to produce subjective probabilities that are too high in the upper range of forecast probabilities and too low in the lower range. In terms of reliability, their verification results tend to look like Fig. 4.

Overconfidence can be overcome to some extent through the mechanism of feedback. In judgment and decision-making literature (e.g., Lichtenstein et al. 1982) the issue of calibrating probability assessments has been studied, again not typically with weather forecasters as subjects. In such studies, results seem to show that subjects may or may not get some limited benefits from feedback, but their reliability tends to remain modest. On the other hand, Murphy and Winkler (1977, 1982) have shown examples where experienced *weather* forecasters demonstrate considerable reliability even when taking on new tasks for which they have had only

limited experience at expression in probabilistic terms. Kay and Brooks (2000) or Vescio and Thompson (2001) have provided more recent examples. It may be that daily experience with decision making under uncertainty makes operational weather forecasters particularly adept at estimating their uncertainty, at least as a group, compared to other subjects of such studies.

Considerable variability of calibration is possible among individual forecasters. Some weather forecasters are better calibrated than others on a consistent basis, but no formal research in this area has been done. Note that calibration is an issue that has nothing to do with meteorological knowledge. Rather, it can be seen as a technical issue associated with translating meteorological knowledge into probability assessments. It is possible to improve weather forecast accuracy through improved calibration and, hence, achieve improved verification scores *without* adding any new meteorological understanding. There is some doubt in the judgment and decision-making literature that the degree of calibration would be easy to measure in a meaningful way if the goal is to compare individuals, however (Lichtenstein et al. 1982).

If we consider the simple 2×2 contingency table (Table 1), Lichtenstein et al. (1982) have suggested that one problem with calibration and other aspects of forecast judgment is that human forecasters might well be ignoring one or more of the cells in even this simplest of forecast contingency tables. Doswell et al. (1990) have shown, for example, that in forecasting rare events, forecast verification measures like the critical success index (CSI; also known as the “threat score”) fail to account for the predominance of correct forecasts of nonevents (element w in Table 1). The asymmetry in the perceived penalty for false negatives versus false positives (discussed at the end of section 2) results in forecasts with a bias—systematic overforecasting of weather events. This is particularly troublesome in the process of producing *hazardous* weather forecasts, because hazardous weather forecasts and warnings have been traditionally dichotomous. Given that the forecaster is forced to issue a binary product for most forecasts (with the exception of probability of precipitation), it is impossible to express uncertainty in its most natural way (see Sanders 1963; Murphy 1991a). One clear benefit of probabilistic warnings is that it encourages forecasting with as little bias as possible (Murphy 1997, p. 35).

4. Consideration of biases associated with selected heuristics

a. Availability

The notion of “availability” as used by cognitive psychologists includes several notions that might not be obvious from the name alone. If the general notion of availability is concerned with the ease of bringing pre-

vious similar instances and conceptual models to mind, then a forecaster with a poor memory will have difficulty in recalling instances that might be perceived to offer value in making a judgment. Some forecasters find it easy to recall instances with considerable accuracy, others might recall pertinent examples, but their memory might be faulty, and, of course, some forecaster might find it difficult to recall specifics confidently enough to use them.

Within this context, a bias associated with egocentric recollection can occur. That is, the problems with which an individual forecaster is directly associated are likely to take precedence over relevant information experienced by someone else. Because the experience base of shift forecasters is never exactly the same, the essentially random association between weather events and any individual forecaster will influence what information that forecaster uses. If it did not happen on *your* forecast shifts, whatever value that experience might potentially have for the situation at hand, that event is not likely to play as much of a role in your decision making in comparison to someone who actually worked the case.

