Journal of Contemporary Ethnography
Volume 35 Number 1
February 2006 3-23
© 2006 Sage Publications
10.1177/0891241605282241
http://jce.sagepub.com
hosted at
http://online.sagepub.com

Ground Truth

Verification Games in Operational Meteorology

Gary Alan Fine
Northwestern University

To gain occupational legitimacy, workers depend on claims of accuracy. How are these claims assessed? Organizations judge competence and effectiveness using metrics, which can be massaged. Using a case study of operational meteorology, the author examines the organizational and self-presentational politics whereby meteorologists verify their predictions. Forecasts of future events are central to occupational work in operational meteorology. However, assessing what really happened "on the ground" is not unproblematic. So, meteorologists and their employers construct measures for assessments, even though these assessments are social and organizational routines. Verification statistics are signals whereby local offices are judged on effectiveness, and as a result, workers develop strategies to increase their verification scores. The author examines the production of verification at three local offices of the National Weather Service using ethnographic observations and in-depth interviews. Similar strategies operate when workers are judged in accountability systems, leading to forms of organizational impression management to demonstrate competence.

Keywords: *science; meteorology; verification; truth; sociology of knowledge*

What does it mean to be "right"? Much scientific work depends on communal criteria by which claims can be assessed as correct (Fleck 1979; Merton 1979; Abbott 1988). But how are these criteria used in practice? It is one matter to predict, quite another to evaluate a prediction. Although I focus on a single occupational domain, that of operational meteorology, this analysis applies to other scientific occupations and beyond. What is true in the case of applied science pertains in part to other work domains in which a logic of accountability is required (Bacharach and Mundell 1993; Ingersoll 2003; Wiener 2000).

Author's Note: I would like to thank Adele Clarke, Tim Hallett, and Jason Owen-Smith for help and advice in the preparation of this article.

The demand for verification is widely recognized within the sciences, and it constitutes part of the scientific ethos (Merton 1979). To achieve institutional legitimacy, technologies and strategies of verification must be established. Although there are several models of verification at issue in science, including experimental, observational, and predictive verification (see Fine 2005), my concern is with the latter. The verification of predictions is often found in applied science. Workers forecast events, and then these predictions are checked for accuracy as the future unfurls. Predictive verification is found in petrology (Bowker 1994), medicine (Christakis 1999), economics (Evans 1999), hydrology, earth science, and meteorology. Errors, when wide enough, can provoke legal liability (e.g., malpractice), needless expense, and mass inconvenience.

For a system of evaluation, a prediction is compared with ostensibly "objective" measures of what has been forecast. These measures are constructed from assessment devices—readings from scientific instruments or other measures—and this assessment is taken as separate from personal judgments. We feel confident that we can compare a prediction and the subsequent event.

To explore verification as an institutional practice, I focus on professional meteorologists employed by the National Weather Service (NWS). Meteorologists are an apt case for analyzing the problem of judging accuracy in that as a matter of occupational routine, forecasters produce claims about future events that are checked and evaluated. Their status, both as individual workers and as a group, results from how well they are able to ascertain through their instruments, models, and applications of scientific theories what transpired. Did the weather occur as predicted, and how, practically and organizationally, can one judge if that is the case? Furthermore, as federal employees, they are constrained by the logic of accountability, found in many governmental and economic domains, making verification institutionally central.

Not every claim is subject to verification. Some are not deemed sufficiently important, and this selection of claims produces bias in judging the accuracy rate. This is particularly salient when we examine predictions that some event will not happen.[1] This bias in which only selected predictions are tested is known as "the forecaster's dilemma" (Erhman and Shugan 1995).

As in all sciences, the truth claims of meteorology depend on imperfect equipment, including observational apparatuses, communication tools, and computer programs. The decision of whether forecasters "got it right" results from their equipment. A gap of uncertain dimensions exists between the claims derived from mechanical inscription and what actually happened. In

addition to this technological gap, a temporal gap exists referring to the difference between when data were gathered and when they were distributed.

Meteorologists speak of these challenges as constituting what they term the problems of *ground truth* and *verification*. The former, referring to the reality of meteorological conditions, is used when a forecaster is attempting to determine what is happening in a short-term severe weather event, whereas the latter covers the organizational assessment of both severe weather and longer term forecasts.

Studying Forecasters

I spent fifteen months conducting ethnographic research at three local offices of the NWS, the U.S. Department of Commerce agency that provides meteorological forecasts for the United States. For each office (Chicago, Belvedere, and Flowerland),[2] I collected field notes and tape-recorded thirty interviews, lasting from one to three hours. At the Chicago office, I spent approximately three days a week observing for a period of six months, and then once or twice a week for the next year. During this period, I attempted to be present during all times of day and night, and I spent two midnight shifts with the forecasters. At the other offices, I spent ten full days observing over the course of three months. In each office, I sat by the forecasters, observing and asking questions about the practices of creating forecasts, about the conditions of work, and about their work identities.

Through the use of computerized models of likely weather patterns, coupled with observations (through radar, satellite images, upper-air balloons, and precipitation and temperature measurements), local operational forecasters have the responsibility to forecast weather conditions for the next ten days, as well as to issue severe weather warnings when dangerous conditions (winter storms, tornadoes, severe thunderstorms, flash flooding, heat waves) approach. Routine forecasts are issued twice a day: in the late afternoon (typically at 3:30 p.m.) and early morning (4:00 a.m.). In most offices, two meteorologists work in tandem; one is in charge of the short-term forecast, while the other manages the longer term forecast. The offices operate on a 24/7/365 schedule, with three shifts (day, evening, and midnight). (For more details about the ethnography and work routines, see Fine 2005).

