Interfaces
Vol. 38, No. 5, September–October 2008, pp. 382–405
ISSN 0092-2102 | EISSN 1526-551X | 08 | 3805 | 0382

DOI 10.1287/inte.1080.0383

Polar Bear Population Forecasts: A Public-Policy Forecasting Audit

J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104,
armstrong@wharton.upenn.edu

Kesten C. Green
Business and Economic Forecasting, Monash University, Victoria 3800, Australia,
kesten@kestencgreen.com

Willie Soon
Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138,
wsoon@cfa.harvard.edu

Calls to list polar bears as a threatened species under the United States Endangered Species Act are based on forecasts of substantial long-term declines in their population. Nine government reports were written to help US Fish and Wildlife Service managers decide whether or not to list polar bears as a threatened species. We assessed these reports based on evidence-based (scientific) forecasting principles. None of the reports referred to sources of scientific forecasting methodology. Of the nine, Amstrup et al. [Amstrup, S. C., B. G. Marcot, D. C. Douglas. 2007. Forecasting the rangewide status of polar bears at selected times in the 21st century. Administrative Report, USGS Alaska Science Center, Anchorage, AK.] and Hunter et al. [Hunter, C. M., H. Caswell, M. C. Runge, S. C. Amstrup, E. V. Regehr, I. Stirling. 2007. Polar bears in the Southern Beaufort Sea II: Demography and population growth in relation to sea ice conditions. Administrative Report, USGS Alaska Science Center, Anchorage, AK.] were the most relevant to the listing decision, and we devoted our attention to them. Their forecasting procedures depended on a complex set of assumptions, including the erroneous assumption that general circulation models provide valid forecasts of summer sea ice in the regions that polar bears inhabit. Nevertheless, we audited their conditional forecasts of what would happen to the polar bear population *assuming*, as the authors did, that the extent of summer sea ice would decrease substantially during the coming decades. We found that Amstrup et al. properly applied 15 percent of relevant forecasting principles and Hunter et al. 10 percent. Averaging across the two papers, 46 percent of the principles were clearly contravened and 23 percent were apparently contravened. Consequently, their forecasts are unscientific and inconsequential to decision makers. We recommend that researchers apply *all* relevant principles properly when important public-policy decisions depend on their forecasts.

Key words: adaptation; bias; climate change; decision making; endangered species; expert opinion; extinction; evaluation; evidence-based principles; expert judgment; forecasting methods; global warming; habitat loss; mathematical models; scientific method; sea ice.
History: This paper was refereed.

Despite widespread agreement that the polar bear population increased during recent years following the imposition of stricter hunting rules (Prestrud and Stirling 1994), new concerns have been expressed that climate change will threaten the survival of some subpopulations in the 21st century. Such concerns led the US Fish and Wildlife Service to consider listing polar bears as a threatened species under the United States Endangered Species Act. To list a species that is currently in good health must surely require valid forecasts that its population would, if it were not listed, decline to levels that threaten the viability of the species. The decision to list polar bears thus rests on long-term forecasts.

The US Geological Survey commissioned nine administrative reports to satisfy the request of the Secretary of the Interior and the Fish and Wildlife Service to conduct analyses. Our objective was to determine if the forecasts were derived from accepted scientific procedures. We first examined the references in the nine government reports. We then assessed the forecasting procedures described in two of the reports relative to forecasting principles. The forecasting principles that we used are derived from evidence obtained from scientific research that has shown the methods that provide the most accurate forecasts for a given situation and the methods to avoid.

Scientific Forecasting Procedures

Scientists have studied forecasting since the 1930s; Armstrong (1978, 1985) provide summaries of important findings from the extensive forecasting literature.

In the mid-1990s, Scott Armstrong established the Forecasting Principles Project to summarize all useful knowledge about forecasting. The evidence was codified as principles, or condition-action statements, to provide guidance on which methods to use under different circumstances. The project led to the *Principles of Forecasting* handbook (Armstrong 2001). Forty internationally recognized forecasting-method experts formulated the principles and 123 reviewed them. We refer to the evidence-based methods as scientific forecasting procedures.

The strongest evidence is derived from empirical studies that compare the performance of alternative methods; the weakest is based on received wisdom about proper procedures. Ideally, performance is assessed by the ability of the selected method to provide useful *ex ante* forecasts. However, some of the principles seem self-evident (e.g., "provide complete, simple, and clear explanations of methods") and, as long as they were unchallenged by the available evidence, were included in the principles list.

The principles were derived from many fields, including demography, economics, engineering, finance, management, medicine, psychology, politics, and weather; this ensured that they encapsulated all relevant evidence and would apply to all types of forecasting problems. Some reviewers of our research have suggested that the principles do not apply to the physical sciences. When we asked them for evidence to support that assertion, we did not receive useful responses. Readers can examine the principles and form their own judgments on this issue. For example, does the principle, "Ensure that information is reliable and that measurement error is low," not apply when forecasting polar bear numbers?

The forecasting principles are available at www.forecastingprinciples.com, a website that the International Institute of Forecasters sponsors. The directors of the site claim that it provides "all useful knowledge about forecasting" and invite visitors to submit any missing evidence. The website also provides forecasting audit software that includes a summary of the principles (which currently number 140) and the strength of evidence for each principle; Armstrong (2001) and papers posted on the website provide details.

General Assessment of Long-Term Polar Bear Population Forecasts

We examined all references cited in the nine US Geological Survey Administrative Reports posted on the Internet. The reports, which included 444 unique references, were Amstrup et al. (2007), Bergen et al. (2007), DeWeaver (2007), Durner et al. (2007), Hunter et al. (2007), Obbard et al. (2007), Regehr et al. (2007), Rode et al. (2007), and Stirling et al. (2007). We were unable to find references to evidence that the forecasting methods described in the reports had been validated.

Forecasting Audit of Key Reports Prepared to Support the Listing of Polar Bears

We audited the forecasting procedures in the reports that we judged provided the strongest support (i.e., forecasts) for listing polar bears. We selected Amstrup et al. (2007), which we will refer to as AMD, because the press had discussed their forecast widely. We selected Hunter et al. (2007), which we will refer to as H6, because the authors used a substantially different approach to the one reported in AMD.

The reports provide forecasts of polar-bear populations for 45, 75, and 100 years from the year 2000 and make recommendations with respect to the polar-bear–listing decision. However, their recommendations do not follow logically from their research because they only make forecasts of the polar bear population. To make policy recommendations based on forecasts, the following assumptions are necessary:

(1) Global warming will occur and will reduce the amount of summer sea ice;

(2) Polar bears will not adapt; thus, they will obtain less food than they do now by hunting from the sea-ice platform;

(3) Listing polar bears as a threatened or endangered species will result in policies that will solve the problem without serious detrimental effects; and

(4) Other policies would be inferior to those that depend on an Endangered Species Act listing.

Regarding the first assumption, both AMD and H6 assumed that general circulation models (GCMs) provide scientifically valid forecasts of global temperature and the extent and thickness of sea ice. AMD stated: "Our future forecasts are based largely on information derived from general circulation model (GCM) projections of the extent and spatiotemporal distribution of sea ice" (AMD: p. 2; p. 83, Figure 2). H6 stated, "We extracted forecasts of the availability of sea ice for polar bears in the Southern Beaufort Sea region, using monthly forecasts of sea-ice concentrations from 10 IPCC Fourth Assessment Report (AR4) fully-coupled general circulation models" (p. 11). (Note: IPCC is the Intergovernmental Panel on Climate Change.) That is, the forecasts of both AMD and H6 are conditional on long-term global warming leading to a dramatic reduction in Arctic sea ice during melt-back periods in spring, late summer, and fall.

Green and Armstrong (2007) examined long-term climate-forecasting efforts and were unable to find a single forecast of global warming that was based on scientific methods. When they audited the GCM climate modelers' procedures, they found that only 13 percent of the relevant forecasting principles were followed properly; some contraventions of principles were critical. Their findings were consistent with earlier cautions. For example, Soon et al. (2001) found that the current generation of GCMs is unable to meaningfully calculate the effects that additional atmospheric carbon dioxide has on the climate. This is because of the uncertainty about the past and present climate and ignorance about relevant weather and climate processes. Some climate modelers state that the GCMs do *not* provide forecasts. According to one of the lead authors of the IPCC's AR4 (Trenberth 2007),

> ...there are no predictions by IPCC at all. And there never have been. The IPCC instead proffers "what if" projections of future climate that correspond to certain emissions scenarios. There are a number of assumptions that go into these emissions scenarios. They are intended to cover a range of possible self consistent "story lines" that then provide decision makers with information about which paths might be more desirable.

AMD and H6 provided no scientific evidence to support their assumptions about any of the four issues that we identified above. Thus, their forecasts are of no value to decision makers. Nevertheless, we audited their polar-bear–population forecasting procedures to assess if they would have produced valid forecasts *if* the underlying assumptions had been valid.

In conducting our audits, we read AMD and H6 and independently rated the forecasting procedures described in the reports by using the forecasting audit software mentioned above. The rating scale ranged from −2 to +2; the former indicated that the procedures contravene the principle; the latter signified that it is properly applied. Following the initial round of ratings, we examined differences in our ratings to reach consensus. When we had difficulty in reaching consensus, we moved ratings toward "0." Principle 1.3 (*Make sure forecasts are independent of politics*) is an example of a principle that was contravened in both reports (indeed, in all nine). By politics, we mean any type of organizational bias or pressure. It is not unusual for different stakeholders to prefer particular forecasts; however, if forecasters are influenced by such considerations, forecast accuracy could suffer. The header on the title page of each of the nine reports suggests how the authors interpreted their task: "USGS Science Strategy to Support US Fish and Wildlife Service Polar Bear Listing Decision." A more neutral statement of purpose might have read "Forecasts of the polar bear population under alternative policy regimes."

While it was easy to code the two reports' procedures against Principle 1.3, the ratings were subjective for many principles. Despite the subjectivity, our ratings after the first round of analyses for each report were substantially in agreement. Furthermore, we readily achieved consensus by the third round.

The two reports did not provide sufficient detail to allow us to rate some of the relevant principles. As a result, we contacted the report authors for additional information. We also asked them to review the ratings that we had made and to provide comments. In their replies, the report authors refused to provide any responses to our requests. (See #2 in the *Author Comments* section at the end of this paper.)

In December 2007, we sent a draft of this article to all authors whose works we cited substantively and asked them to inform us if we had misinterpreted their findings. None objected to our interpretations. We also invited each author to review our paper but received no reviews from our requests.

Audit Findings for AMD

In auditing AMD's forecasting procedures, we first agreed that 24 of the 140 forecasting principles were irrelevant to the forecasting problem they were trying to address. We then examined principles for which our ratings differed. The process involved three rounds of consultation; after two rounds, we were able to reach consensus on ratings against all 116 relevant principles. We were unable to rate AMD's procedures against 26 relevant principles (Table A.3) because the paper lacked the necessary information. Tables A.1, A.2, A.3, and A.4 provide full disclosure of our AMD ratings.

