

# Communicating scientific uncertainty

Baruch Fischhoff<sup>a,b,1</sup> and Alex L. Davis<sup>a</sup>

Departments of <sup>a</sup>Engineering and Public Policy and <sup>b</sup>Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890

Edited by Dietram A. Scheufele, University of Wisconsin–Madison, Madison, WI, and accepted by the Editorial Board February 24, 2014 (received for review November 5, 2013)

**All science has uncertainty. Unless that uncertainty is communicated effectively, decision makers may put too much or too little faith in it. The information that needs to be communicated depends on the decisions that people face. Are they (i) looking for a signal (e.g., whether to evacuate before a hurricane), (ii) choosing among fixed options (e.g., which medical treatment is best), or (iii) learning to create options (e.g., how to regulate nanotechnology)? We examine these three classes of decisions in terms of how to characterize, assess, and convey the uncertainties relevant to each. We then offer a protocol for summarizing the many possible sources of uncertainty in standard terms, designed to impose a minimal burden on scientists, while gradually educating those whose decisions depend on their work. Its goals are better decisions, better science, and better support for science.**

science communication | expert judgment | expert elicitation | risk | mental models

**D**ecision making involves uncertainty. Some of that uncertainty concerns facts: What will happen if we make a choice? Some of that uncertainty concerns values: What do we want when we cannot have everything?

Scientific research can reduce both kinds of uncertainty. Research directed at clarifying facts can provide imperfect answers to questions such as how well an oil pipeline will be maintained and monitored, how long the recovery period will be after bariatric surgery, and how much protection bicycle helmets afford. Research directed at clarifying values can provide imperfect answers to questions such as which charitable contributions give donors the greatest satisfaction, what happens when people commit to long-term goals, how much pleasure people can expect from windfalls (e.g., lottery winnings) and how much pain from misfortunes (e.g., severe accidents) (1, 2).

Taking full advantage of scientific research requires knowing how much uncertainty surrounds it. Decision makers who place too much confidence in science can face unexpected problems, not realizing how wary they should have been. Decision makers who place too little confidence in science can miss opportunities, while wasting time and resources gathering information with no practical value. As a result, conveying uncertainty is essential to science communication.

Much of scientists' own discourse is about uncertainty. Journals require authors to disclose the assumptions and ambiguities underlying their work. Scientific debates focus on uncertainties requiring attention. Peer review scrutinizes the uncertainty in individual studies, protecting science from unwarranted faith in flawed results. A healthy scientific community rewards members who raise problems before their critics and penalizes those who overstate results.

By revealing uncertainties, scientific discourse is an essential resource for communications about them. However, it typically provides both more and less detail than decision makers need (3). At one extreme, scientists' discourse can involve minutiae that overwhelm even experts. At the other, it can omit vital uncertainties that are common knowledge within a field, hence go without saying, and uncertainties that a field routinely ignores, either because they appear unimportant or because its scientists have nothing to say about them.

As a result, communicating scientific uncertainty requires both simplifying and complicating normal scientific discourse. On the

one hand, the uncertainties that it addresses must be reduced to their decision-relevant elements. On the other hand, the uncertainties that scientists fail to mention must be uncovered. Which uncertainties to subtract and add depends on the decisions that the communications are meant to serve. Some decisions (e.g., whether to reexamine disaster plans) may be robust to any plausible uncertainty (e.g., how storm surge estimates are revised). Others (e.g., whether to buy monetized securities) may be highly sensitive to details (e.g., the assumptions made about the liquidity of capital markets).

Once communicators know what science is worth knowing, they can study what facts decision makers know already and how best to convey missing knowledge. As with all communication, though, the first task is determining what to say. That will depend on the decision, which might fall into one of three broad categories. (i) Decisions about action thresholds: Is it time to act? (ii) Decisions with fixed options: Which is best? (iii) Decisions about potential options: What is possible?

Communicating uncertainty for each class of decision requires (i) characterizing uncertainty, by identifying the issues most relevant to the choice; (ii) assessing uncertainty, by summarizing that information in a useful form; and (iii) conveying uncertainty, by creating messages that afford decision makers the detail that their choices warrant. After considering the scientific foundations for addressing these tasks, we offer a protocol for eliciting uncertainty from the scientists who know it best, in a standard form designed to serve those who need to learn about it.