The NWS operates 122 local offices, with each having a limited geographical area (the County Warning Area [CWA]) for which it is responsible. The sizes of these areas are tied to what can be covered by the current generation of Doppler radar. For instance, the Chicago office serves twenty-three

counties, stretching from Rockford, Illinois, northwest of Chicago, to five counties in Indiana on the east. Each NWS office employs approximately two dozen workers, including ten forecasters. I also spent two weeks at the Storm Prediction Center (SPC), located in Norman, Oklahoma, the office with the responsibility for putting out watches for tornadoes and severe thunderstorms. The SPC, one of seven centers the NWS operates, employs a staff of approximately forty, including twenty forecasters.

Knowing Ground Truth

Sitting in a comfortable, secure office, forecasters are required by their institutional position to judge potentially dynamic and tempestuous conditions. Surrounding them are numerous computer screens on which are displayed the output of various measuring devices and modeling systems. These images are supposed to—and do—tell forecasters what they need to know about the world that they have the responsibility to divine.

Although mechanical representations are often taken for granted, such a stance is not inevitable (Latour 1987), and the mechanical claims must fit atmospheric understandings: the "commonsense assumptions" of a trained professional. Consider a case from the Belvedere office:

> A little after 9:16 a.m. (14:16 Zulu—or global time) the thermometer reports a brief temperature spike for Midwest City to 60 degrees, although the temperature on either side of the spike was about 55 degrees. Garth and Dom discuss whether this temperature is "real" or whether it is an "equipment error." Since it is a reading from Midwest Airport, they speculate that it might be jet exhaust, although they have never seen this problem before. They notice that the temperature in North Boyer was 60 degrees at 14:23, which made sense because clouds over North Boyer were breaking up, and might have explained the temperature spike in Midwest City. However, they then realize that 14:23 was on a local 24-hour clock, representing 2:23 p.m., five hours from the Midwest City spike. Dom comments about the Midwest City temperature: "No one on site could tell you that the observation was accurate. There is nothing meteorologically sound that said that this did or did not happen. There is no way of knowing it." They decide that the North Boyer temperature reading is correct, and the Midwest City temperature is an equipment error. Garth comments about automatic equipment replacing observers, "Now you have greater detail, more data, but you have more doubt when it is anomalous, because there is no one there to interrogate the data." (Field notes)

To know a meteorological reality, these forecasters rely on their shared assumptions and experiences of how weather systems operate. If one elimi-

nates human observers, one also eliminates the possibility of discretionary questioning of the data other than through challenging the machines. The search for ground truth, as forecasters describe it, is focused on deciding "what is real," given organizational demands to produce useful information. On some level, forecasters are hostage to mechanical claims that rest outside their control. Facts, seemingly objective, become claims that are locally produced through organizational choice, but by being seemingly objective, they serve to create a hegemonic zone, preventing questioning. It is surely unfair to suggest that all that exists is a patina of truth, but adherence to the mechanical claims constitutes a bureaucratic strategy, evident when the otherwise taken-for-granted mechanically produced truth is challenged by lived experience.

Verification Games

Surely it is reasonable to inquire of any activity whether it meets the goals that participants have set for themselves and that others have set. How are we to assess work? When a government forecaster makes a prediction, citizens have a reasonable expectation that there will be accountability, and this accountability shifts organizational priorities.[3] Yet what does a claim that a forecaster is accurate mean? Are we substituting precision for validity (Bowker and Star 1999:24)? John King's observation about policy arguments is pithy and true: "some numbers beat no numbers every time" (Bowker and Star 1999:103; see Porter 1995).

Consider the following claims:

> Accuracy improved but little in the first half century of the weather service, hovering around the 75 to 85 percent mark—and this was for the next day's weather. . . . In 1941, five-day Weather Bureau forecasts had an accuracy rate of only 48 percent for temperatures and 16 percent for precipitation. (Laskin 1996:144-45)

> A three-day forecast in many areas of the country is only about 50% accurate. . . . It means that the forecaster has "blown the forecast" three days into the future just about as many times as he's gotten it right! (Fishman and Kalish 1994:2-3)

These passages are from books by science writers sympathetic to operational meteorology. Yet the authors do not indicate what "accuracy" means. The numbers sound precise, but precise for what? What does it mean for a forecast to be 50 percent accurate or even accurate 50 percent of the time? Consider temperatures.[4] If a forecaster predicts that the temperature will reach 54

degrees, and the high is 53 degrees, is it wrong (current thermometers have reliability of plus or minus 2 degrees)? Suppose the high is 51, 45, or 36. At what point does the forecast become wrong? Even though forecasts are made for particular locations, they are taken by the public as referring to the city as a whole. In Chicago, the high temperature by the lake might be 42 degrees, while the temperature in the southwestern corner of the city is 64 degrees, a range that is by no means impossible. Averaging these two measurements produces a high of 53 degrees. Will the public consider this forecast correct? Are we to judge high temperatures, low temperatures, or average temperatures? If average temperatures, how is this to be figured? The official Chicago reading is from O'Hare International Airport. Should that be the measuring point, even though it is far from the population center of Cook County? Do we average the high and low temperature or average the reading for each hour? These are organizational choices.