Overall, we found that AMD definitely contravened 41 principles and apparently contravened an additional 32 principles. The authors provided no justifications for the contraventions. Of the 116 relevant principles, we could find evidence that AMD properly applied only 17 (14.7 percent) (Table A.4).

In the remainder of this section, we will describe some of the more serious problems with the AMD forecasting procedures by listing a selected principle and then explaining how AMD addressed it.

Principle 6.7: Match the forecasting method(s) to the situation.

The AMD forecasts rely on the opinions of a single polar bear expert. The report authors transformed these opinions into a complex set of formulae without using evidence-based forecasting principles. In effect, the formulae were no more than a codification of the expert's unaided judgments, which are not appropriate for forecasting in this situation.

One of the most counterintuitive findings in forecasting is that judgmental forecasts by experts who ignore accepted forecasting principles have little value in complex and uncertain situations (Armstrong 1978, pp. 91–96; Tetlock 2005). This finding applies whether the opinions are expressed in words, spreadsheets, or mathematical models. In relation to the latter, Pilkey and Pilkey-Jarvis (2007) provide examples of the failure of domain experts' mathematical models when they are applied to diverse natural science problems including fish stocks, beach engineering, and invasive plants. This finding also applies regardless of the amount and quality of information that the experts use because of the following:

(1) Complexity: People cannot assess complex relationships through unaided observations.

(2) Coincidence: People confuse correlation with causation.

(3) Feedback: People making judgmental predictions typically do not receive unambiguous feedback that they can use to improve their forecasting.

(4) Bias: People have difficulty in obtaining or using evidence that contradicts their initial beliefs. This problem is especially serious among people who view themselves as experts.

Despite the lack of validity of expert unaided forecasts, many public-policy decisions are based on such forecasts. Research on persuasion has shown that people have substantial faith in the value of such forecasts and that faith increases when experts agree with one another. Although they may seem convincing at the time, expert forecasts can, a few years later, serve as important cautionary tales. Cerf and Navasky's (1998) book contains 310 pages of examples of false expert forecasts, such as the Fermi award-winning scientist John von Neumann's 1956 prediction that "A few decades hence, energy may be free." Examples of expert climate forecasts that turned out to be wrong are easy to find, such as UC Davis ecologist Kenneth Watt's prediction during an Earth Day speech at Swarthmore College (April 22, 1970) that "If present trends continue, the world will be about four degrees colder in 1990, but eleven degrees colder in the year 2000. This is about twice what it would take to put us into an ice age."

Tetlock (2005) recruited 284 people whose professions included "commenting or offering advice on political and economic trends." He picked topics (geographic and substantive) both within and outside of their areas of expertise and asked them to forecast the probability that various situations would or would not occur. By 2003, he had accumulated more than 82,000 forecasts. The experts barely, if at all, outperformed nonexperts; neither group did well against simple rules.

Despite the evidence showing that expert forecasts are of no value in complex and uncertain situations, people continue to believe in experts' forecasts. The first author's review of empirical research on this problem led him to develop the "seer-sucker theory," which states that "No matter how much evidence exists that seers do not exist, seers will find suckers" (Armstrong 1980).

Principle 7.3: Be conservative in situations of high uncertainty or instability.

Forecasts should be conservative when a situation is unstable, complex, or uncertain. Being conservative means moving forecasts towards "no change" or, in cases that exhibit a well-established, long-term trend and where there is no reason to expect the trend to change, being conservative means moving forecasts toward the trend line. A long-term trend is one that has been evident over a period that is *much longer* than the period being forecast. Conservatism is a fundamental principle in forecasting.

The interaction between polar bears and their environment in the Arctic is complex and uncertain. For example, AMD associated warmer temperatures with lower polar bear survival rates; yet, as the following quote illustrates, colder temperatures have also been found to be associated with the same outcome: "Abnormally heavy ice covered much of the eastern Beaufort Sea during the winter of 1973–1974. This resulted in major declines in numbers and productivity of polar bears and ringed seals in 1975" (Amstrup et al. 1986, p. 249). Stirling (2002, pp. 68, 72) further expanded on the complexity of polar bear and sea-ice interactions:

> In the eastern Beaufort Sea, in years during and following heavy ice conditions in spring, we found a marked reduction in production of ringed seal pups and consequently in the natality of polar bears.... The effect appeared to last for about three years, after which productivity of both seals and bears increased again. These clear and major reductions in productivity of ringed seals in relation to ice conditions occurred at decadal-scale intervals in the mid-1970s and 1980s... and, on the basis of less complete data, probably in the mid-1960s as well.... Recent analyses of ice anomalies in the Beaufort Sea have now also confirmed the existence of an approximately 10-year cycle in the region... that is roughly in phase with a similar decadal-scale oscillation in the runoff from the Mackenzie River.... However, or whether, these regional-scale changes in ecological conditions have affected the reproduction and survival of young ringed seals and polar bears through the 1990s is not clear.

Regional variability adds to uncertainty. For example, Antarctic ice mass has been increasing while sea and air temperatures have also been increasing (Zhang 2007). At the same time, depth-averaged oceanic temperatures around the Southeastern Bering Sea (Richter-Menge et al. 2007) have been cooling since 2006. Despite the warming of local air temperatures by 1.6 ± 0.6 °C, there was no consistent mid-September (the period of minimal ice extent) ice decline in the Canadian Beaufort Sea over the continental shelf, which had been ice-covered for the 36 years between 1968 and 2003 (Melling et al. 2005).

In their abstract, AMD predicted a loss of "... 2/3 of the world's current polar bear population by mid-century." The 2/3 figure is at odds with the output from the authors' "deterministic model" as they show in Table 6 in their report. The model's "ensemble mean" prediction is for a more modest decline of 17 percent in the polar bear population by the year 2050. Even the GCM minimum ice scenario, which the authors used as an extreme input, provides a forecast decline of 22 percent—much less than the 2/3 figure they state in their abstract. We believe that the authors derived their 2/3 figure informally from the outputs of their Bayesian network modeling exercise. The Bayesian network output of interest is in the form of probabilities (expressed as percentages) for each of five possible population states: "larger," "same as now," "smaller," "rare," and "extinct" (AMD, Table 8, pp. 66–67). There is, however, no clear link between the sets of probabilities for each population state for each of the authors' four Arctic eco-regions and the dramatic 2/3 population-reduction figure.

AMD made predictions based on assumptions that we view as questionable. They used little historical data and extreme forecasts rather than conservative ones.

Principle 8.5: Obtain forecasts from heterogeneous experts.

AMD's polar bear population forecasts were the product of a single expert. Experts vary in their knowledge and in how they approach problems. A willingness to bring additional information and different approaches to bear on a forecasting problem improves accuracy. When researchers use information from a single source only, the validity and reliability of the forecasting process is suspect. In addition, in situations in which experts might be biased, it is important to obtain forecasts from experts with different biases. Failing to follow this principle increases the risk that the forecasts obtained will be extreme when, in this situation, forecasts should be conservative (see Principle 7.3 above).

Principle 10.2: Use all important variables.

Dyck et al. (2007) noted that scenarios of polar bear population decline from changing sea-ice habitat alone grossly oversimplify the complex ecological relationships of the situation. In particular, AMD did not adequately consider the adaptability of polar bears. They mentioned that polar bears evolved from brown bears 250,000 years ago; however, they appear to have underrated the fact that polar bears probably experienced much warmer conditions in the Arctic over that extended period, including periods in which the sea-ice habitat was less than the amount predicted during the 21st century by the GCM projections that AMD used. A dramatic reduction of sea ice in both the northwest Alaskan coast and northwest Greenland part of the Arctic Ocean during the very warm interglacial of marine isotope stage 5e *ca*. 130,000 to 120,000 years ago was documented by Hamilton and Brigham-Grette (1991), Brigham-Grette and Hopkins (1995), and Norgaard-Pedersen et al. (2007). Brigham-Grette and Hopkins (1995, p. 159) noted that the "winter sea-ice limit was north of Bering Strait, at least 800 km north of its present position, and the Bering Sea was perennially ice-free" and that "[the more saline] Atlantic water may have been present on the shallow Beaufort Shelf, suggesting that the Arctic Ocean was not stratified and the Arctic sea-ice cover was not perennial for some period." The nature and extent of polar bear adaptability seem crucial to any forecasts that assume dramatic changes in the bears' environment.

Audit Findings for H6

H6 forecast polar bear numbers and their survival probabilities in the Southern Beaufort Sea for the 21st century.

Of the 140 forecasting principles, we agreed that 35 were irrelevant to the forecasting problem. We found that H6's procedures clearly contravened 61 principles (Table A.5) and probably contravened an additional 19 principles (Table A.6). We were unable to rate H6's procedures against 15 relevant principles (Table A.7) because of a lack of information. Perhaps the best way to summarize H6's efforts is to say that the authors properly applied only 10 (9.5 percent) of the 105 relevant principles (Table A.8).

Many of the contraventions in H6 were similar to those in AMD. We describe some of the more serious problems with the H6 forecasting procedures by examining their contraventions of 13 important principles that differed from contraventions discussed in AMD.

Principles 1.1–1.3: Decisions, actions, and biases.

The H6 authors did not describe alternative decisions that might be taken (as Principle 1.1 requires), nor did they propose relationships between possible forecasts and alternative decisions (as Principle 1.2 requires). For example, what decision would be implied by a forecast that predicts that bear numbers will increase to where they become a threat to existing human settlements?

Principle 4.2: Ensure that information is reliable and that measurement error is low.

H6 relied heavily on five years of data with unknown measurement errors. Furthermore, we question whether the capture data on which they relied provide representative samples of bears in the Southern Beaufort Sea given the vast area involved and difficulties in spotting and capturing the bears. Bears wander over long distances and do not respect administrative boundaries (Amstrup et al. 2004). The validity of the data was also compromised because H6 imposed a speculative demographic model on the raw capture-recapture data (Amstrup et al. 2001, Regehr et al. 2006).

Principle 4.4: Obtain all important data.

H6 estimated their key relationship—between ice-free days and the polar bear population—by using data that appear to be unreliable primarily because of the difficulty of estimating the polar bear population, but also because of the measurements of ice. Experts in this field, including the authors of the nine reports, are aware of these problems. In addition, they rely on only five years of data with a limited range of climate and ecology combinations. They might, for example, have independently estimated the magnitude of the relationship by obtaining estimates of polar bear populations during much warmer and much colder periods in the past. The supplementary information in Regehr et al. (2007, Figure 3) shows that 1987, 1993, and 1998 were exceptional seasons with more than 150 ice-free days (i.e., substantially above the 135 ice-free days documented for 2004–2005) in the Southern Beaufort Sea. Yet, there were no apparent negative impacts on the polar bear population and well-being (Amstrup et al. 2001).

Because they used only five observations, the above points are moot. It is impossible to estimate a causal relationship in a complex and uncertain situation by using only five data points.

Principle 7.3: Be conservative in situations of high uncertainty or instability.