From this perspective, communications are effective when they help people identify choices that serve their own, self-defined best interests. In contrast to such nonpersuasive communications, persuasive messages are effective when they convince people to behave in ways that someone else has chosen (e.g., public health officials, marketers, politicians). With persuasive communications, shading or hiding uncertainty might be justified. With nonpersuasive ones, honesty is the only policy.

## Decisions About Action Thresholds: Is It Time to Act?

The simplest but sometimes most fateful communications inform decisions to act, triggered by evidence passing a threshold for action. Science may provide that trigger, in communications such as those telling people to head for (or leave) storm shelters, start (or stop) medical treatments, and sell (or buy) securities. However, setting a threshold for action inevitably poses value questions. Which tornado warning level best balances protection and disruption? Which medical guideline embodies the best tradeoff between the uncertain risk and benefits of a treatment? Which portfolio rebalancing policy is best for a retirement plan? As a result, communications must address the uncertainties surrounding both scientists' facts and decision makers' values.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "The Science of Science Communication II," held September 23–25, 2013, at the National Academy of Sciences in Washington, DC. The complete program and video recordings of most presentations are available on the NAS website at [www.nasonline.org/science-communication-II](http://www.nasonline.org/science-communication-II).

Author contributions: B.F. and A.L.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.A.S. is a guest editor invited by the Editorial Board.

<sup>1</sup>To whom correspondence should be addressed. Email: [baruch@cmu.edu](mailto:baruch@cmu.edu).

**Characterizing Uncertainty.** People will judge science's value by the apparent wisdom of recommendations based on it. Do calls to action seem to lead to saving lives, health, and profits or to needlessly taking shelter, undergoing surgery, or selling valuable securities? Those evaluations should depend on both the uncertainty in the science and the appropriateness of the decision rule translating it into a recommendation. In retrospect, such evaluations can be colored by outcome bias, arising when decisions are judged by their outcomes rather than by their wisdom (4), and by hindsight bias, arising when people exaggerate how well those outcomes could have been foreseen (5).

Even when decision makers resist such biases, they can judge science fairly only if they can see its evidence through the filter of the recommendations. If a call to action proves unnecessary, then the science might have been highly uncertain, making mistakes inevitable, or the science might have been strong but the threshold overly cautious. Conversely, if a needed call to action was missed, it could mean uncertain science or insufficient caution.

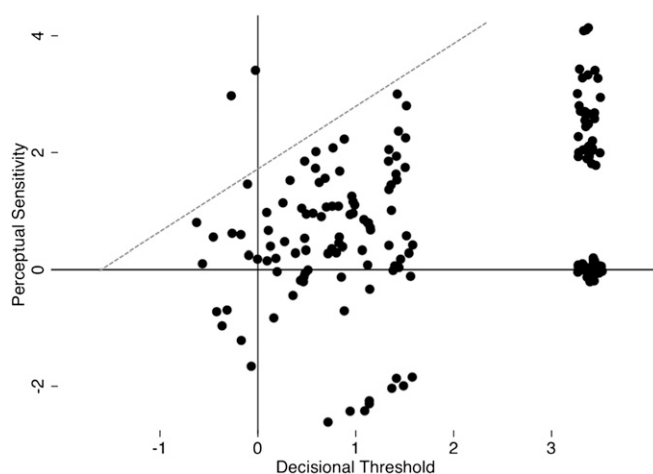
With a single event, decision makers cannot disentangle the evidence and the decision rule on their own but must be told how the advice was reached. That explanation might be, "We recommend this risky, painful surgery because we are confident that it increases 5-y survival rates" or, "The evidence for a market collapse is weak, but it is our fiduciary responsibility to issue a 'sell' call for retirement accounts." With multiple events, though, decision makers can infer the quality of the science and the decision rule from the pattern of decisions and outcomes. As formalized by signal detection theory, quality is represented by  $d'$  (how well experts can discriminate among states of the world), whereas decision rules are represented by  $\beta$  (the tradeoffs made among possible outcomes) (6).

Fig. 1 illustrates the kind of research needed to estimate  $d'$  and  $\beta$  (7) from a study examining physicians' decisions about whether to transfer emergency room patients from regional hospitals to medical centers. As with all empirical studies, it has heterogeneity in both parameters. Experts vary in their discrimination ability and diagnoses vary in their difficulty. Experts vary in their decision rules and cases pose different tradeoffs. Statistical analyses are needed to estimate these parameters. Unless decision makers receive such estimates, they must rely on their own intuitive statistics for inferring the uncertainty of the science and the caution of the decision rule.