Judging precipitation is as problematic. Suppose a forecaster claims that there is a 40 percent chance of rain. Is the forecaster correct if it rains or if it doesn't? Suppose there are occasional or scattered showers. Should O'Hare serve as the measuring point for rain and snow, as the organizational practices require? It may not rain at O'Hare, but it may in downtown Chicago. So what does accuracy mean? When one moves from temperature and precipitation to the accuracy of a forecast as a whole, things become even more complex. As weather consumers, we sense when a forecast is right, but this refers to whether the forecast inconveniences us (getting us cold, hot, or wet or needlessly canceling our plans) and is not linked directly to numerical precision.

Given these interpretive challenges, verification is construed by the NWS as an organizational practice. As a government agency at a moment at which *institutional accountability* is a watchword (Ingersoll 2003), the ability to claim institutional effectiveness is important. The head of the NWS at the time of this research, retired general Jack Kelly, has been successful bureaucratically by emphasizing "objective" (or at least numerical) indicators of organizational competence.

That the bureaucratic emphasis on touting numerical indicators of effectiveness and using them for organizational assessment has not always characterized the NWS is suggested by the claim of an administrator that "I was a manager who believed in verification before verification became cool" (interview). A similar view is evident in the claim of a forecaster from an earlier period:

> We don't keep track of our accuracy rates and compare our scores with the other stations. That is childish. We're all using the same materials. Sooner or later

> you'll be right and they'll be wrong and vice versa. When you're wrong you don't want to commit suicide. It's really not that big a deal. (Laskin 1996:189)

Such a remark would no longer be spoken publicly. Today, performance-based measures are demanded, so the question is how staff actions can be constructed to demonstrate success. Analogous to the case of evidence-based medicine or accountability in education, we find performance-based meteorology. Crafting measures of organizational competency is a bureaucratic strategy, and a successful one in that President George W. Bush and his aides, as well as congressional leaders, cite the NWS as an agency that works. They demonstrate their confidence by pointing to the statistics that the agency has compiled from their bureaucratic self-interest.

As with any organizational practice, the doing of verification is complex. However, local offices are fundamentally concerned with two forms of verification: those assessing routine forecasts (temperature and precipitation) and those assessing severe weather warnings.[5]

For every public forecast (two per day), forecasters fill out a Coded Cities Forecast, in which they specify the high and low temperature for each time period and indicate the probability of precipitation. Each office has a set of observing locations for which it makes internally designated forecasts. In Chicago, meteorologists forecast for O'Hare and Rockford airports. The test is whether a human forecaster can beat the predictions made by statistical models (what are called model output statistics; see Oreskes, Shrader-Frechette, and Belitz 1994; Edwards 1999). The organizational goal is to beat the model output statistics.[6] When the models are substantially wrong on forecasted temperature, they are easy for a skilled meteorologist to beat. But even when the models are effective, strategies permit forecasters to beat them. If the model suggests a 90 percent chance of rain, a clever forecaster can best the model by specifying a 100 percent chance of rain. Assuming that it rains, the forecaster will get credit in the verification statistics as having greater confidence than the model prediction. Precipitation verification compares the percentage likelihood of precipitation in light of what happened at the recording station.

The verification of severe weather operates differently. Once severe weather is forecast, the test is whether the predicted event occurred.[7] Was there a tornado, or did a thunderstorm reach severe criteria (winds of fifty-eight miles per hour or three-quarter-inch hail)? How can one judge whether a forecast was correct, considering that images on the radar are not sufficient to verify? A human observer is necessary unless the storm passes directly over a wind gauge. On occasion, storms are verified after the fact through damage to trees or property. If forecasters learn that a tornado or severe thun-

derstorm occurred that they hadn't predicted (and needless to say, searching for such a storm would not be a high priority), this counts against their verification statistics, because the NWS counts the number of severe storms for which no warnings were issued.

Problems of Inscription

In verifying forecasts, meteorologists are at the mercy of their machines. The automatic temperature-recording machine is considered accurate to within two degrees. Measuring snow depends on whether one considers the water content of snow and the degree to which it is packed.[8] The height at which wind is measured affects wind speed. Even small changes in where equipment is sited can alter the results. As one forecaster explained,

> I worked at the airport. Sensors ideally were supposed to be placed close to midfield. Representative of the air field. We were then losing the temperature and dew point there, and they found out it was because the cable had been cut and spliced so many times water was getting in, kind of shorting it out. So they just moved the equipment, the temperature dew point sensor closer to where the office was. Instead of being out on the center of the airfield where it was level. . . . So it dropped off maybe twelve, fifteen feet. . . . You had a lot of cold air drainage. (Interview)

Whenever equipment is replaced or moved, the continuity of records becomes confounded, and the meaning of the measurements becomes problematic.