The situation regarding polar bears in the Southern Beaufort Sea is complex and uncertain. On the basis of five years of data, H6 associated warmer temperatures (and hence more ice-free days) with lower polar bear survival rates. Yet, as we noted in relation to AMD, cold temperatures have also been found to be associated with the same outcome. In addition, regional variability (e.g., sea ice increases while sea and air temperatures increase) adds to uncertainty.

There is general agreement that polar bear populations have increased or remained stable in the Alaska regions in recent decades (Amstrup et al. 1995, Angliss and Outlaw 2007). H6 assumed that there are downward forces that will cause the trend to reverse. However, studies in economics have shown little success in predicting turning points. Indeed, Armstrong and Collopy (1993) proposed the principle that one should not extrapolate trends if they are contrary to the direction of the causal forces as judged by domain experts. They tested the principle on four data sets involving 723 long-range forecasts and found that it reduced forecast error by 43 percent. Therefore, even if one had good reason to expect a trend to reverse, being conservative and avoiding the extrapolation of *any* trend will increase the accuracy of forecasts.

Principle 9.2: Match the model to the underlying phenomena.

Because of the poor spatial resolution of the GCMs, it is important that readers know the meaning of the "Southern Beaufort Sea" (SB) in the H6 report. H6 states:

> Because GCMs do not provide suitable forecasts for areas as small as the SB, we used sea ice concentration for a larger area composed of 5 IUCN (International Union for Conservation of Nature) polar bear management units (Aars et al. 2006) with ice dynamics similar to the SB management unit (Barents Sea, Beaufort Sea, Chukchi Sea, Kara Sea, and Laptev Sea; see Rigor and Wallace 2004, Durner et al. 2007). We assumed that the general trend in sea ice availability in these 5 units was representative of the general trend in the Southern Beaufort region. (p. 12).

Given the unique ecological, geographical, meteorological, and climatological conditions in each of the five circumpolar seas, this assumption by H6 is not valid or convincing.

Principle 9.5: Update frequently.

When they estimated their model, H6 did not include data for 2006, the most recent year that was then available. From the supplementary information that Regehr et al. (2007, Figure 3) provide, one finds that the number of ice-free days for the 2006 season was approximately 105—close to the mean of the "good" ice years.

Principle 10.2: Use all important variables.

When using causal models, it is important to incorporate policy variables if they might vary or if the purpose is to decide which policy to implement. H6 did not include policy variables, such as seasonal protection of bears' critical habitat or changes to hunting rules.

Other variables, such as migration, snow, and wind conditions, should also be included. For example, Holloway and Sou (2002), Ogi and Wallace (2007), and Nghiem et al. (2007) suggested that large-scale atmospheric winds and related circulatory and warming and cooling patterns play an important role in causing—in some situations with substantial time delays—both the decline in extent and thinning of Arctic sea ice. The GCM forecasts of sea ice did not correctly include those effects; hence, the forecasts of the quality of the polar bear habitats also did not.

In addition, as Dyck et al. (2007) noted, forecasts of polar bear decline because of dramatic changes in their environment do not take proper account of the extent and type of polar bear adaptability.

Principle 10.5: Use different types of data to measure a relationship.

This principle is important when there is uncertainty about the relationships between causal variables (such as ice extent) and the variable being forecast (polar bear population), and when large changes are expected in the causal variables. In the case of the latter condition, H6 accepted the GCM model predictions of large declines in summer ice throughout the 21st century.

Principle 10.7: Forecast for alternate interventions.

H6 did not explicitly forecast the effects of different policies. For example, if the polar bear population came under stress because of inadequate summer food, what would be the costs and benefits of protecting areas by prohibiting marine and land-based activities, such as tourism, capture for research, and hunting at critical times? In addition, what would be the costs and benefits of a smaller but stable population of polar bears in some polar subregions? And how would the net costs of such alternative policies compare with the net costs of listing polar bears?

Principle 13.8: Provide easy access to the data.

The authors of the reports that we audited did not include all of the data they used in their reports. We requested the missing data, but they did not provide it.

Principle 14.7: When assessing prediction intervals, list possible outcomes, and assess their likelihoods.

To assess meaningful prediction intervals, it is helpful to think of diverse possible outcomes. The H6 authors did not appear to consider, for example, the possibility that polar bears might adapt to terrestrial life over summer months by finding alternative food sources (Stempniewicz 2006, Dyck and Romberg 2007) or by successfully congregating in smaller or localized ice-hunting areas. Consideration of these and other possible adaptations and outcomes would have likely led the H6 authors to be less confident (e.g., provide wider prediction intervals) about the outcome for the bear population. Extending this exercise to the forecasts of climate and summer ice extent would have further widened the range of possible outcomes.

Discussion

Rather than relying on untested procedures to forecast polar bear populations, the most appropriate approach would be to rely upon prior evidence of which forecasting methods work best under which conditions. Thus, one could turn to empirical evidence drawn from a wide variety of forecasting problems. This evidence is summarized in the Forecasting Method Selection Tree at http://forecastingprinciples.com.

Armstrong (1985) provided an early review of the evidence on how to forecast given high uncertainty. Schnaars (1984) and Schnaars and Bavuso (1986) concluded that the random walk was typically the most accurate model in their comparative studies of hundreds of economic series with forecast horizons of up to five years. This principle has a long history. For example, regression models "regress" towards a no-change forecast when the estimates of causal relationships are uncertain.

Because of the enormous uncertainty involved in long-term forecasts of polar bear populations, the lack of accurate time-series data on these populations, and the complex relationships that are subject to much uncertainty, prior evidence from forecasting research calls for simple and conservative methods. Therefore, one should follow a trend if such a trend is consistent and if there are no strong reasons to expect a change in the trend. Even then, however, it is wise to dampen the trend towards zero given the increasing uncertainty as the forecast horizon is extended. Empirical evidence supports this notion of "damping trends" (Armstrong 2001). Lacking a trend, forecasters should turn to the so-called "random walk" or no-change model. Given the upward trend in polar bear numbers over the past few decades, a modest upward trend is likely to continue in the near future because the apparent cause of the trend (hunting restrictions) remains. However, the inconsistent long-term trends in the polar bear population suggest that it is best to assume no trend in the long-term.

Summary

We inspected nine administrative reports that the US government commissioned. Because the current polar bear population is not at a level that is causing concern, the case for listing depends upon forecasts of serious declines in bear numbers in future decades. None of these reports included references to scientific works on forecasting methods.

We found that the two reports that we judged most relevant to the listing decision made assumptions rather than forecasts. Even if these assumptions had been valid, the bear population forecasting procedures described in the reports contravened many important forecasting principles. We did forecasting audits of the two key reports (Table 1).

Decision makers and the public should require scientific forecasts of both the polar bear population *and* the costs and benefits of alternative policies before making a decision on whether to list polar bears as threatened or endangered. We recommend that important forecasting efforts such as this should properly apply all relevant principles and that their procedures be audited to ensure that they do so. Failure to apply any principle should be supported by evidence that the principle was not applicable.

Principles	AMD	H6
Contravened	41	61
Apparently contravened	32	19
Not auditable	26	15
Properly applied	17	10
Totals	116	105

Table 1: We summarize our forecasting audit ratings of the AMD and H6 reports against relevant forecasting principles.

Author Comments

1. Our interest in the topic of this paper was piqued when the State of Alaska hired us as consultants in late September 2007 to assess forecasts that had been prepared "to Support US Fish and Wildlife Service Polar Bear Listing Decision." We received $9,998 as payment for our consulting. We were impressed by the importance of the issue; therefore, after providing our assessment, we decided to continue work on it and to prepare a paper for publication. These latter efforts have not been funded. We take responsibility for all judgments and for any errors that we might have made.

2. On November 27, 2007, we sent a draft of our paper to the authors of the US Geological Survey administrative reports that we audited; it stated:

> As we note in our paper, there are elements of subjectivity in making the audit ratings. Should you feel that any of our ratings were incorrect, we would be grateful if you would provide us with evidence that would lead to a different assessment. The same goes for any principle that you think does not apply, or to any principles that we might have overlooked. There are some areas that we could not rate due to a lack of information. Should you have information on those topics, we would be interested. Finally, we would be interested in peer review that you or your colleagues could provide, and in suggestions on how to improve the accuracy and clarity of our paper.

We received this reply from Steven C. Amstrup on November 30, 2007: "We all decline to offer preview comments on your attached manuscript. Please feel free, however, to list any of us as potential referees when you submit your manuscript for publication."

3. We invite others to conduct forecasting audits of Amstrup et al., Hunter et al., or any of the other papers prepared to support the endangered-species listing, or any other papers relevant to long-term forecasting of the polar bear population. Note that the audit process calls for two or more raters. The audits can be submitted for publication on pubicpolicyforecasting.com with the auditors' bios and any information relevant, potential sources of bias.

Appendix: Full Disclosure of the Codings

Table A.1: Principles contravened in Amstrup et al. (AMD).

Setting objectives
- 1.2 Prior to forecasting, agree on actions to take assuming different possible forecasts.
- 1.3 Make sure forecasts are independent of politics.
- 1.4 Consider whether the events or series can be forecasted.
- 1.5 Obtain decision makers' agreement on methods.

Identifying data sources
- 3.5 Obtain information from similar (analogous) series or cases. Such information may help to estimate trends.

Collecting data
- 4.2 Ensure that information is reliable and that measurement error is low.

Selecting methods
- 6.1 List all the important selection criteria before evaluating methods.
- 6.2 Ask unbiased experts to rate potential methods.
- 6.7 Match the forecasting method(s) to the situation.
- 6.8 Compare track records of various forecasting methods.
- 6.10 Examine the value of alternative forecasting methods.

Implementing methods: General
- 7.3 Be conservative in situations of high uncertainty or instability.

Implementing judgmental methods
- 8.1 Pretest the questions you intend to use to elicit judgmental forecasts.
- 8.2 Frame questions in alternative ways.
- 8.5 Obtain forecasts from heterogeneous experts.
- 8.7 Obtain forecasts from enough respondents.
- 8.8 Obtain multiple forecasts of an event from each expert.

Implementing quantitative methods
- 9.1 Tailor the forecasting model to the horizon.
- 9.3 Do not use "fit" to develop the model.
- 9.5 Update models frequently.

Implementing methods: Quantitative models with explanatory variables
- 10.6 Prepare forecasts for at least two alternative environments.
- 10.8 Apply the same principles to forecasts of explanatory variables.
- 10.9 Shrink the forecasts of change if there is high uncertainty for predictions of the explanatory variables.

Combining forecasts
- 12.1 Combine forecasts from approaches that differ.
- 12.2 Use many approaches (or forecasters), preferably at least five.
- 12.3 Use formal procedures to combine forecasts.
- 12.4 Start with equal weights.

Evaluating methods
- 13.6 Describe potential biases of forecasters.
- 13.10 Test assumptions for validity.
- 13.32 Conduct explicit cost-benefit analyses.