Another topic tied to availability is the unique set of task structures that individual forecasters use. I have observed that forecasters usually develop a personal “forecast rote” and so the diagnostic tools and prognostic models used routinely become a sort of filter giving each forecaster a unique and selective window on the available information. Because differences in the task structures can be substantial and diverse, it seems obvious that certain situations will give good results with a particular set of task structures, whereas other situations will yield poor results with that same task structure. The task structures preferred by a particular forecaster might change with time, or might be very nearly constant over long periods. It is widely accepted that our expectations shape our perceptions and influence our judgments, and those expectations are part of the process by which an individual develops a forecast rote. In principle, if all forecasters are well calibrated and have roughly equal meteorological knowledge, they would arrive at similar probability assessments in all situations, irrespective of how they structure their tasks. In reality, of course, such an ideal is not usually achieved.

b. Representativeness

The representativeness heuristic includes a number of topics that should be familiar to weather forecasters. If a particular weather situation is viewed as representing a class of similar situations, this might be referred to in forecaster jargon as a “pattern recognition” approach to forecasting. This topic is clearly connected to the availability heuristic, because forecaster experience will vary. However, another issue that can be grouped under the “representativeness” heuristic is the notion of sam-

ple size. It has been learned that most subjects of judgment and decision-making studies are unaware of the impact of sample size on the probability that a small sample is representative of the population as a whole (Kahneman and Tversky 1972). Many subjects implicitly believe that a small sample is as likely to represent the whole population as a large sample. Even when they are warned about making this error beforehand, they often ignore that warning when called upon to make a judgment in a specific case.

When applying this heuristic to operational weather forecasting at least two distinctly different topics are suggested. For a given situation, the uncertainties associated with sampling errors can lead a forecaster to consider the data to be suggesting that the situation belongs to one class of events when it actually belongs in a different class. This would generally reflect diagnosis error. When confronted with a similar dataset for a different situation, a forecast may or may not make the same error.

However, another sort of error associated with the representativeness heuristic would be to assume that the behavior of a small sample of events belonging to a certain class will be adequate to infer the behavior of most such events. This would likely manifest itself as a prognosis error, although it can also influence the diagnosis.

In weather forecasting, each day can be considered unique and a day exactly like it in every aspect will never be encountered again. Hence, it can be argued that weather forecasters are trapped with trying to deal with the smallest possible sample size—a sample of *one*—every day. To some extent, of course, particular days can share many characteristics with other, similar days. This leads to the notion of pattern recognition in weather forecasting. Cognitive psychology has only begun to address the notion of pattern recognition (Hammond 1996, 196–200), but it has long been a traditional tool in human weather forecasting (see Johns and Doswell 1992; Moller 2001). Even if it is argued that pattern recognition can be a powerful method for making weather forecasts, there are many issues associated with sample size tied to its application. If the sample of those related cases is small, then the confidence with which that experience can be applied to a particular new event should be less than if many similar cases have been observed. And, of course, because each day is different, it could well be the case that many forecast days would fail to fit within *any* of the forecaster’s collection of patterns stored in memory. Mentioning memory should remind us that a forecaster’s memory for details might well be faulty, and an objective pattern comparison between the day in question and the archetype as recalled by the forecaster might not be all that good. The number of samples of a particular pattern that is needed is a function of the variability within the archetype. In general, statistical logic asserts that the greater the variability, the larger the sample needs to be. In using pat-

tern recognition, it seems likely that many forecasters may not be accounting properly for the sample size problem in their judgments.

5. Modes of forecaster cognition

a. Analysis versus intuition

In the judgment and decision-making literature (e.g., Hammond 1996) it is recognized that there are two distinct ways to make judgments under uncertainty: 1) *analysis*, which is defined as a stepwise, conscious, logically defensible argument, and 2) *intuition*, which is defined as a process that is the opposite of analysis. Hammond (1996, chapter 3) has made a strong argument that humans use *both* thought processes to solve problems, and so human forecaster cognition might be seen as forming a continuum between these apparent polar opposites. Hammond also suggests that human cognition can *oscillate* between analysis and intuition. Some tasks might be predominantly done analytically, whereas others might be done primarily in an intuitive mode, and still others would require both in roughly equal proportion. When human weather forecasting is considered in this light, it is apparent that forecast judgments typically involve *both* analytical and intuitive elements [described in Doswell (1986b) as left-brain versus right-brain thinking], and individual forecasters may vary as to where they fall on the continuum between the polar extremes of purely analytical versus purely intuitive. Moreover, some weather situations might be more amenable to analysis than others.