If decision makers guess incorrectly about  $d'$ , then they may place inappropriate confidence in the science, pay inappropriate attention to new information, and hold inappropriate resources in reserve. If they guess incorrectly about  $\beta$ , then they may take undesired risks, follow advice unsuited to their needs, and feel misled by the experts. Thus, although calls to action (or inaction) can demonstrate science's usefulness, they can also generate distrust when those messages are misinterpreted or misaligned with decision makers' values. Communications increase the chance of better outcomes when they make the uncertainties underlying recommendations transparent. Doing so requires both analytical research for assessing uncertainty and empirical research for conveying it.

**Assessing Uncertainty.** The receiver operating curve is the standard summary of the uncertainty underlying categorical messages (e.g., whether to shelter from a storm or get a medical treatment) (6). It shows the tradeoffs possible with a given ability to discriminate states of the world. The greater that ability is, the more favorable the decision options become. For example, as medical imaging improves, a decision maker who tolerates a 5% chance of missing a cancerous tumor will face less risk of exploratory surgery that finds nothing.

How precise estimates of  $d'$  and  $\beta$  must depend on the decision. Sometimes decision makers need only a rough idea of how much experts know and what values the decision rule embodies. Sometimes they need greater precision. Estimation is harder when they must infer both parameters, compared with when they know one and must estimate the other. Experts



**Fig. 1.** Signal detection theory in evaluating decisions to transfer ER patients. Emergency physicians provided written recommendations for the next steps in treating patients depicted in detailed profiles drawn from actual records. Higher values for the decisional threshold indicate more cautious decisions ( $\beta$ ). Higher values for perceptual sensitivity indicate better discrimination ability ( $d'$ ). American College of Surgeons—Committee on Trauma (8) guidelines were applied to the profiles to identify the appropriate decision. (Reproduced with permission from ref. 7.)

can eliminate uncertainty about values ( $\beta$ ) by applying and communicating explicit decision rules, leaving just uncertainty about their knowledge ( $d'$ ).

Estimating  $d'$  or  $\beta$  means evaluating recommendations in the light of subsequent events. Sometimes, those events can be observed directly, for example, seeing how often floods (or tumors) follow warnings and reassurances. Doing so requires defining the events clearly (e.g., precisely what "tumor" means), lest performance be judged too harshly or leniently ("we did not mean to include benign tumors"). Sometimes, the evidence is indirect, as when diseases are inferred from biomarkers or tornadoes from damage patterns. The weaker the science supporting those inferences, the greater the uncertainty is about what scientists knew ( $d'$ ) and valued ( $\beta$ ). For example, (7) used American College of Surgeons guidelines (8) to define the event "patient requires transfer to a major medical center." If those guidelines predict medical outcomes, using them simplifies the analysis (compared with evaluating cases individually). If not, then using them may bias the analysis, perhaps even imposing a standard that some physicians reject.

An alternative to observing experts' beliefs and values is to elicit them (9). Like survey research, expert elicitation must balance decision makers' desire for precise answers with respondents' ability to translate their knowledge into the requested terms. Research into how people access their own beliefs can guide that process (10). For example, rather than asking people why they reached a conclusion, it is better to ask them to think aloud as they make the inference, in order to avoid having the answers contaminated by their intuitive theories of mental processes.

**Conveying Uncertainty.** Effective communication requires clear, mutually understood terms. One threat arises when scientists use common words in uncommon ways. For example, probability-of-precipitation forecasts can mislead people who do not know that "precipitation" means at least 0.01 inches at the weather station (11). Survey researchers have documented the difficulty of conveying what experts mean by such seemingly simple terms as "room," "safe sex," and "unemployed" (12, 13). Clarity is an empirical question, answered by user testing. When scientists' favored term fails to communicate, a more familiar one may be needed (e.g., using "given up trying to find a job" rather than "out of the labor market"). After Hurricane Sandy, the National

Oceanic and Atmospheric Administration concluded that it had confused the public by downgrading the storm when it fell below formal hurricane status, despite being more powerful than many people could imagine (14).

Once the terms are clear, decision makers can consider  $d'$  and  $\beta$ . As mentioned, given multiple observations, decision makers can estimate these parameters, subject to the biases afflicting such intuitive statistics (e.g., flawed memory, illusory correlation, wishful thinking, denial). Given a single observation, though, they must rely on their beliefs about experts' knowledge and values when asking questions such as "Why is there so much apparent controversy in that field?" and "How much do those experts know, or care, about people like me, when giving advice?" (15, 16).