Verification and Value

Verification is tied to a system of values given institutional priority. Although the NWS does not currently use verification statistics in setting salaries for individual forecasters,[9] offices are required to gather, and to display, their collective verification statistics, transforming these numbers into a means by which reputations are established. Regional and national administrators treat these numbers as an "objective" basis on which to distribute resources to successful offices. Verification statistics are taken as signals, among the few systematic signals available to permit a hierarchical ranking of effectiveness. The Belvedere office, with an office culture that focuses on accountability, emphasizes its statistics by placing charts in a public area. Staff members claim that the emphasis on accountability is beneficial. As one told me about verification statistics, "It is a good check on our mind. If there was no verification, I would still forecast as I did, but it's still important to know" (field notes). A paradox is evident. Meteorologists realize that

often enough, forecasts do not verify and that forecasting with verification in mind may make one overly cautious, ignoring possible, but unlikely, risks. Yet this is their only "objective" measure of ability.[10] Forecasters are exhorted not to worry about verification—to do the best they can—but are then faced with the numerical representation of what is ostensibly a secondary concern. Verification statistics, however secondary they might be claimed, are a rhetorical resource used by administrators. One urged his forecasters to be concerned with the science but simultaneously that "we're constantly watching our performance measures. . . . We've been top gun for winter weather" (field notes), revealing the multiple, cross-cutting concerns involved.

Others are rhetorically far more suspicious of verification, although they too bow to the organizational demands for a metric accounting of their abilities. The basic complaint is that verification is a game and doesn't provide a "true" measure of forecasting. Rather than seeing how human and model combine (the "man-machine mix": "how the forecaster interprets the model against what actually happens" [interview]), the verification of daily weather involves comparing the human forecast against a machine-generated model. The organizational question is whether a human forecaster improves the model predictions, but for the public, the question is different: Is the prediction seen as useful? As one meteorologist explained,

> I want to know what I've done compared to reality, because my ultimate goal in the perfect world is to put out the perfect forecast, which I know is an unrealistic goal, but I think the effort should always be striving in that direction. Higher-ups in the weather service compare us more to the numerical models with the idea that if more of the numerical models are better than a forecaster than what do we need a forecaster for? . . . With a lot of this fair weather that we've had this winter, the models do as good as we do, maybe at times a little better. For fair weather, we're not as cost effective perhaps. But with this big winter storm coming up, I can almost guarantee that time in and time out, we will have a tendency to beat the models. You don't pay the fireman for when he's sitting in at the fire station waiting for the bell to ring. You pay him when he's at your house putting the fire out. (Interview)

The image of verification as an organizational game is routine (Long 1958), even among those resigned to the practice. I was told that "everyone plays the system," playing a "numbers game" (field notes):

> They're playing numbers against the models. They're not forecasting. . . they say, "Well, if the numbers say this, so that's what I'm going to say, so my numbers match." . . . [They're trying to] beat the numbers. And I think that is a terri-

> ble pitfall that we can fall into very easily. And I've seen forecasters here do that. (Interview)

An administrator, noting that the game involves impression management, instructed his staff, "We're supposed to post these things [verification statistics], so find a place where we can post them so no one will see them, unless we're doing better than average [comparing themselves to other offices]" (field notes). I never learned if they were doing better than average, but the sheets were posted in an obscure location.

Another administrator commented, half joking, that perhaps they should not attempt to warn for small F0 tornadoes, because "you're doomed to failure as far as the numbers. . . . We just don't get the verification on F0 tornadoes." Indeed, of the 282 tornado warnings in this CWA over fifteen years, only 121 were verified, and most of these were through hail or wind damage, with only 38 verified by visual sightings. The same was true of severe thunderstorms. Of 1,924 warnings, only 798 were verified. The number of "actual" tornadoes or severe thunderstorms is anyone's guess. Presumably, at least some of the false alarms were correct warnings in which no one saw the severe weather, and some verified storms were accepted on the claims of mistaken eyewitnesses. A warning, even for a real event, may have negative organizational consequences if staff members cannot find legitimate evidence to verify the warning. An administrator commented, "We've got to verify them and hope that we hit on a few of them" (field notes). A distinction is drawn between what is important and what is organizationally necessary, as in this sarcastic comment about verifying a heat warning: "We're going to verify our warning with a body count. You've got to verify somehow. . . . You verify twice. Once by the heat and then by the number of people killed" (field notes).

Creating Verification

The mission of any science is to present the contours of the "real world" in a way that audiences accept, and meteorology is no exception. As a result, having an organizational gauge of how successful one is in predicting severe weather is crucial, even if, as noted above, most warnings do not verify. But how does one know what has happened, given the absence of data? How can meteorologists discover ground truth?

A tornado is real enough if one is caught in its midst, but most tornadoes are small events, not much stronger than a heavy wind. If the vortex skips over a field or lake it may leave only minor marks. Sparsely populated land tells no tales (Monmonier 1997:91). If the timing is wrong—when the world

sleeps—in effect, the storm never happened. The same is true even if the storm was observed but the observer doesn't call. If the public doesn't notice, only the physical impact of the storm remains: holes in a tilled garden from hail, broken branches from high wind, or a circular pattern of debris from a tornado. Looking at a small storm that produced one-inch hail, a forecaster commented, "Those storms don't look like anything on radar. Those [hail] stones must have just fallen where someone was looking because they don't look like a big area" (field notes). In such cases, there is a strong temptation not to warn for storms that might barely meet the criteria for severe weather. As one forecaster explained,

> What we had was really pulse storms. Even when you warned it would have weakened by the time the warning went up. My preference is not to warn because it is on the border. The chance of getting a sighting would be rare. (Field notes)

Whether this is desirable for public safety, it is justified as an organizational strategy to maximize verification.