Assessing uncertainty
- 14.1 Estimate prediction intervals (PIs).
- 14.2 Use objective procedures to estimate explicit prediction intervals.
- 14.3 Develop prediction intervals by using empirical estimates based on realistic representations of forecasting situations.
- 14.5 Ensure consistency over the forecast horizon.
- 14.7 When assessing PIs, list possible outcomes and assess their likelihoods.
- 14.8 Obtain good feedback about forecast accuracy and the reasons why errors occurred.
- 14.9 Combine prediction intervals from alternative forecasting methods.
- 14.10 Use safety factors to adjust for overconfidence in the PIs.
- 14.11 Conduct experiments to evaluate forecasts.
- 14.13 Incorporate the uncertainty associated with the prediction of the explanatory variables in the prediction intervals.
- 14.14 Ask for a judgmental likelihood that a forecast will fall within a predefined minimum-maximum interval.

Table A.2: Principles apparently contravened in AMD.

Structuring the problem
- 2.1 Identify possible outcomes prior to making forecasts.
- 2.7 Decompose time series by level and trend.

Identifying data sources
- 3.2 Ensure that the data match the forecasting situation.
- 3.3 Avoid biased data sources.
- 3.4 Use diverse sources of data.

Collecting data
- 4.1 Use unbiased and systematic procedures to collect data.
- 4.3 Ensure that the information is valid.

Selecting methods
- 6.4 Use quantitative methods rather than qualitative methods.
- 6.9 Assess acceptability and understandability of methods to users.

Implementing methods: General
- 7.1 Keep forecasting methods simple.

Implementing quantitative methods
- 9.2 Match the model to the underlying phenomena.
- 9.4 Weight the most relevant data more heavily.

Implementing methods: Quantitative models with explanatory variables
- 10.1 Rely on theory and domain expertise to select causal (or explanatory) variables.
- 10.2 Use all important variables.
- 10.5 Use different types of data to measure a relationship.

Combining forecasts
- 12.5 Use trimmed means, medians, or modes.
- 12.7 Use domain knowledge to vary weights on component forecasts.
- 12.8 Combine forecasts when there is uncertainty about which method is best.
- 12.9 Combine forecasts when you are uncertain about the situation.
- 12.10 Combine forecasts when it is important to avoid large errors.

Evaluating methods
- 13.1 Compare reasonable methods.
- 13.2 Use objective tests of assumptions.
- 13.7 Assess the reliability and validity of the data.
- 13.8 Provide easy access to the data.

13.17 Examine all important criteria.
13.18 Specify criteria for evaluating methods prior to analyzing data.
13.27 Use ex post error measures to evaluate the effects of policy variables.

Assessing uncertainty
14.6 Describe reasons why the forecasts might be wrong.

Presenting forecasts
15.1 Present forecasts and supporting data in a simple and understandable form.
15.4 Present prediction intervals.

Learning to improve forecasting procedures
16.2 Seek feedback about forecasts.
16.3 Establish a formal review process for forecasting methods.

Table A.3: Principles not rated because of lack of information in AMD.

Structuring the problem
2.5 Structure problems to deal with important interactions among causal variables.

Collecting data
4.4 Obtain all of the important data.
4.5 Avoid the collection of irrelevant data.

Preparing data
5.1 Clean the data.
5.2 Use transformations as required by expectations.
5.3 Adjust intermittent series.
5.4 Adjust for unsystematic past events.
5.5 Adjust for systematic events.
5.6 Use multiplicative seasonal factors for trended series when you can obtain good estimates for seasonal factors.
5.7 Damp seasonal factors for uncertainty.

Selecting methods
6.6 Select simple methods unless empirical evidence calls for a more complex approach.

Implementing methods: General
7.2 The forecasting method should provide a realistic representation of the situation.

Implementing judgmental methods
8.4 Provide numerical scales with several categories for experts' answers.

Implementing methods: Quantitative models with explanatory variables
10.3 Rely on theory and domain expertise when specifying directions of relationships.
10.4 Use theory and domain expertise to estimate or limit the magnitude of relationships.

Integrating judgmental and quantitative methods
11.1 Use structured procedures to integrate judgmental and quantitative methods.
11.2 Use structured judgment as inputs to quantitative models.
11.3 Use prespecified domain knowledge in selecting, weighting, and modifying quantitative methods.
11.4 Limit subjective adjustments of quantitative forecasts.

Evaluating methods
13.4 Describe conditions associated with the forecasting problem.
13.5 Tailor the analysis to the decision.
13.9 Provide full disclosure of methods.
13.11 Test the client's understanding of the methods.
13.19 Assess face validity.

Assessing uncertainty
14.12 Do not assess uncertainty in a traditional (unstructured) group meeting.

Learning to improve forecasting procedures
16.4 Establish a formal review process to ensure that forecasts are used properly.

Table A.4: Principles properly applied or *apparently properly applied (italics)* in AMD.

Setting objectives
1.1 Describe decisions that might be affected by the forecasts.

Structuring the problem
2.2 *Tailor the level of data aggregation (or segmentation) to the decisions*.
2.3 *Decompose the problem into parts*.
2.6 *Structure problems that involve causal chains*.

Identifying data sources
3.1 Use theory to guide the search for information on explanatory variables.

Collecting data
4.6 *Obtain the most recent data*.

Preparing data
5.8 *Use graphical displays for data*.

Selecting methods
6.3 Use structured rather than unstructured forecasting methods.
6.5 *Use causal methods rather than naive methods if feasible*.

Implementing methods: General
7.5 *Adjust for events expected in the future*.
7.6 *Pool similar types of data*.
7.7 *Ensure consistency with forecasts of related series and related time periods*.

Implementing judgmental methods
8.3 *Ask experts to justify their forecasts in writing*.

Implementing methods: Quantitative models with explanatory variables
10.7 Forecast for alternate interventions.

Presenting forecasts
15.2 Provide complete, simple, and clear explanations of methods.
15.3 Describe your assumptions.

Learning to improve forecasting procedures
16.1 Consider the use of adaptive forecasting models.

Table A.5: Principles contravened in Hunter et al. (H6).

Setting objectives
- 1.3 Make sure forecasts are independent of politics.
- 1.4 Consider whether the events or series can be forecasted.

Structuring the problem
- 2.6 Structure problems that involve causal chains.

Identifying data sources
- 3.4 Use diverse sources of data.
- 3.5 Obtain information from similar (analogous) series or cases. Such information may help to estimate trends.

Collecting data
- 4.4 Obtain all of the important data.

Preparing data
- 5.2 Use transformations as required by expectations.
- 5.4 Adjust for unsystematic past events.
- 5.5 Adjust for systematic events.

Selecting methods
- 6.1 List all the important selection criteria before evaluating methods.
- 6.2 Ask unbiased experts to rate potential methods.
- 6.6 Select simple methods unless empirical evidence calls for a more complex approach.
- 6.7 Match the forecasting method(s) to the situation.
- 6.8 Compare track records of various forecasting methods.
- 6.10 Examine the value of alternative forecasting methods.

Implementing methods: General
- 7.1 Keep forecasting methods simple.
- 7.2 The forecasting method should provide a realistic representation of the situation.
- 7.3 Be conservative in situations of high uncertainty or instability.
- 7.4 Do not forecast cycles.

Implementing quantitative methods
- 9.1 Tailor the forecasting model to the horizon.
- 9.2 Match the model to the underlying phenomena.
- 9.3 Do not use “fit” to develop the model.
- 9.5 Update models frequently.

Implementing methods: Quantitative models with explanatory variables:
- 10.2 Use all important variables.
- 10.5 Use different types of data to measure a relationship.
- 10.7 Forecast for alternate interventions.
- 10.9 Shrink the forecasts of change if there is high uncertainty for predictions of the explanatory variables.

Integrating judgmental and quantitative methods
- 11.1 Use structured procedures to integrate judgmental and quantitative methods.
- 11.2 Use structured judgment as inputs to quantitative models.
- 11.3 Use prespecified domain knowledge in selecting, weighting, and modifying quantitative methods.

Combining forecasts
- 12.1 Combine forecasts from approaches that differ.
- 12.2 Use many approaches (or forecasters), preferably at least five.
- 12.3 Use formal procedures to combine forecasts.
- 12.8 Combine forecasts when there is uncertainty about which method is best.
- 12.9 Combine forecasts when you are uncertain about the situation.
- 12.10 Combine forecasts when it is important to avoid large errors.

Evaluating methods
- 13.1 Compare reasonable methods.
- 13.2 Use objective tests of assumptions.
- 13.3 Design test situations to match the forecasting problem.
- 13.5 Tailor the analysis to the decision.
- 13.6 Describe potential biases of forecasters.
- 13.7 Assess the reliability and validity of the data.
- 13.8 Provide easy access to the data.
- 13.10 Test assumptions for validity.
- 13.12 Use direct replications of evaluations to identify mistakes.
- 13.13 Replicate forecast evaluations to assess their reliability.
- 13.16 Compare forecasts generated by different methods.
- 13.17 Examine all important criteria.
- 13.18 Specify criteria for evaluating methods prior to analyzing data.
- 13.26 Use out-of-sample (ex ante) error measures.
- 13.27 Use ex post error measures to evaluate the effects of policy variables.
- 13.31 Base comparisons of methods on large samples of forecasts.

Assessing uncertainty
- 14.3 Develop prediction intervals by using empirical estimates based on realistic representations of forecasting situations.
- 14.5 Ensure consistency over the forecast horizon.
- 14.9 Combine prediction intervals from alternative forecasting methods.
- 14.10 Use safety factors to adjust for overconfidence in the PIs.
- 14.11 Conduct experiments to evaluate forecasts.
- 14.13 Incorporate the uncertainty associated with the prediction of the explanatory variables in the prediction intervals.
- 14.14 Ask for a judgmental likelihood that a forecast will fall within a predefined minimum-maximum interval (not by asking people to set upper and lower confidence levels).

Presenting forecasts
- 15.1 Present forecasts and supporting data in a simple and understandable form.
- 15.2 Provide complete, simple, and clear explanations of methods.

Table A.6: Principles apparently contravened in H6.

Setting objectives
- 1.1 Describe decisions that might be affected by the forecasts.
- 1.2 Prior to forecasting, agree on actions to take assuming different possible forecasts.

Structuring the problem
- 2.1 Identify possible outcomes prior to making forecasts.
- 2.3 Decompose the problem into parts.

Identifying data sources
- 3.2 Ensure that the data match the forecasting situation.
- 3.3 Avoid biased data sources.

Collecting data
- 4.2 Ensure that information is reliable and that measurement error is low.
- 4.3 Ensure that the information is valid.

Preparing data
- 5.3 Adjust intermittent series.
- 5.7 Damp seasonal factors for uncertainty.
- 5.8 Use graphical displays for data.

Implementing methods: General
- 7.6 Pool similar types of data.

Implementing methods: Quantitative models with explanatory variables
- 10.4 Use theory and domain expertise to estimate or limit the magnitude of relationships.
- 10.8 Apply the same principles to forecasts of explanatory variables.

Evaluating methods
- 13.4 Describe conditions associated with the forecasting problem.
- 13.9 Provide full disclosure of methods.