Weather forecasting has proceeded along a path that began with *entirely* intuitive cognition (see Nebecker 1995, chapter 4). The first forecasters had no scientific basis for what they were attempting to do and proceeded entirely on the basis of their intuition, which would be highly dependent on their personal experience. Even today, it is not necessary to be an educated and trained meteorologist to be a forecaster; farmers, pilots, and other people whose livelihood depends on the weather often use intuitive forecasting methods without any analytical knowledge whatsoever. The accuracy of their forecasts is essentially undocumented, of course. With the development of the science of meteorology and its use of mathematical, physical, and statistical logic, the possibility of analytical approaches was introduced. Generally speaking, it is accepted in judgment and decision-making literature that analysis replaces intuition wherever possible (Hammond 1996, chapter 3). Nebecker (1995, p. 40) describes this as a drive to replace "art" with "science." Thus, the ascendancy of NWP and objective forecasting methodologies seem to be inevitably replacing human intuitive judgment. Of course, the continuing incompleteness of meteorological theory leaves room for human intuition (see the preface of Petterssen 1956; Schwerdtfeger 1981), but the path to the future might be interpreted to suggest the inevitable dominance

of analysis over intuition. This certainly is the vision described by Nebecker (1995, chapter 11) that pervaded the early history of NWP and apparently persists widely today.

However, Lorenz's (1963) insights regarding nonlinear dynamics and chaos also make it clear that, barring some presently unforeseen breakthrough, the weather is going to remain resistant to becoming deterministically predictable, even as NWP models and observational technology continue to improve. This suggests to me that it remains to be seen to what extent human intuition ultimately can be replaced by analysis. Can human forecasters provide useful predictability beyond the limits imposed by nonlinear dynamics? This is an interesting issue that invites further study.

Judgment and decision-making studies (Hammond 1996, chapter 3) have shown that analytic reasoning tends to produce highly accurate results, but occasionally produces very large errors because the analytic model is being applied outside of the range of conditions for which it is suited. On the other hand, intuitive reasoning produces results that may have small *average* error but are more widely dispersed. This is illustrated schematically in Fig. 5.

Well-tuned objective forecasting methods, like MOS, produce results that look much more like Fig. 5b than Fig. 5a (see Brooks and Doswell 1996), but under some circumstances analysis can fail to give any answer at all, because the conditions under which it would appropriately be applied are obviously not satisfied at all. Occasionally, the systems supporting the objective methods (notably, computers and broadband communications links) fail. Intuition is robust in that it *always* can provide an answer under virtually any circumstances, although that answer might not be a particularly accurate one and its consistency may at times be poor. Working with weather data is how forecasters derive their intuition, and it is doubtful that anyone is literally born with "instincts" about meteorology. Instead, human weather forecasters accumulate experience by working with what are called *multiple fallible indicators* (or "cues" based on the data). Because every forecaster's experiences are different, it is not surprising that different forecasters can arrive at divergent judgments when using the same data (e.g., Uccellini et al. 1992), whereas objective methods always produce the same results when given identical data. If dependence on objective methods discourages forecasters from working with meteorological data, then they likely will not develop the proper "intuition" about atmospheric structure and behavior (see Doswell and Maddox 1996).

b. Coherence versus correspondence thinking

Hammond (1996, chapter 4) also refers to a dichotomy in the approaches to making judgments and decisions: the distinction between abstract knowledge (*coherence*) and knowledge gained from empirical data