Other forecasters detail the challenge of determining the true weather conditions in the face of an opaque reality and a public that doesn't always report what it knows:

> We say we can use radar to initiate a forecast for a tornado. We cannot use the radar to verify [the warning]. A lot of these things happened in open areas where there is no opportunity to get a verification. So it happens to people where they have not issued because they know they can't verify. . . . Burke County had a severe thunderstorm going through there, and one of the forecasters said straight out, "I'm not going to put out a warning for that because nobody's ever reported anything from there, and I've never been able to get a verification." (Interview)

> Not everything gets reported. If you see your tree blows down in front of your yard, and you don't happen to report it into the police, and they don't happen to report it in to us, then I don't hear about it. So then it's like we missed the warning, but it wasn't missing the warning. So, if Farmer Joe was there, [and] he doesn't report "I had three-quarter-inch hail and part of my cornfield was flattened." Well, if he doesn't pick up the phone. . . . Even if he picked up the phone and reported it to somebody that somebody now doesn't report it back here. So if we don't hear about it, then it didn't occur. But did it occur? Of course it occurred. Just because we didn't get a report doesn't mean it didn't occur. (Interview)

The absence of a report is not the absence of a storm. Even if a report is made, a decision must be made as to how it is to be treated. Forecasters want reports, but not if these reports do not support their claims. For weather spotters to be organizationally useful, they must relate what forecasters want to hear. Forecasters desire public responses, but calls from the public can be a nuisance, providing incorrect information, creating a distraction from required tasks, or contributing to forecasting errors.

The Public as Problem

The verification process assumes a direct relation between what is claimed and what actually happened, but this is problematic even when dealing with eyewitness testimony. These claims are often from individuals who have not been trained, those observing weather conditions at a distance and/or under conditions of poor visibility, and those in situations in which excitement or fear may cloud judgment.

To create a corps of trained observers, local offices arrange several dozen "spotter" talks each spring, training employees from governmental agencies and the interested public to watch for severe weather. In two hours, only rudimentary material can be covered. The goal, at least in theory, is to produce a public that can provide authority for the warnings through increased verification, providing mortar for an edifice of scientific forecasts. Yet these external observers can also complicate the task of the forecasters, providing incorrect information or verifying what should not be verified. One forecaster estimated that only one in twenty will become a useful spotter. As a result, forecasters take a jaundiced view of public reports:

> We don't want the public to report. The vast number of people have no idea what they're looking at. . . . You got to rely on the so-called reliable people [police, fire or emergency management workers]. . . . Most people are basically deadwood. [He refers to a tornado report from an outlying county.] You see how some people panic over nothing. They just let their imagination run wild. . . . The public is going to become totally useless now that they all have cell phones in the cars. They see everything. . . . That's why we're paid big bucks. Not to panic. Here's a weather law. If it don't spin, don't turn it in. When I go to a spotter talk, I hammer that in. (Field notes)

> I never put out [a warning] unless I see it on the screen. It's bogus. I ask, "Did you see it?" Anyone who picks up the phone and doesn't question the call is not doing a service. I tell people at spotter training that I will question them. (Field notes)

In practice, spotter reports are given variable weight on the basis of reputations and the assumption of competence. Keli Tarp (2001) observed,

> A spotter gave a report on amateur radio of a tornado from a storm that did not look particularly threatening. When the meteorologist who took the report told the warning forecaster, he responded, "I don't believe that." The first meteorologist then said, "That was X, I would believe it," and the short term forecaster concurred, "I don't think we have a better spotter." The warning forecaster immediately began issuing a tornado warning for that county, even as he was saying, "It doesn't look like a tornadic storm." (p. 11)

As Tarp noted, spotters gain reputations that influence how their reports are treated, part of the way that all scientific claims are evaluated (Barnes, Bloor, and Henry 1996; Collins 1992). Such reputations are also tied to their organizational role. Calls from "Joe Public" are treated with less respect than those with recognized institutional positions. When the public calls, the report is likely to be accepted if it supports the forecaster's idea or if it verifies a warning, but it may be dismissed in the absence of other human or mechanical reports if it claims something that a forecaster doubts, doesn't match other claims, or means that the forecaster hasn't verified a severe storm:

> Randy informs Bert about an unwarned thunderstorm, noting that the County Emergency Management office "has a [public] report of ping-pong ball size hail, but they have two spotters up there who didn't see anything." Because the report would have meant that they missed severe weather and because it was an indirect public report, the forecasters don't take it seriously. Burt jokes, "Right place at the right time. . . . We got calls from the media, 'What did happen?'" Randy says, "The County heard on the radio that they had ping-pong size hail, but they think not." They treat the storm as though it was not severe. (Field notes)

Institutional standing provides the presumption of legitimacy, and as a result, in contrast to the variable acceptance of public reports, those whose role provides institutional standing are given deference. The validity of a report is a situated evaluation, on the basis of organizational credibility and authority.