Assessing uncertainty
- 14.6 Describe reasons why the forecasts might be wrong.
- 14.7 When assessing PIs, list possible outcomes and assess their likelihoods.
- 14.8 Obtain good feedback about forecast accuracy and the reasons why errors occurred.

Table A.7: Principles not rated because of lack of information in H6.

Setting objectives
- 1.5 Obtain decision makers' agreement on methods.

Structuring the problem
- 2.7 Decompose time series by level and trend.

Identifying data sources
- 3.1 Use theory to guide the search for information on explanatory variables.

Collecting data
- 4.1 Use unbiased and systematic procedures to collect data.
- 4.5 Avoid the collection of irrelevant data.

Preparing data
- 5.1 Clean the data.

Selecting methods
- 6.4 Use quantitative methods rather than qualitative methods.
- 6.5 Use causal methods rather than naive methods if feasible.
- 6.9 Assess acceptability and understandability of methods to users.

Evaluating methods
- 13.11 Test the client's understanding of the methods.
- 13.19 Assess face validity.

Presenting forecasts
- 15.3 Describe your assumptions.

Learning to improve forecasting procedures
- 16.2 Seek feedback about forecasts.
- 16.3 Establish a formal review process for forecasting methods.
- 16.4 Establish a formal review process to ensure that forecasts are used properly.

Table A.8: Principles properly applied or *apparently properly applied* in H6.

Structuring the problem
- 2.2 *Tailor the level of data aggregation (or segmentation) to the decisions.*

Collecting data
- 4.6 *Obtain the most recent data.*

Selecting methods
- 6.3 Use structured rather than unstructured forecasting methods.

Implementing methods: Quantitative models with explanatory variables
- 10.1 Rely on theory and domain expertise to select causal (or explanatory) variables.
- 10.3 *Rely on theory and domain expertise when specifying directions of relationships.*
- 10.6 *Prepare forecasts for at least two alternative environments.*

Assessing uncertainty
- 14.1 *Estimate prediction intervals (PIs).*
- 14.2 *Use objective procedures to estimate explicit prediction intervals.*

Presenting forecasts
- 15.4 Present prediction intervals.
- 15.5 *Present forecasts as scenarios.*

Acknowledgments

We thank Don Esslemont, Milton Freeman, Paul Goodwin, Benny Peiser, Orrin Pilkey, Tom Stewart, Mitchell Taylor, and two anonymous reviewers for their comments on earlier drafts. Janice Dow and Kelly Jin provided editorial assistance.

References

Amstrup, S. C., G. W. Garner, G. M. Durner. 1995. Polar bears in Alaska. E. T. La Roe, ed. *Our Living Resources: A Report to the Nation on the Abundance, Distributions, and Health of the US Plants, Animals, and Ecosystems*. US Department of the Interior–National Biological Sciences, Washington, D.C., 351–353.

Amstrup, S. C., B. G. Marcot, D. C. Douglas. 2007. Forecasting the rangewide status of polar bears at selected times in the 21st century. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Amstrup, S. C., T. L. McDonald, G. M. Durner. 2004. Using satellite radiotelemetry data to delineate and manage wildlife populations. *Wildlife Soc. Bull.* **32** 661–679.

Amstrup, S. C., T. L. McDonald, I. Stirling. 2001. Polar bears in the Beaufort Sea: A 30-year mark-recapture case history. *J. Agricultural Biol. Environ. Statist.* **6** 221–234.

Amstrup, S. C., I. Stirling, J. W. Lentfer. 1986. Past and present status of polar bears in Alaska. *Wildlife Soc. Bull.* **14** 241–254.

Angliss, R. P., R. B. Outlaw. 2007. Alaska marine mammal stock assessments, 2006. Retrieved June 4, 2008, http://www.nmfs.noaa.gov/pr/pdfs/sars/ak2006.pdf.

Armstrong, J. S. 1978. *Long-Range Forecasting: From Crystal Ball to Computer*. Wiley-Interscience, New York.

Armstrong, J. S. 1980. The seer-sucker theory: The value of experts in forecasting. *Tech. Rev.* **83**(June–July) 16–24.

Armstrong, J. S. 1985. *Long-Range Forecasting*: *From Crystal Ball to Computer*, 2nd ed. Wiley-Interscience, New York.

Armstrong, J. S. 2001. *Principles of Forecasting*: *A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA.

Armstrong, J. S., F. Collopy. 1993. Causal forces: Structuring knowledge for time-series extrapolation. *J. Forecasting* **12** 103–115.

Bergen, S., G. M. Durner, D. C. Douglas, S. C. Amstrup. 2007. Predicting movements of female polar bears between summer sea ice foraging habitats and terrestrial denning habitats of Alaska in the 21st century: Proposed methodology and pilot assessment. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Brigham-Grette, J., D. M. Hopkins. 1995. Emergent marine record and paleoclimate of the last interglaciation along the Northwest Alaskan coast. *Quaternary Res.* **43** 159–173.

Cerf, C., V. Navasky. 1998. *The Experts Speak*. Random House, New York.

DeWeaver, E. 2007. Uncertainty in climate model projections of Arctic sea ice decline: An evaluation relevant to polar bears. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Durner, G. M., D. C. Douglas, R. M. Nielson, S. C. Amstrup, T. L. McDonald. 2007. Predicting the future distribution of polar bear habitat in the polar basin from resource selection functions applied to 21st century general circulation model projections of sea ice. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Dyck, M. G., S. Romberg. 2007. Observations of a wild polar bear (*Ursus maritimus*) successfully fishing Arctic charr (*Salvelinus alpinus*) and Fourhorn sculpin (*Myoxocephalus quadricornis*). *Polar Biol.* **30** 1625–1628.

Dyck, M. G., W. Soon, R. K. Baydack, D. R. Legates, S. Baliunas, T. F. Ball, L. O. Hancock. 2007. Polar bears of western Hudson Bay and climate change: Are warming spring air temperatures the "ultimate" survival control factor? *Ecological Complexity* **4** 73–84.

Green, K. C., J. S. Armstrong. 2007. Global warming: Forecasts by scientists versus scientific forecasts. *Energy Environ.* **18** 997–1021.

Hamilton, T. D., J. Brigham-Grette. 1991. The last interglaciation in Alaska: Stratigraphy and paleoecology of potential sites. *Quaternary Internat.* 10–12, 49–71.

Holloway, G., T. Sou. 2002. Has Arctic sea ice rapidly thinned? *J. Climate* **15** 1691–1701.

Hunter, C. M., H. Caswell, M. C. Runge, S. C. Amstrup, E. V. Regehr, I. Stirling. 2007. Polar bears in the Southern Beaufort Sea II: Demography and population growth in relation to sea ice conditions. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Melling, H., D. A. Riedel, Z. Gedalof. 2005. Trends in the draft and extent of seasonal pack ice, Canadian Beaufort Sea. *Geophysical Res. Lett.* **32**(24) L24501, doi:10.1029/2005GL024483.

Nghiem, S. V., I. G. Rigor, D. K. Perovich, P. Clemente-Colon, J. W. Weatherly, G. Neumann. 2007. Rapid reduction of Arctic perennial sea ice. *Geophysical Res. Lett.* **34** L19504, doi:10.1029/2007GL031138.

Norgaard-Pedersen, N., N. Mikkelsen, S. J. Lassen, Y. Kristoffersen, E. Sheldon. 2007. Reduced sea ice concentrations in the Arctic Ocean during the last interglacial period revealed by sediment cores off northern Greenland. *Paleoceanography* **22** PA1218, doi:10.1029/2006PA001283.

Obbard, M. E., T. L. McDonald, E. J. Howe, E. V. Regehr, E. S. Richardson. 2007. Trends in abundance and survival for polar bears from Southern Hudson Bay, Canada, 1984–2005. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Ogi, M., J. M. Wallace. 2007. Summer minimum Arctic sea ice extent and the associated summer atmospheric circulation. *Geophysical Res. Lett.* **34** L12705, doi:10.1029/2007GL029897.

Pilkey, O. H., L. Pilkey-Jarvis. 2007. *Useless Arithmetic*. Columbia University Press, New York.

Prestrud, P., I. Stirling. 1994. The international polar bear agreement and the current status of polar bear conservation. *Aquatic Mammals* **20** 113–124.

Regehr, E. V., S. C. Amstrup, I. Stirling. 2006. Polar bear population status in the Southern Beaufort Sea. Retrieved November 17, 2006, http://pubs.usgs.gov/of/2006/1337/pdf/ofr20061337.pdf.

Regehr, E. V., C. M. Hunter, H. Caswell, S. C. Amstrup, I. Stirling. 2007. Polar bears in the Southern Beaufort Sea I: Survival and breeding in relation to sea ice conditions, 2001–2006. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Richter-Menge, J., J. Overland, A. Proshutinsky, V. Romanovsky, R. Armstrong, J. Morison, S. Nghiem, et al. 2007. State of the climate in 2006: Arctic. *Bull. Amer. Meteorological Soc.* **88** S62–S71.

Rode, K. D., S. C. Amstrup, E. V. Regehr. 2007. Polar bears in the Southern Beaufort Sea III: Stature, mass, and cub recruitment in relationship to time and sea ice extent between 1982 and 2006. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Schnaars, S. P. 1984. Situational factors affecting forecasting accuracy. *J. Marketing Res.* **21** 290–297.

Schnaars, S. P., R. J. Bavuso. 1986. Extrapolation models on very short-term forecasts. *J. Bus. Res.* **14** 27–36.

Soon, W., S. Baliunas, S. B. Idso, K. Y. Kondratyev, E. S. Posmentier. 2001. Modeling climatic effects of anthropogenic carbon dioxide emissions: Unknowns and uncertainties. *Climate Res.* **18** 259–275.

Stempniewicz, L. 2006. Polar bear predatory behavior toward moulting barnacle geese and nesting glaucous gulls on Spitsbergen. *Arctic* **59** 247–251.

Stirling, I. 2002. Polar bears and seals in the Eastern Beaufort Sea and Amundsen Gulf: A synthesis of population trends and ecological relationships over three decades. *Arctic* **55** (Suppl. 1) 59–76.

Stirling, I., T. L. McDonald, E. S. Richardson, E. V. Regehr. 2007. Polar bear population status in the Northern Beaufort Sea. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Tetlock, P. E. 2005. *Expert Political Judgment*: *How Good Is It? How Can We Know?* Princeton University Press, Princeton, NJ.

Trenberth, K. 2007. Predictions of climate. Retrieved June 2, 2008, http://blogs.nature.com/climatefeedback/2007/06/predictions_of_climate.html.

Zhang, J. 2007. Increasing Antarctic sea ice under warming atmospheric and oceanic conditions. *J. Climate* **20** 2515–2529.