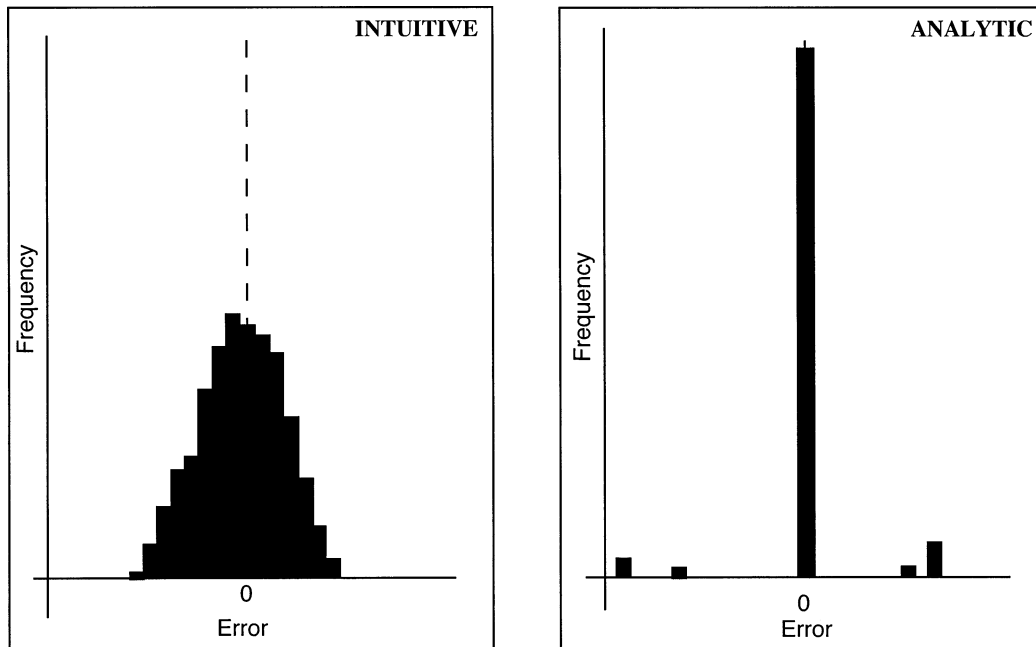


FIG. 5. Schematic illustration of the error distributions for (left) intuitive and (right) analytic cognition (after Stewart 2001).

analysis (*correspondence*). Unlike the continuum that humans develop between the polar opposites of analysis and intuition, coherence and correspondence approaches remain dichotomous, according to Hammond. Correspondence approaches are concerned with the empirically determined accuracy of the predictions made, whereas coherence approaches are concerned with the logical consistency of the judgments made. Note that logical consistency is no guarantee of empirical accuracy. Hammond (1996) sees no middle ground between such viewpoints, although he argues that both can be of use in making judgments under uncertainty.

Meteorological theory clearly represents abstract knowledge, but such theory is not always in the form of *mathematical* models. A conceptual model, such as the Norwegian Cyclone Model (NCM), is a nonmathematical form of abstraction—a so-called *mental model* (see Rouse and Morris 1986). It is a prototypical example of a conceptual model in meteorology. In the particular case of the NCM, the reasoning follows from the particular to the general, or so-called *inductive* reasoning. Many cases were analyzed and then a generalized conceptual model was constructed. The NCM includes not only the fronts and general pattern of isobars, but also the distribution of weather in a cyclone-relative framework.

When applied in practice, the NCM continues to be a strong influence on subjective weather forecasts, despite its recognized deficiencies (e.g., Mass 1991). Of course, the experience of most forecasters in using this model to forecast the weather is that even though the conceptual evolution of a cyclone may indeed resemble, more or less closely, the actual evolution of fronts and

isobars, the distribution of the sensible weather (e.g., clouds and precipitation) can be very different from that associated with the NCM. A forecast based entirely on such a model is coherent (in that it is internally consistent), but it may not correspond well to the observations. That is, when the conceptual model is used *deductively* to argue from the general to the particular, its accuracy can be limited, indeed. Nevertheless, forecasters continue to apply the NCM despite repeated failures in using it in the past. The tendency to prefer coherence to correspondence is common. People are loath to abandon their abstract understanding even in the face of repeated empirical failures when applying that understanding. This can lead forecasters to interpret data contradictory to their mental models as being *supportive* of those models through elaborate rationalizations, rather than acknowledging that the data are simply inconsistent with their model (Hammond 1996, chapter 3) and then seeking a revised model.

This issue is of considerable interest in judgment and decision making, and it indeed is of considerable importance in weather forecasting. Even research scientists of all sorts, including meteorologists, are vulnerable to preferring coherence to correspondence. A coherent conceptual model can persist for many years in spite of its poor performance in terms of making accurate weather forecasts (see Evans and Doswell 2001). Most meteorologists will find it easy to think of their own examples of this problem.