Manufacturing Verification

Given the demand for forecasters to verify their predictions, strategies are developed to massage and generate verification. The local office verifies its own reports, so, as a meteorologist put it, "it's like the fox guarding the

chicken coop." Another claimed, "You prepare the man to cheat." A third commented about the verification of flood warnings, "Since we're the ones who put the data on the national system for verification, we can take poetic license" (Field notes). Is an impartial verification system possible when evidence is ambiguous and subjective? How can validity be generated when the evaluator is not a disinterested party? Because the verification of severe weather is not checked, it is easy for forecasters to claim that the public verified each of their warnings: a public of thin air. In fact, suspicion is expressed when other offices verify every warning. One forecaster commented, "I hate to say it, but I think they may have made it up." Forecasters search for what is in their interest to believe:

> The Chicago office is in the middle of a severe weather outbreak, including a tornado warning.
>
> Lewis calls out: "I need a severe hail report from McKinley County, if anyone's going up there."
>
> Sid: "Spotter says half inch size hail, east side of Medford" [They need 3/4" to verify].
>
> Ritchie: "We couldn't get a quarter inch when we need it."
>
> Sid: "That's what the man said. You can influence what I write down."
>
> Ritchie calls Medford to see if he can verify the thunderstorm.
>
> They issue a tornado warning for Shelton County, but do not receive confirming reports from spotters: "All these spotters, and we don't hear squat back from any of them." They receive a call from a spotter in Shelton County who sees dime size hail and a green sky [an indication of a tornado]. Sid reports: "No sign of a tornado, but then he's in the basement. Smart guy."
>
> Sid: "There's a citizen report of a funnel over Afton."
>
> Lewis: "That's a good place for it to be" [given the radar image].
>
> Sid: "I'll put it down on the form, but funnel clouds don't count" [to verify tornadoes].
>
> They receive word of a waterspout in Lake Michigan near Afton. Sid remarks: "We know this is our tornado. It wasn't a waterspout, it was a tornado that hit the water." He decides that it was a tornado that hit the beach and then went into the water, and he tells others, "We had a confirmed touchdown right along the beach, then it went off-shore. Or it was right off-shore, because we had a spotter who watched it right off-shore. We had really good rotation." Although the spotter says that it was in Shorewood, they decide that it was actually in Afton, adding "That makes more sense with what the spotter saw."
>
> Bert says about another tornado report, "Two residents say they saw a tornado touch down. We should take it with a grain of salt."
>
> Sid responds: "It makes sense. Eyeballs work. It all fits in." This report from the public fits with what they expect, based on the radar, and so this report is given the credibility that it might not otherwise have. (Field notes)

As this episode indicates, verification is not self-evident but must be processed organizationally to fit into the logic of accountability. Ground truth is created as a work practice. One forecaster put the matter directly:

> Now there's so much pressure on proving the numbers. I think we're way too liberal in how we interpret things. Someone [a member of the public] says fifty- to sixty-mile-an-hour winds. If we didn't have a warning, we'll call it fifty. If we do, we'll call it sixty. Because, of course, you want to verify. You want to be right. If you put out a warning for Mohawk County, the next day you'll spend two hours making fifty phone calls trying to find something to verify the warning. If you didn't put out a warning, are you going to spend three hours making fifty phone calls? No. (Interview)

As is often the case in organizational culture, strain is expressed through joking (Fine and DeSoucey 2005). Forecasters kid about making the world fit their forecast by whatever means necessary:

> During severe weather, we send [an intern] out with a chain saw [to cut down a tree for verification]. . . . After a big snow, we send interns out to measure snow, and if they don't come back with the amounts we forecast, we send them out again. (Field notes)

This remark, part of the office's joking culture, uncovers occupational realities at the same time as it preserves credibility by being "only joking."

The Politics of Political Boundaries

If outright cheating is not often evident, frequently a proactive search for supportive information occurs. This search is built into the procedures, and given that local offices issue warnings by counties, political units, verification also is judged by county. A tornado that touches down just outside of a county boundary does not verify for the county that it barely missed. Being on the wrong side of the road, it constitutes an organizational failure. Given that counties have different resources and different attitudes to communicating with the NWS, responses to warnings vary, leading to forecasters not warning some counties of the possibility of minor severe weather:

> Sean reports at the staff meeting: "The day shift has to make the callbacks the next day to find out if the events are verified. It has to be done the next day. It needs to be done every time there are unverified warnings. You have to be specific. Do you have reports of structural damage? Do you have any reports of trees down? Do you have any power lines down? The more specific you are the more likely you will get a response. One needs only one form of verification. If

> you've got a severe weather warning for Orange County, [the head of Emergency Services] has computers all over the county." Sean emphasizes that Orange County is likely to verify severe weather, but other counties are not, based on the structure of emergency services. One suburban county attempted to integrate fifteen counties into a centralized warning system, but most counties in the less-populated areas have since dropped out, feeling that the results were not worth the organizational effort. Sean notes, "McKinley County, you're not going to get anything." Sean notes that even if they warn for a part of a county (e.g., eastern Orange County) and there is a confirmed tornado in another part of the county (e.g., western Orange County) that verifies the warning. He suggests that by issuing warnings for longer periods, they might increase their verification scores. He concludes, "I realize that verification is not supposed to be in the front of our heads, but that is what Congress looks at." (Field notes)

The location of a storm—and the organizational commitments that are tied to that location—contribute to the likelihood that events will become verified. Key to knowing are those practices that are linked to the politics of place.

Proactive Work

As Sean's quotation emphasizes, forecasters can be proactive, but they are proactive in verifying their warnings, not for the areas for which they didn't warn. A forecaster at the SPC explained, "When the events are marginal, [local offices] have a certain influence over them. If they have issued a warning, they will ask around. If it is an event they haven't warned for, they may not ask as hard" (field notes). One forecaster, commenting on the relatively low severe weather verification scores of his office, noted, "We definitely need to be making more calls. We definitely need to be more proactive in finding out what is going on. Sometimes it's just the nature of the game" (interview). Another added, "That's something that other offices do to pump up their verification" (interview). Consider these cases:

> Ritchie, the college student intern, calls local police departments and emergency offices in areas for which warnings were issued. Without his presence these calls would not have been made. As a result, almost all the warnings were verified.
>
> Sid remarks: "I think we've hit every county we've had a warning for. . . . I think we're going to bust in Dilling and Smythe." However, Ritchie says that he will keep calling Smythe until they admit they have a tree down. Eventually he gets verification from both counties.
>
> Sid says, "So far I'm batting 1000."
>
> Ritchie tells him, "You missed Blaine County. 990."