A Commentary on Polar Bear Population Forecasts: A Public-Policy Forecasting Audit

James J. Cochran
Department of Marketing and Analysis, College of Business, Louisiana Tech University, Ruston, Louisiana 71272,
jcochran@latech.edu

In February 2005, the Center for Biological Diversity (CBD) filed a petition (Center for Biological Diversity 2005) with the US Fish and Wildlife Service (USFWS) requesting that the polar bear (*ursus maritimus*) be listed as a threatened species under the Endangered Species Act (http://www.fws.gov/laws/lawsdigest/ESACT.HTML). In response, USFWS Director Dale Hall formally proposed listing the polar bear as threatened based on the species' threatened habitat (receding sea ice). Hall cited nine studies (Amstrup et al. 2007, Bergen et al. 2007, DeWeaver 2007, Durner et al. 2007, Hunter et al. 2007, Obbard et al. 2007, Regehr et al. 2007, Rode et al. 2007, and Stirling et al. 2007), which were conducted under the auspices of the US Geological Survey (USGS) to broaden the understanding of the species' current circumstances.

A contentious debate ensued. Supporters of the petition accused the USFWS of delaying the decision to avoid obstructing the sales of oil and gas leases in the Chukchi Sea. In a *Newsweek* article (Adler 2008), USFWS Director Dale Hall denied these allegations, stating with respect to the studies of the polar bears' plight that, "This is cutting edge science. I needed time to understand it."

Opponents of the petition accused the petitioners and their supporters of using their petition to surreptitiously impede the sales of Chukshi Sea oil and gas leases and force the United States (through ESA restrictions on activities in the species' habitat that would result) to adopt a stronger response to global warming. Kassie Siegel, the environmental attorney who filed the original petition on behalf of the CBD, maintains that the goal of the petition is solely to protect polar bears, but that the protection of endangered or threatened species and protection of these species' habitats are inseparable issues.

In May 2008, US Secretary of the Interior Dirk Kempthorne announced that he concurred with Hall and had decided to list the polar bear as a threatened species (US Fish & Wildlife Service 2008, US Department of the Interior 2008). The official announcement included the statement:

> The listing is based on the best available science, which shows that loss of sea ice threatens and will likely continue to threaten polar bear habitat. This loss of habitat puts polar bears at risk of becoming endangered in the foreseeable future, the standard established by the ESA for designating a threatened species.

In "Polar Bear Population Forecasts: A Public-Policy Forecasting Audit," J. Scott Armstrong, Kesten C. Green, and Willie Soon (AGS in this paper) contribute *ex post* to this debate by challenging the scientific validity of the nine USGS studies that Hall cited to support his recommendation. Specifically, AGS assess the degree to which two of the studies that the USGS commissioned (i.e., Amstrup et al. 2007 (AMD) and Hunter et al. 2007 (H6)) adhere to the forecasting principles that the Forecasting Principles Project developed in the mid-1990s. *Principles of Forecasting* (Armstrong 2001) includes these principles (of which there are currently 140); they are also available at www.forecastingprinciples.com, a website that the International Institute of Forecasters sponsors.

After eliminating the 24 principles that they judged as not relevant or applicable to the AMD and H6 efforts, AGS evaluated these studies based on the remaining principles and summarized their challenges by assigning each of these studies a grade of *contravened*, *apparently contravened*, *not auditable*, or *properly applied* for each remaining principle. The purpose of this commentary is to reflect on AGS's challenges.

When policy makers use science to support their recommendations and decisions, we must insist that these scientific efforts meet three criteria. First, the research and researchers must adhere to established scientific standards and practices. Second, they must undertake and execute the scientific research with the sole intent of informing rather than explicitly advocating a specific position on the policy in question (i.e., they must be free of external influence, particularly from the policy and decision makers). Third, the scientific findings must withstand repeated rigorous challenges by other scientists and be consistent with findings of other independent research efforts. While these three criteria are important, they are not sacrosanct. Extenuating circumstances, such as lack of time or unavailability of relevant historical data, might compel scientists engaged in efforts to aid policy makers to strike a balance between rigor and pragmatism. AGS challenge the USGS studies cited by USFWS Director Hall on the basis of each of these three criteria; in doing so, they implicitly argue that neither AMD nor H6 has adequately balanced rigor and pragmatism.

With respect to adherence to established scientific standards and practices of the USGS studies, AGS first challenge that:

> The reports . . . make recommendations with respect to the polar bear listing decision. However, their recommendations do not follow logically from their research because they only make forecasts of the polar bear population. To make policy recommendations based on forecasts, the following assumptions are necessary:
> (1) Global warming will occur and will reduce the amount of summer sea ice;
> (2) Polar bears will not adapt; thus, they will thus obtain less food than they do now by hunting from the sea-ice platform;
> (3) Listing polar bears as a threatened or endangered species will result in policies that will solve the problem without serious detrimental effects; and
> (4) Other policies would be inferior to those based on the Endangered Species Act listing.

Later AGS states:

> AMD and H6 provided no scientific evidence to support their assumptions about any of the four issues that we identified above.

AGS's discussion of the first two assumptions on their list is relevant to the AMD and H6 forecasts and deserves careful consideration. However, counter to AGS's statements, I find no specific policy recommendations in either AMD or H6; each report carefully avoids making such recommendations. Furthermore, the third and fourth assumptions that AGS list are not assumptions of the AMD or H6 forecasts; they are issues that policy makers must consider; therefore, this portion of AGS's list is not relevant to a discussion of the scientific merits of AMD and H6.

With regard to AMD and H6, AGS also maintain:

> Their forecasting procedures depended on a complex set of assumptions, including the erroneous assumption that general circulation models provide valid forecasts of summer sea ice in the regions that polar bears inhabit. Nevertheless, we audited their conditional forecasts of what would happen to the polar bear population *assuming*, as the authors did, that the extent of summer sea ice would decrease substantially during the coming decades. We found that Amstrup et al. properly applied 15 percent of relevant forecasting principles and Hunter et al. 10 percent. Averaging across the two papers, 46 percent of the principles were clearly contravened and 23 percent were apparently contravened. Consequently, their forecasts are unscientific and inconsequential to decision makers. We recommend that researchers apply *all* relevant principles properly when important public-policy decisions depend on their forecasts.

By AGS's assessment, AMD contravened 41 principles and apparently contravened 32 principles, while properly applying only 17 of the remaining 116 principles. In AGS's judgment, H6 contravened 61 principles and apparently contravened 19 principles, while properly applying only 10 of the remaining 105 principles. AGS summarize their grades for the AMD and H6 research efforts on every forecasting principle in appendices. They also discuss in some detail "... some of the more serious problems with the AMD forecasting procedures ..." i.e., AMD's apparent contraventions of the following principles:

- Match the forecasting method(s) to the situation (Principle 6.7)
- Be conservative in situations of high uncertainty or instability (Principle 7.3)
- Obtain forecasts from heterogeneous experts (Principle 8.5)
- Use all important variables (Principle 10.2)

AGS follow with a similar discussion of H6, focusing on the following principles:

- Decisions, actions, and biases (Principles 1.1–1.3)
- Ensure that information is reliable and that measurement error is low (Principle 4.2)
- Obtain all important data (Principle 4.4)

- Be conservative in situations of high uncertainty or instability (Principle 7.3)
- Match the model to the underlying phenomena (Principle 9.2)
- Update frequently (Principle 9.5)
- Use all important variables (Principle 10.2)
- Use different types of data to measure a relationship (Principle 10.5)
- Forecast for alternate interventions (Principle 10.7)
- Provide easy access to the data (Principle 13.8)
- When assessing prediction intervals, list possible outcomes and assess their likelihoods (Principle 14.7)

AMD and H6 unquestionably contravened several of the 140 principles to some degree and should be held accountable for these contraventions. However, it is doubtful that scientific projections provided to inform policy makers with respect to any decision to list a species as threatened or endangered (or any other policy decision) have successfully avoided all contravention of these principles. To adequately assess the research used in support of the polar-bear–listing decision, we must consider two issues (neither of which AGS have addressed):

- After considering the potential gravity of the circumstances and consequences of the ultimate decision, how egregious are the contraventions of AMD and H6? That is, what level of adherence to the forecasting principles is practical in this case?
- How do the contraventions of AMD and H6 compare to the contraventions committed by scientists who have produced projections for similar ESA listing decisions? Did such contraventions lead to substantially inferior forecasts?

With respect to the freedom of the scientific research from influence by the policy makers, AGS state:

> Principle 1.3 (*Make sure forecasts are independent of politics*) is an example of a principle that was contravened in both reports (indeed, in all nine). By politics, we mean any type of organizational bias or pressure. It is not unusual for different stakeholders to prefer particular forecasts; however, if forecasters are influenced by such considerations, forecast accuracy could suffer. The header on the title page of each of the nine reports suggests how the authors interpreted their task: "USGS Science Strategy to Support US Fish and Wildlife Service Polar Bear Listing Decision." A more neutral statement of purpose might have read "Forecasts of the polar bear population under alternative policy regimes."

Given the contentious nature of this issue and the current noxious political environment, would Kempthorne, Hall, or USGS Director Mark D. Myers brazenly flaunt such an obvious and egregious bias? The heading to which AGS refer is indeed included on the title page of each of the nine USGS reports. However, a review of the abstracts and introductions reveals one or more statements in each USGS report that explicitly indicate the research was undertaken to inform the USFWS polar-bear–listing recommendation and not with the intent to support a specific outcome, conclusion, or policy. Other than the titles of the individual USGS research projects, the title pages are identical for each of the nine USGS studies; this is likely because the USGS, USFWS, or Department of the Interior imposed a specific format requirement. Based on the evidence that AGS cited, this challenge lacks merit.

With respect to withstanding repeated rigorous challenges by other scientists and consistence with findings of other independent research efforts, we must temper any assessment by the potential gravity of the circumstances and consequences of the ultimate decision. If this decision is made before other scientists have an opportunity to reflect on and challenge the methodology and conclusions of AMD and H6, we risk incurring the opportunity costs of imposing restrictions on economic activity in the polar bear habitat. On the other hand, we risk the loss of the polar bear species and the resulting (mostly unknown) consequences if this decision is inordinately delayed.

Again, the scientific community has undeniably had relatively little time to ruminate over the nine USGS-sponsored research studies that USFWS cites in its recommendation. However, to adequately assess the research used to support the polar-bear–listing decision, we must consider the following issues, which AGS does not address:

- Do the potential gravity of the circumstances and consequences of the ultimate decision justify the relatively brief opportunity the scientific community was given to critically evaluate the AMD and H6 research efforts?

• How does the opportunity the scientific community was given to critically evaluate the research efforts of AMD and H6 compare to the opportunities given to scientists to critically evaluate projections for similar ESA listing decisions?

A related issue that AGS did not explicitly raise should also be of concern. Forecasting Principle 8.5 states: "Obtain forecasts from heterogeneous experts." The nine USGS papers list 36 authors; however, only 17 different authors contributed to these papers. Amstrup was coauthor of six of these papers; Regehr was coauthor of five; Douglass, McDonald, and Sterling each coauthored three of these papers; Caswell, Durner, Hunter, and Richardson each coauthored two; and Bergen, Howe, Marcot, Nielson, Obbard, Rode, and Runge each coauthored a single paper (DeWeaver was sole author on his only contribution). The lack of independent perspectives for the scientific study of a complex problem with such broad potential implications is disturbing (even if this is a standard government practice).