6. Discussion

The preceding suggests that a potentially fruitful collaboration could develop between weather forecasters

and those who study cognition as well as judgment and decision making. Forecasters continue to be able to improve on the guidance forecasts they receive from various objective forecasting systems. It stands to reason that learning more about *how* human forecasters achieve this, perhaps with special attention paid to cognitive styles between those forecasters who consistently perform best and those who perform poorly, might result in an improved human forecast product overall. There is some evidence that weather forecasters have some characteristics that make them quite different from the typical subjects chosen for judgment and decision-making studies (e.g., Lichtenstein et al. 1982; Stewart et al. 1997). There also is evidence that weather forecasters may *not* be so different from the usual subjects of such research (Hammond 1996, p. 283; Stewart et al. 1992). Resolution of the somewhat contradictory results from the existing limited studies using human weather forecasters as subjects is necessary to determine the extent to which the existing understanding within the field of judgment and decision making can be applied to human weather forecasting.

It should be clear that no one is going to be able to apply certain analytical modes of cognition without education and training. For example, the principles of Bayesian logic are central to the tasks undertaken by forecasting, even if they are not formally applied in practice. Simply having a reasonable level of understanding of Bayesian principles would almost surely be of assistance in making probability assessments, even though a formal quantitative application of them might be impractical. Comparable statements could be made regarding topics such as quasigeostrophic theory. Studies cited by Hammond (1996) have shown that education and training can indeed be useful in making analytical reasoning and coherence-based judgments. These works have shown that naïve subjects without such education and training do considerably worse at probability assessment than those who do have the requisite understanding. Coherence-based cognition is not possible without the requisite education and training (see Roebber and Bosart 1996).

To date, the process of weather forecasting by humans has not been subjected to a thorough and comprehensive study, perhaps because it is widely believed that weather forecasting by humans will eventually disappear in favor of wholly objective processes. The early history of NWP created the dream of a purely objective forecast procedure, with human subjectivity removed. The gap in verification between humans and objective methods has been decreasing, especially as seen by the developers of objective guidance forecasts (e.g., Charba and Klein 1980). If that trend continues, it is inevitable that subjective forecasting will eventually be overtaken and rendered obsolete by objective methods. However, we do not know what systematic differences in cognitive styles might exist between forecasters who consistently perform well and those whose performance is from me-

diocre to poor. We do not know how best to combine analytical methods and products with human intuitive approaches to produce the most accurate human-produced forecasts. Thus, we do not know how to go about raising the overall performance level of human forecasts in comparison to objective methods, nor does there seem to be any evidence for a commitment to learn about such things.

If the operative assumption is that analysis will drive out intuition entirely, then the absence of research aimed specifically at human weather forecasting is a moot point. If, on the other hand, there is a commitment to having humans involved in weather forecasting into the indefinite future, the general dearth of such studies is inconsistent with any envisioned future role for human weather forecasters. For the management of forecasting organizations to be *demonstrably* committed to a future for humans in the process, the dedication of resources to this critically important task is essential. A consistent collaboration between meteorologists, cognitive psychologists, and others involved in judgment and decision-making research will be necessary if the goal of improving human weather forecasting is to be achieved. Such interdisciplinary work is often underfunded and, consequently, usually has more lip service than results. The failure to commit significant resources to this collaboration is tantamount to conceding the forecasting role to purely objective methods in the near future; then, the only issue is *when* it will happen, not *if* it will happen.

Acknowledgments. I have benefited from innumerable discussions with many colleagues on this and related topics over many years. Notable contributors, in no particular order, include Harold Brooks, Steve Weiss, Bob Johns, Charlie Liles, Tom Stewart, Barry Schwartz, Mike Foster, Rich Thompson, Larry Wilson, Jim Johnson, Roger Edwards, Lance Bosart, Les Lemon, Fred Sanders, Don Baker, Cliff Mass, Paul Roebber, Dave Schultz, and especially Al Moller, Bob Maddox, and the late Allan Murphy. This listing is by no means comprehensive and I value all such discussions, whether or not I managed to recall a particular name at the time of this writing. Tom Stewart kindly assisted with valuable comments on an early draft, as did Romu Romero, but any mistakes that might remain concerning the interpretation of judgment and decision-making studies are purely my own. Tom Stewart, Neil Stuart, and an anonymous reviewer offered valuable reviews of this manuscript. This work was partially supported by the Ministerio de Educación, Cultura y Deporte, of Spain, Grant SAB2002-0084.