> Sid responds, "We might get a report tomorrow. We can scrounge up something. We can find something."
>
> Ritchie jokes, "No one should bat 1000. . . . Trees and powerlines down. I think that is a generic thing they are telling us. It counts as verification."
>
> Sid laughs, "Blaine, they'll find something."
>
> Ritchie tells me, "He'll go out with his car and hit a tree." (Field notes)

> After issuing a warning for southern Foster County, Bert calls the local police and asks if damaging winds occurred. Ritchie then calls a nearby reporting station and reports that in a previous storm cell (that wasn't warned for) there was a tree down and one-inch hail. Ritchie suggests, "We could fudge the time."
>
> Bert responds, humorously but as social control, "That's a big fudge."
>
> Sean tells Ritchie with certainty, "That's a bust. We didn't get a warning out."
>
> Bert jokes about southern Foster, "Anybody else who might have that information. Maybe Miss Cleo?" (Field notes)

The strategy involves both investigating and not investigating, hoping for confirmation. Although contact is often made after the fact, offices sometimes call counties before severe weather warnings are issued to prime them for impending bad weather. Like lawyers with guilty clients, forecasters learn to ask the proper leading questions, "Would you say that the hail was quarter size [not dime size]?" or "Would you say that the tornado was spotted five to ten [not fifteen] minutes ago?" The naively "wrong" answer can create organizational distress.

Although verification might appear a simple matter, the absence of the material effects of short-lived weather events, coupled with the inability to verify these events through the same technology by which they were originally noticed, challenges institutional practice. That organizations are committed to accountability makes decisions social and political as well as scientific.

The Politics of Verification

Although some peculiarities are associated with operational meteorology because of its role as an applied science and because of the immediacy and the human level quality of its predictions, issues of verification are found in many occupational worlds that make explicit or implicit claims about the effects of natural processes or human interventions. If weather has more salient issues because of its routinized and everyday qualities, it is not alone. We search for ground truth on which we can judge what has occurred and then create procedures of verification by which we can share those assessments with others in ways that reveal our credibility and demonstrate our

accountability as in required testing in elementary schools, in evidence-based medicine, or in assessing economic forecasting. Our claims to competency must be verified, establishing the legitimacy of expertise (Collins and Evans 2003). Techniques must be developed to determine the adequacy of these claims, and then strategies are developed to "work" these techniques for best advantage, providing favorable assessments of practitioners.

When teachers "teach to the test," they operate in the same world in which meteorologists predict in light of the limits of models. When doctors choose to perform only certain tests on their patients, they are operating in the world of the meteorologist, who only calls those agencies from which he or she wishes to hear. All are searching for their own form of ground truth, that which serves as a proper measure of presentation of self.

Strategies and practices develop to judge the adequacy of professional work products. Each occupation hopes to determine its own rules of practice in light of those forms of social control under which they operate. Accountability and verification become political matters, tied to occupational authority and institutional power. As operational meteorology teaches us, workers develop strategies that they consider legitimate and ethical and simultaneously support their self-presentational and organizational needs. One of the goals of weather service offices is to persuade the public that their claims should be treated as indications of their competence (Fine 2005). If meteorologists are to be allowed to predict, their claims must have a degree of legitimation, established through occupational, organizational, and self-presentational strategies. Trust, thus, becomes a central goal of the forms of organizational communication. This becomes particularly salient given that the NWS is a governmental agency, and forecasters claim expertise not only for the agency but on behalf of the entire state system. When government science is accepted, the state as a whole is legitimated.

When, as in the case of the NWS, rewards are given to offices on the basis of their "success," strategies develop to ensure that these success rates are as high as possible. From the standpoint of larger organizations, this is not necessarily a bad outcome, because they too are judged on the claims of their local offices or laboratories.

Choosing a single domain of truth claims limits the analysis, because of the effects of local work conditions. Meteorological verification practices must be compared and contrasted with other worlds: measurements of environmental pollution, petrology, climate modeling, or earthquake measurements, not to mention education, economics, and medicine. Each has institutional demands placed on the production of predictive claims and strategies for measurement of those claims.

Ultimately, verification is an organizational practice. Although it is linked to a physical obdurate reality, it is not transparent. The adjudication of truth is a collective practice on which social confidence depends as crafted measures of competence serve as bureaucratic strategy. Workers who claim a vision of the future view their creations through a window whose assumed transparency directs attention away from its distortions and shadows.

Notes

1. In the case of potential severe weather, a tornado warning must be verified to judge the effectiveness of a meteorological office. The forecast by itself does not demonstrate that a tornado actually occurred. In contrast, if an office chooses not to issue a warning, that belief that no tornado will hit will not be verified unless there is clear evidence that a tornado occurred. Correct assessments that severe weather will not occur do not contribute to verification rates.