In conclusion, AGS state in their abstract:

> We recommend that researchers apply *all* relevant principles properly when important public-policy decisions depend on their forecasts.

While AGS raise several valid concerns, questions remain. Was the lack of strict adherence to forecasting principles by AMD and H6 due to their zealous support of the listing of polar bears as an endangered species? Or rather, were these scientists' contraventions merely practical and minor infractions that were justifiable based on the critical nature of the polar bears' circumstances? Have AMD and/or H6 adequately balanced adherence to scientific principles and pragmatism? The final assessment of the scientific merit of the research efforts of AMD and H6 (as well as the other seven USGS studies that USWFS cited) depends on the answers to these questions. Ultimately, what do we choose? Do we protect the polar bear (and its environment) and risk needlessly burdening those who wish to use the arctic region for economic purposes, or do we allow economic development of the polar bears' arctic habitat and risk the eventual extinction of the species? One's opinion about which concern is more critical to the long-term viability of mankind and the planet likely drives one's opinion of how strictly the USGS reports must adhere to the forecasting principles that AGS reference. Which is more critical—the environmental crisis or energy crisis? Your response to this question ultimately determines your assessment of the validity of AGS's challenges to the AMD and H6 studies.

Ultimately and somewhat paradoxically, when forming my opinion on the AMD and H6 studies and the resulting policy decision, I find myself returning to Forecasting Principle 7.3 and AGS's ensuing discussion:

Be conservative in situations of high uncertainty or instability (*Principle 7.3*)

> Forecasts should be conservative when a situation is unstable, complex or uncertain. Being conservative means moving forecasts towards "no change" or, in cases that exhibit a well-established, long-term trend and where there is no reason to expect the trend to change, being conservative means moving forecasts toward the trend line. A long-term trend is one that has been evident over a period that is *much longer* than the period being forecast. Conservatism is a fundamental principle in forecasting.

In the face of such high uncertainty or instability, policy makers should be held to the same standard of conservatism. And policies that are intended to preserve *Ursus maritimus* through attempts to safeguard its habitat are certainly more conservative than policies that allow for potential changes or disruption of this habitat.

References

Adler, J. 2008. The race for survival. *Newsweek* (June 9) 40–50.

Amstrup, S. C., B. G. Marcot, D. C. Douglas. 2007. Forecasting the rangewide status of polar bears at selected times in the 21st century. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Armstrong, J. S. 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA.

Bergen, S., G. M. Durner, D. C. Douglas, S. C. Amstrup. 2007. Predicting movements of female polar bears between summer sea ice foraging habitats and terrestrial denning habitats of Alaska in the 21st century: Proposed methodology and pilot assessment. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Center for Biological Diversity. 2005. Petition to list the polar bear (*ursus maritimus*) as a threatened species under the Endangered Species Act. Retrieved May 30, 2008, http://www.eswr.com/latest/polarbearpetition.pdf.

DeWeaver, E. 2007. Uncertainty in climate model projections of Arctic sea ice decline: An evaluation relevant to polar bears. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Durner, G. M., D. C. Douglas, R. M. Nielson, S. C. Amstrup, T. L. McDonald. 2007. Predicting the future distribution of polar bear habitat in the polar basin from resource selection functions applied to 21st century general circulation model projections of sea ice. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Hunter, C. M., H. Caswell, M. C. Runge, S. C. Amstrup, E. V. Regehr, I. Stirling. 2007. Polar bears in the southern Beaufort Sea II: Demography and population growth in relation to sea ice conditions. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Obbard, M. E., T. L. McDonald, E. J. Howe, E. V. Regehr, E. S. Richardson. 2007. Trends in abundance and survival for polar bears from Southern Hudson Bay, Canada, 1984–2005. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Regehr, E. V., C. M. Hunter, H. Caswell, S. C. Amstrup, I. Stirling. 2007. Polar bears in the Southern Beaufort Sea I: Survival and breeding in relation to sea ice conditions, 2001–2006. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Rode, K. D., S. C. Amstrup, E. V. Regehr. 2007. Polar bears in the Southern Beaufort Sea III: Stature, mass, and cub recruitment in relationship to time and sea ice extent between 1982 and 2006. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Stirling, I., T. L. McDonald, E. S. Richardson, E. V. Regehr. 2007. Polar bear population status in the Northern Beaufort Sea. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

US Department of the Interior. 2008. Secretary Kempthorne announces decision to protect polar bears under Endangered Species Act. Retrieved June 2, 2008, http://www.doi.gov/issues/polar_bears.html.

US Fish and Wildlife Service. 2008. Secretary Kempthorne announces decision to protect polar bears under Endangered Species Act. Retrieved June 2, 2008, http://www.fws.gov/news/NewsReleases/showNews.cfm?newsId=ECB61DD1-0D74-1D7B-4A67E9B51FB1626B.

Unscientific Forecasts and Wise Decisions: Commentary on Armstrong, Green, and Soon

Paul Goodwin
University of Bath, The Management School, Claverton Down, Bath, United Kingdom,
mnspg@bath.ac.uk

Armstrong, Green, and Soon argue that forecasts of a serious decline in the population of polar bears should be discounted because they are not based on scientific forecasting principles. They go on to imply that, because the argument for listing polar bears as an endangered species is based on these forecasts, there is no current case for such a listing. In this commentary, I examine two aspects of their argument. First, should the principles established in the forecasting principles project (Armstrong 2001) be applied to forecasting in the natural sciences and, second, should an absence of credible forecasts of declining polar bear populations necessarily preclude a decision to add them to the endangered species list?

I looked at the brief biographies, which Armstrong (2001) lists, of the "forty internationally recognised experts" who formulated the forecasting principles. Of these, 31 (nearly 78 percent) were primarily associated with business or economic forecasting. The remainder were either psychologists or statisticians. None of the experts had a predominant interest or background in the natural sciences. This raises the question of whether there are fundamental differences between business and economic forecasting and forecasting in the natural sciences, which might render the principles inappropriate for evaluating the quality of the polar bear forecasts that Armstrong, Green, and Soon report.

If one compares studies of business and economic forecasting (e.g., Fildes and Hastings 1994, Fildes and Goodwin 2007, Turner 1990) with reports of natural science forecasting (e.g., Amstrup et al. 2007, Green and Armstrong 2007), it is difficult to find any fundamental distinctions between the characteristics of the two types of forecasts. Indeed, it is easier to find similarities than differences. Both forecast types often involve the simple extrapolation of past time series patterns or the use of models of complex systems to predict outputs from inputs. Both normally involve the need to filter noise so that we can identify underlying systematic patterns. Both often involve substantial amounts of judgments by experts;

and both can be significantly affected by political considerations. In the natural sciences these political considerations might include the desire to support the prevailing wisdom of the day, perhaps to secure additional research funding, or the wish to confirm one's own theories (Mahoney 1979, Green and Armstrong 2007). In a business setting, these considerations might involve the need to please senior managers or the inclination to enhance the interests of one's own department (Fildes and Hastings 1994, Goodwin 1998). The misuse of mathematical or statistical methods in the two domains is also sometimes surprisingly similar with an expert producing a forecast based on judgment and then creating or adapting a model to confirm this judgment to provide a veneer of scientific validity for the desired forecast (Goodwin et al. 2007, Green and Armstrong 2007).

Nevertheless, a large part of the case that Armstrong, Green, and Soon present is based on doubts about the role of the experts in the polar bear forecasts. Perhaps expert judgments are more reliable in the natural sciences? When Shanteau (1992) compared the effectiveness of experts' judgments in 20 fields, he found no systematic distinction between the quality of judgments made in science and business. While weather forecasters, accountants, auditors, and physicists displayed good judgment, the judgments of physicians, psychiatrists, stockbrokers, and personnel selectors were often poor. Shanteau concluded that expert judgment tends to be reliable when it relates to something that is relatively constant, when it involves physical stimuli rather than human behavior, and when regular and rapid feedback allows the expert to learn from past performance. While the human-behavior element may be relatively disadvantageous to business and economic forecasters, it appears that we should also be skeptical of the forecasts of expert scientists when there is a lack of constancy in the variable to be forecast (e.g., global temperatures and polar bear populations) and an absence of rapid and reliable feedback.

This suggests that, in general, the principles apply just as much to forecasting in the natural sciences as they do to business and economic forecasts. While different principles might have varying degrees of relevance and importance in different forecasting situations, this is true whether one is comparing different business situations or business and scientific forecasting. Indeed, many of the principles are self-evidently sensible. Estimating a causal relationship from five observations is not advisable in any field. On that basis, the critique that Armstrong, Green, and Soon made of the polar-bear–population forecasts is compelling.

Can the same be said about the authors' implication that polar bears should not be listed as an endangered species? I believe that we should be more cautious. Normative decision theory separates the process of making a forecast from the process of evaluating the possible outcomes of each course of action. Conventional decision models represent a forecast as a probability distribution of the outcomes. Unlike a point forecast (e.g., a forecast that the current polar bear populations will increase by 10 percent by mid century), a forecast that is expressed as a probability distribution provides information on the level of uncertainty that is associated with a course of action. For example, a 10 percent increase may simply be the mean of the probability distribution; however, there may be a 0.99 probability that the increase in the population could range between −20 percent and +40 percent. The authors point out that, in conditions of high uncertainty, forecasts should be conservative; hence, no change forecasts are advisable. While this may apply to the estimate of the central tendency of the probability distribution, we should clearly increase the dispersion of this distribution when uncertainty is greater. Consider a probability distribution of the polar bear population in 2030. Even if we apply a no-change forecast to the distribution's central tendency, there is still likely to be a greater-than-zero probability of a significant decline in polar bear numbers. Suppose that we estimate this probability to be only 1 in 1,000. Can we say that the risk is so low that the listing of polar bears is unjustified? Without an evaluation of the consequences of a significant population decline, we cannot make such a decision. The value that society attaches to a healthy polar bear population might make such a risk unacceptable, even when balanced against the advantages of not listing polar bears (such as greater freedom to exploit resources in the polar regions). Few people would tolerate a nuclear power plant near their town that had a 1 in 1,000 probability of a meltdown in a 20-year period. In short, it is not possible to use a forecast to establish that a decision is wrong. This applies even if we express the forecast as a probability distribution, let alone a point forecast.

Armstrong, Green, and Soon are right; the forecasts they criticize tell us nothing about which course of action is appropriate. However, while the evidence-based forecasting process that they recommend *would* inform the decision, it would, on its own, be insufficient to tell us whether or not polar bears should be protected.

References

Amstrup, S. C., B. G. Marcot, D. C. Douglas. 2007. Forecasting the rangewide status of polar bears at selected times in the 21st century. Administrative Report, USGS Alaska Science Center, Anchorage, AK.

Armstrong, J. S. 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA.

Fildes, R., P. Goodwin. 2007. Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces* **37** 570–576.

Fildes, R., R. Hastings. 1994. The organization and improvement of market forecasting. *J. Oper. Res. Soc.* **45** 1–16.

Goodwin, P. 1998. Enhancing judgmental sales forecasting: The role of laboratory research. G. Wright, P. Goodwin, eds. *Forecasting with Judgment*. Wiley, Chichester, UK, 91–111.