REFERENCES

- Brooks, H. E., 2004: Tornado-warning performance in the past and the future: A perspective from signal detection theory. *Bull. Amer. Meteor. Soc.*, **85**, 837–843.

- , and C. A. Doswell III, 1996: A comparison of measures-oriented and distributions-oriented approaches to forecast verification. *Wea. Forecasting*, **11**, 288–303.
- Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on the skill and the potential skill of an ensemble prediction system. *Mon. Wea. Rev.*, **126**, 2503–2518.
- , M. Miller, and T. N. Palmer, 1999: Stochastic simulation of model uncertainties. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.
- Charba, J. P., and W. H. Klein, 1980: Skill in precipitation forecasting in the National Weather Service. *Bull. Amer. Meteor. Soc.*, **61**, 1546–1555.
- Doswell, C. A., III, 1985: Storm scale analysis. Vol. II, The operational meteorology of convective weather, NOAA Tech. Memo. ERL ESG-15, 240 pp.
- , 1986a: Short range forecasting. *Mesoscale Meteorology and Forecasting*, P. Ray, Ed., Amer. Meteor. Soc., 689–719.
- , 1986b: The human element in weather forecasting. *Natl. Wea. Dig.*, **11** (2), 6–17.
- , and R. A. Maddox, 1986: The role of diagnosis in weather forecasting. Preprints, *11th Conf. on Weather Forecasting and Analysis*, Kansas City, MO, Amer. Meteor. Soc., 177–182.
- , and —, 1996: A review of student performance on pretests given at the flash flood forecasting course. Preprints, *13th Conf. on Weather Analysis and Forecasting*, Vienna, VA, Amer. Meteor. Soc., 403–410.
- , R. Davies Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585.
- , D. V. Baker, and C. A. Liles, 2002: Recognition of negative mesoscale factors for severe weather potential: A case study. *Wea. Forecasting*, **17**, 937–954.
- Evans, J. S., and C. A. Doswell III, 2001: Examination of derecho environments using proximity soundings. *Wea. Forecasting*, **16**, 329–342.
- Fischhoff, B., 1982a: For those condemned to study the past: Heuristics and biases in hindsight. *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, Eds., Cambridge University Press, 335–351.
- , 1982b: Debiasing. *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, Eds., Cambridge University Press, 422–444.
- Hahn, B. B., E. Rall, and D. W. Klinger, 2002: Cognitive analysis of the warning forecaster task. Klein Associates, Inc., Final Rep. RA1330-02-SE-0280, NOAA/NWS Office of Climate, Water, and Weather Services, 26 pp. [Available from Klein Associates, Inc., 1750 Commerce Center Blvd North, Fairborn, OH 45324-6362.]
- Hammond, K. R., 1996: *Human Judgment and Social Policy*. Oxford University Press, 436 pp.
- Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612.
- Joliffe, I. T., and D. B. Stephenson, Eds., 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, 254 pp.
- Kahneman, D., and A. Tversky, 1972: Subjective probability: A judgment of representativeness. *Cognit. Psychol.*, **3**, 430–454.
- , and —, 1973: On the psychology of prediction. *Psychol. Rev.*, **80**, 237–251.
- , P. Slovic, and A. Tversky, Eds., 1982: *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, 555 pp.
- Kay, M. P., and H. E. Brooks, 2000: Verification of probabilistic severe storm forecasts at the SPC. Preprints, *20th Conf. on Severe Local Storms*, Orlando, FL, Amer. Meteor. Soc., 285–288.
- Lichtenstein, S., B. Fischhoff, and L. D. Phillips, 1982: Calibration of probabilities: The state of the art to 1980. *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, Eds., Cambridge University Press, 306–334.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- Lusk, C. M., T. R. Stewart, K. R. Hammond, and R. J. Potts, 1990: Judgment and decision making in dynamic tasks: The case of forecasting the microburst. *Wea. Forecasting*, **5**, 627–639.