2. The last two of these are pseudonyms. I conducted most of my observations at the Chicago office, and because so much of the data are linked to the geographical and meteorological conditions of the area, it would be disingenuous to use a pseudonym for that office. I take care to use pseudonyms to disguise the identities of informants as well as geographical locations.

3. Yesterday's forecast might be a curiosity, but today's forecast matters. Large numbers of Americans access weather news on a daily basis, using it to organize their immediate futures. Weather forecasts are consequential. As one meteorologist claimed, "If you blow the weather, you affect a lot of people and a lot of money" (field notes).

4. For the problems of determining temperature, see Christy (2002).

5. There are also assessments of aviation forecasts, but I do not discuss this side of meteorology.

6. In one area, these predictions are compared with actual weather observations, although office statistics are not compiled as a result. If forecasters predict the exact high and low temperatures for the following three forecast periods, they win a $50 bonus, called a "triple zero." These perfect forecasts are rare, occurring only a few times annually.

7. In most cases, these warnings occur in conditions in which the SPC has issued a watch for an area, priming the forecasters (and the public) to be aware of the possibility of severe weather.

8. One forecaster told me about how in one storm, three inches of snow fell, even though none stuck to the ground but melted immediately. However, they were measuring the amount of water, and the melted snow amounted to a three-inch snowfall (field notes).

9. There are many reasons for this decision. Some regions of the country have weather that is easier to verify. Furthermore, forecasters told me that sometimes, one hits a string of bad or complex weather, which could lead to lower verification statistics. Meteorologists (even the most skilled) believed that using verification statistics for setting salary would be counterproductive, although of course this is not all that dissimilar from teachers who complain about merit pay schedules, noting that some schools are harder to teach at and that sometimes teachers will be assigned disruptive children.

10. Forecasters also write prognostic discussions (called Area Forecast Discussions) in which they explain their rationales for the choices they make. However, these are not reviewed systematically for indications of competence.

References

Abbott, Andrew. 1988. *The system of professions*. Chicago: University of Chicago Press.

Bacharach, Samuel B., and Bryan L. Mundell. 1993. Organizational politics in schools: Micro-macro, and logics of action. *Educational Administration Quarterly* 29:423-52.

Barnes, Barry, David Bloor, and John Henry. 1996. *Scientific knowledge: A sociological analysis*. Chicago: University of Chicago Press.

Bowker, Geoffrey. 1994. *Science on the run: Information management and industrial geophysics at Schlumberger, 1920-1940*. Cambridge, MA: MIT Press.

Bowker, Geoffrey, and Leigh Star. 1999. *Sorting things out*. Cambridge, MA: MIT Press.

Christakis, Nicholas. 1999. *Death foretold: Prophecy and prognosis in medical care*. Chicago: University of Chicago Press.

Christy, John R. 2002. When was the hottest summer? A state climatologist struggles for an answer. *Bulletin of the American Meteorological Society* 83 (5): 723-34.

Collins, H. M. 1992. *Changing Order: Replication and induction in scientific practice*. Chicago: University of Chicago Press.

Collins, H. M., and Robert Evans. 2003. The third wave of science studies: Studies of expertise and experience. *Social Studies of Science* 32:235-96.

Edwards, Paul N. 1999. Global climate science: Uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture* 8:437-72.

Erhman, Chaim M., and Steven M. Shugan. 1995. The forecaster's dilemma. *Marketing Science* 14:123-47.

Evans, Robert. 1999. *Macroeconomic forecasting: A sociological appraisal*. London: Routledge.

Fine, Gary Alan. 2005. Authors of the storm: Operational meteorology and the production of the future. Unpublished manuscript.

Fine, Gary Alan, and Michaela DeSoucey. 2005. Joking cultures: Humor themes as social regulation in group life. *Humor* 18:1-24.

Fishman, Jack, and Robert Kalish. 1994. *The weather revolution: Innovations and imminent breakthroughs in accurate forecasting*. New York: Plenum.

Fleck, Ludwik. 1979. *Genesis and development of scientific fact*. Chicago: University of Chicago Press.

Ingersoll, Richard M. 2003. *Who controls teachers' work? Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.

Laskin, David. 1996. *Braving the elements: The stormy history of American weather*. New York: Anchor.

Latour, Bruno. 1987. *Science in action*. Cambridge, MA: Harvard University Press.

Long, Norton. 1958. The local community as an ecology of games. *American Journal of Sociology* 64:251-61.

Merton, Robert K. 1979. *The sociology of science: An episodic memoir*. Carbondale: Southern Illinois University Press.

Monmonier, Mark. 1997. *Cartographies of danger: Mapping hazards in America*. Chicago: University of Chicago Press.

Oreskes, Naomi, Kristin Shrader-Frechette, and Kenneth Belitz. 1994. Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641-46.

Porter, Theodore. 1995. *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Tarp, Keli Pirtle. 2001. Communication in the distributed cognition framework: An ethnographic study of a National Weather Service forecast office. Unpublished manuscript, University of Oklahoma.

Wiener, Carolyn. 2000. *The elusive quest: Accountability in hospitals*. Hawthorne, NY: Aldine.

Gary Alan Fine is John Evans Professor of Sociology at Northwestern University. He is currently a fellow at the Russell Sage Foundation. He is most recently the author of *Everyday Genius: Self-Taught Art and the Culture of Authenticity* (University of Chicago Press, 2004). His next ethnographic project will be an examination of the worlds of professional and amateur chess.