Goodwin, P., W. Y. Lee, R. Fildes, K. Nikolopoulos, M. Lawrence. 2007. Understanding the use of forecasting systems: An interpretive study in a supply-chain company. Working paper, Bath University Management School, Bath, UK.

Green, K. C., J. S. Armstrong. 2007. Global warming: Forecasts by scientists versus scientific forecasts. *Energy Environ.* **18** 997–1021.

Mahoney, M. J. 1979. Psychology of the scientist: An evaluative review. *Soc. Stud. Sci.* **9** 349–375.

Shanteau, J. 1992. The psychology of experts: An alternative view. G. Wright, F. Bolger, eds. *Expertise and Decision Support*. Plenum, New York, 11–23.

Turner, D. S. 1990. The role of judgment in macroeconomic forecasting. *J. Forecasting* **9** 315–345.

Commentary: Polar Bear Population Forecasts: A Public-Policy Forecasting Audit

Frederic Murphy
Fox School of Business and Management, Temple University, Philadelphia, Pennsylvania 19122,
fmurphy@temple.edu

Armstrong, Green, and Soon make very powerful points about the reliability of public-policy forecasts. What do you do when you must address an important policy question but there is no available data to fit with any statistical method? The major problem is that every expert carries around some mental model and argues using that model. Mental models have more hidden assumptions than explicit models do; verification consists of quoting everything possible that supports that model. This reduces policy-making to argument, something that Washington lawyers relish to the point where they come to believe their mental models are fact.

People with a singular lack of insight often dominate the debates on how to frame a problem. Jack Anderson, a noted Washington columnist during the 1970s energy crisis, framed the energy problem as an internal government conspiracy; he then proceeded to ruin careers through innuendo. Barry Commoner, one of the early ecologists, also articulated conspiracy theories, essentially acting as the Don Rickles of energy-policy analysis. The most interesting feature of Commoner's career is that he left biology, became an ecologist, after he used his polemical skills to argue against Watson and Crick, the discoverers of DNA, and the existence of DNA. He demonstrated that the need to debate is more important for some people than the subject debated. What has saved us from conspiracy theorists in this round of energy-price increases is that no lurking conspirators have ever been found.

Another skewed framing of the energy debate was and is the notion of energy independence. The public figures who now talk of "independence" forget that Richard Nixon framed the problem with his quote "if we can put a man on the moon, we can achieve energy independence." Thus, they ally themselves with one of the worst US presidents. The concept of energy independence suffers from more than guilt by association: two aspects of the perspective were and are dead wrong. First, it bounds the problem as an engineering and technology problem and ignores that it is also a lifestyle problem. When we talk about lifestyle, we should not talk about "needs." As long as we are not starving in Darfur, we should talk of "wants." The use of "needs" too often frames debates in ways that are not helpful; and "needs" is an empirical claim that demand elasticities are 0, but stated as if absolute truth. The words (and mental models) of lawyers and engineers too often consist mainly of beliefs that they articulate as immutable truths. Because lawyers dominate politics and engineers dominate the Department of Energy, the high-level discussions are regularly framed in terms of needs to be met by overly subsidized new technologies. Yet, having only a technology focus ignores the adaptability of the populace. Armstrong, Green, and Soon point out that polar bears have a history of adaptation. If polar bears can adapt, so too can people.

Second, we must not forget the formative experiences of the Nixon-era leaders. Nixon, the last of the New Deal presidents, frequently talked about his "running the country." All of the formative experiences of the nation's leaders in the early 1970s were rooted in the Depression and World War II—periods in which government, not markets, solved problems. I and my then-young colleagues studied microeconomics as it was developed in the 1950s and 1960s. The glory days of the New Deal were something we read about in history books and our parents fought in WWII. Thus, the energy debates of the 1970s were a contest between the markets view and a "running the country" view that had an intergenerational twist.

At the time of the legislative debates, John Dingell, the head of the House Energy and Commerce Committee and a prime believer in the New Deal and big government, found himself weather stripping his windows to lower his heating bills. He realized that people can and do adapt. Thus, he came to accept that demand elasticity is greater than 0 through his own actions. The market-oriented energy legislation of the Carter years passed because of that simple experience. History also came out on the side of markets working in most, but not all, cases. That markets have worked in most areas, but not all, illustrates that the virtues of markets are also not an absolute.

Armstrong, Green, and Soon argue too blithely about staying independent of politics. As I have described, we all operate with belief systems, and those belief systems lead us to self-select where we stand on issues and with whom we associate. Our beliefs lead us to the point where we know our side is right. We have all seen the maps of who buys what books from Amazon and the disjunction between "lefty" and "righty" readers. Both sets of purchasers think they are right—the people on the other side are "political." After all, those with whom we choose to associate reinforce our perception of cultural and social norms. My views are shaped by hanging out with colleagues who have analytic skills, which I see as all too rare in the general public.

Thompson (1984), a sociologist, writes of visiting the International Institute on Applied Systems Analysis, outside of Vienna, Austria. After observing the researchers there, he described the different energy tribes and how they made assumptions in their models that led to results that validated their tribal beliefs. That process is repeating itself in the global-warming debates. We have new Barry Commoners engaged in the same battles for prominence over a different subject. After all, the personal desire is prominence; the topic at hand is the vehicle. Independent of the truths that are yet to be revealed, we are seeing the law of conservation on debate intensity: Once one issue quiets down, those in need of an argument will find another issue to debate with equal fervor.

The papers that Armstrong, Green, and Soon examine are tribal statements and must be seen as such. The authors of those papers may be in the winning tribes. However, if they are, polar bears will be gone before the full evidence is in.

What I do not see in the Armstrong, Green, and Soon paper, or read in the newspapers, is the consequence of the decision to list polar bears as endangered species. What are the benefits of listing them, other than restricted hunting? To what extent is polar bear mortality due to hunting? Is there anything short of ending global warming or putting all of them in zoos that will help polar bears? Furthermore, I do not know what the real decision situation is. For example, if the models are right and the polar bears will die, then the policy that maximizes social welfare is to hunt them and eat them before they drown from lack of sea ice. If global-warming decisions are to be made solely for the preservation of polar bears, then we must address the reality that fossil fuels have been essential in holding off the Malthusian outcome of mass starvation. Are the authors just members of the global-warming tribe who are using polar bears as another vehicle to beat up on an energy-consuming world where there is no real decision that can change this situation other than addressing global warming head on? Then listing the bears as endangered has only symbolic value and no bears will be saved. If the cost of a bear-preserving solution is cheap, then the forecast errors will not cost much if the solution is implemented.

The real decision problem needs to be articulated before the forecasts are made. The tribal assumptions must be vetted. At the same time, we should be aware that it is easier to pick apart the details of an explicitly stated analytical model than a mental model. Going after specified models has the unintended consequence of giving more credence to untestable mental models.

Reference

Thompson, M. 1984. Among the energy tribes: A cultural framework for the analysis and design of energy policy. *Policy Sci.* **17**(3, September) 321–339.

What Is the Appropriate Public-Policy Response to Uncertainty?

J. Scott Armstrong
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104,
armstrong@wharton.upenn.edu

Kesten C. Green
Business and Economic Forecasting, Monash University, Victoria 3800, Australia,
kesten@kestencgreen.com

Willie Soon
Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138,
wsoon@cfa.harvard.edu

Scientific forecasting methods can help to inform decisions by providing forecasts and prediction intervals about the costs and benefits of *alternative policies*. We believe that this is an underlying message in the commentaries of Professors Goodwin and Murphy.

We have a minor point to add to Professor Murphy's commentary. In discussing mental models, he might also have mentioned the importance of judgmental bootstrapping for revealing how decision makers and forecasters think about a situation (Armstrong 2001). Two of this paper's authors were recently involved in a study that showed that people tend to vote for competent-looking presidential candidates (Armstrong et al. 2008). We suspect that if polar bears were ugly, there would be no mass movement to support them. In Senator Boxer's hearings, the only exhibits presented to support the listing were attractive photographs of polar bears.

Professor Cochran provides a good summary of the process involved in the listing of polar bears. However, he characterizes our contribution as *ex post*. In fact, we made considerable efforts to ensure that our analysis was available to decision makers prior to the time that they made their decisions. Dr. Armstrong presented testimony at Senator Boxer's US Senate Hearing on the listing of polar bears on January 30, 2008. The website, www.theclimatebet.com, includes a video of this testimony; www.publicpolicyforecasting.com made available draft versions of our paper prior to the testimony. We kept officials at the Department of Interior, as well as authors of the administrative reports, informed of our work as it progressed. This is especially important because our forecast, which we based on evidence-based methods, is that the polar bear population will continue to grow slowly, whereas the government reports forecast a large and rapid decline in the population.

Professor Cochran is wrong to suggest that the petition that the Center for Biological Diversity (CBD) filed is solely for polar bear protection. In fact, the CBD is using the polar bear listing as a means to reduce emissions of carbon dioxide from fossil fuels (Center for Biological Diversity 2007).

We also have the following issues with Professor Cochran's logic:

(1) He chooses to use standard practice rather than scientific procedures as a basis for judging the scientific basis for the forecasts. In other words, if previous endangered species studies have failed to follow scientific procedures (and we suspect that he is correct in this assumption) then, he implies, it is appropriate to ignore scientific procedures when forecasting polar bear populations.

(2) He claims that the US Department of the Interior's administrative reports did not have a position on this policy. He argues that the heading of each report ("to support US Fish and Wildlife Service Polar Bear Listing Decision") does not imply a position.

(3) He refers to our principle for forecasting, "Be conservative in situations of high uncertainty or instability," as being equivalent to his position, which is effectively the "precautionary principle." To the contrary, we believe the positions are diametrically opposed. Our principle refers to forecasting small changes when uncertainty is high. Important changes should be based on a scientific study of the costs and benefits of alternative policies; to the extent that there is uncertainty, one should avoid major policy changes. The precautionary principle argues that uncertainty is a basis for action; if one lacks knowledge, then some action should be taken—just in case. This happens when interest groups identify an issue that can help them to achieve their ends. If the interest group is successful in lobbying for an issue, politics replaces science, and government dictates follow. It brings to mind the slogan on the Ministry of Truth building in George Orwell's *1984*: "Ignorance is Strength."

We believe that proper scientific principles will lead to better decisions than will political principles, and that people will be better off if politicians have the courage to resist calls to action when uncertainty is high.

References

Armstrong, J. S. 2001. Judgmental bootstrapping: Inferring experts' rules for forecasting. J. S. Armstrong, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA.

Armstrong, J. S., K. Green, R. Jones, M. Wright. 2008. Predicting elections from politicians' faces. Retrieved June 11, 2008, http://forecastingprinciples.com/Faces_24_OctR.pdf.

Center for Biological Diversity. 2007. Not too late to save the polar bear: A rapid action plan to address the Arctic meltdown. Retrieved July 7, 2008, http://www.biologicaldiversity.org/publications/papers/ArcticMeltdown.pdf.