- Mason, I. B., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–203.
- Mass, C., 1991: Synoptic frontal analysis. Time for a reassessment? *Bull. Amer. Meteor. Soc.*, **72**, 348–363.
- Moller, A. R., 2001: Severe local storms forecasting. *Severe Convective Storms, Meteor. Monogr.*, No. 50, Amer. Meteor. Soc., 433–480.
- Murphy, A. H., 1985: Probabilistic weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy and R. Katz, Eds., Westview Press, 337–377.
- , 1991a: Probabilities, odds, and forecasting rare events. *Wea. Forecasting*, **6**, 302–307.
- , 1991b: Forecast verification: Its complexity and dimensionality. *Mon. Wea. Rev.*, **119**, 1590–1601.
- , 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293.
- , 1997: Forecast verification. *Economic Value of Weather and Weather Forecasts*, R. W. Katz and A. H. Murphy, Eds., Cambridge University Press, 19–74.
- , and R. L. Winkler, 1977: Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *Natl. Wea. Dig.*, **2**, 2–9.
- , and —, 1982: Subjective probabilistic tornado forecasts: Some experimental results. *Mon. Wea. Rev.*, **110**, 1288–1297.
- , and —, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330–1338.
- Nebeker, F., 1995: *Calculating the Weather: Meteorology in the 20th Century*. Academic Press, 251 pp.
- Oskamp, S., 1965: Overconfidence in case-study judgments. *J. Consult. Psychol.*, **29**, 261–265.
- Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.
- Petterssen, S., 1956: *Weather Analysis and Forecasting*. 2d ed. McGraw-Hill, 428 pp.
- Roebber, P. J., and L. F. Bosart, 1996: The contributions of education and experience to forecast skill. *Wea. Forecasting*, **11**, 21–40.
- Rouse, W. B., and N. M. Morris, 1986: On looking into the black box: Prospects and limits in the search for mental models. *Psychol. Bull.*, **100**, 349–363.
- Sanders, F., 1963: On subjective probability forecasting. *J. Appl. Meteor.*, **2**, 191–201.
- Schwerdtfeger, W., 1981: Comments on Tor Bergeron's contributions to synoptic meteorology. *Pure Appl. Geophys.*, **119**, 501–509.
- Stensrud D. J., H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.*, **127**, 433–446.
- Stewart, T. R., 2000: Uncertainty, judgment, and error in prediction. *Science, Decision Making and the Future of Nature*, D. Sarewitz, R. A. Pielke, and R. Bierly, Eds., Island Press, 41–57.
- , 2001: Improving reliability of judgmental forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. Armstrong, Ed., Kluwer, 81–106.
- , W. R. Moninger, K. F. Heideman, and P. Reagan-Cirincione, 1992: Effect of improved information on the components of skill in weather forecasting. *Organiz. Behav. Hum. Decis. Proc.*, **53**, 107–134.
- , P. R. Roebber, and L. F. Bosart, 1997: The importance of the task in analyzing expert judgment. *Organiz. Behav. Hum. Decis. Proc.*, **69**, 205–219.
- Uccellini, L. W., S. F. Corfidi, N. W. Junker, P. J. Kocin, and D. A. Olson, 1992: Report on the surface analysis workshop at the National Meteorological Center 25–28 March 1991. *Bull. Amer. Meteor. Soc.*, **73**, 459–471.
- Vescio, M. D., and R. L. Thompson, 2001: Subjective tornado probability forecasts in severe weather watches. *Wea. Forecasting*, **16**, 192–195.
- Wikle, C. K., and C. J. Anderson, 2003: Climatological analysis of tornado report counts using a hierarchical Bayesian spatiotemporal model. *J. Geophys. Res.*, **108**, 9005, doi:10.1029/2002JD002806.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.