Serious miscommunication can arise when recipients do not realize that categorical advice always reflects both parameters. Physicians who fail to transfer seriously ill patients to medical centers might be faulted for their decision rule (e.g., wanting to keep the business at their facility) when the problem lies with their judgment (e.g., overlooking critical symptoms). Observers may overestimate the uncertainty of scientists whose professional norms preclude giving simple answers to complex questions (e.g., "Is this storm due to climate change?"). They may place unwarranted faith in pundits who face no sanctions for going far beyond the evidence. Observers familiar with the norms of science might misinterpret scientists who abandon their customary caution for advocacy or style themselves as public intellectuals (17).

As with all communications, the content of science-based messages should depend on what recipients know already. At times knowing the gist of  $\beta$  might allow inferring  $d'$  (18). For example, "FDA approved" (Food and Drug Administration) means more to people who know that the agency's  $\beta$  for approving a drug means "potentially suitable for some patients," implying a range of acceptable tradeoffs. At other times, decision makers need more detail. For example, recommendations against routine mammography screening for younger women evoked public furor in 1997 and again in 2009 (19, 20). They might have fared better had they communicated more clearly the  $d'$  for mammography (which has difficulty detecting treatable cancers in younger women) and the parties whose welfare its  $\beta$  considered (patients? physicians? insurers?) (21).

### Decisions with Fixed Options: Which Is Best?

Categorical communications recommend choices. Other communications merely seek to inform choices, allowing recipients to apply their values by themselves. Information about scientific uncertainty can serve that end, for example, by telling patients how well new and old drugs are understood; by telling investors how predictable bonds or stocks are, given their use in complex financial products; or by telling climate change advocates what surprises to expect from future research, with campaigns focused on ocean acidification or sea-level rise.

**Characterizing Uncertainty.** Sensitivity analysis is the common formalism for assessing how much people need to know about how much is known (22). It asks questions like, "Would any plausible value for expected mean global temperature change affect your preferred energy portfolio?" and "Would any prostate-specific antigen test result affect your prostatectomy decision?" If the answer is no, then that uncertainty is immaterial to the decision, whatever the value of having reduced it that far.

If the answer is yes, then one can elicit experts' judgments of the uncertainties that matter to the decision. A standard representation of uncertainty is a probability distribution over possible parameter values (23). For example, Morgan and Keith elicited the judgments of 16 climate experts for questions such as globally averaged surface temperatures given a doubling of atmospheric  $\text{CO}_2$  (24). For some decisions (and decision makers), knowing an extreme fractile will suffice (e.g., "with even a 5% chance of a 3 °C increase, I support an aggressive carbon tax"). For others, such as investors in carbon permits, knowing the full probability distribution may have value.

Whether such probability distributions capture all useful uncertainty has been a topic of lively academic debate (25–28). Some decision theorists argue for the usefulness of higher-order uncertainties (e.g., ranges of probability distributions) when probabilities are not well known. Others contend that all beliefs should be reducible to a single distribution. One practical reason for assessing higher-order uncertainties is preparing decision makers for surprises. For example, one of Morgan and Keith's climate experts gave two probability distributions (for temperature with doubling of atmospheric  $\text{CO}_2$ ), depending on whether the North Atlantic thermohaline circulation collapses. That expert could not assign a probability to that event, thereby allowing a weighted combination of the two conditional distributions, arguing that it was beyond current scientific knowledge. On other questions, some experts predicted that scientific uncertainty would increase over time, as research revealed unforeseen complications. Decision makers need to know when uncertainty assessments include all factors or just factors that scientists are comfortable considering.

A second practical reason for going beyond a single summary distribution is identifying opportunities to reduce uncertainty. Decision makers can then ask whether the benefits of that additional information outweigh the direct costs of collecting it and the opportunity costs of waiting for it. Those costs will depend on the source of uncertainty. When it arises from sampling variation, more observations (e.g., additional clinical trial patients) will provide greater precision, with readily calculated costs and benefits. When uncertainty arises from poor measurement, calculating the cost of reducing it might be straightforward (e.g., using a better test) or difficult (e.g., developing a better test). When uncertainty arises from unstudied issues, the costs and benefits of new knowledge are matters for expert judgment. Whatever the case, decision makers need to know how certain the science could be, as well as how certain it currently is.

**Assessing Uncertainty.** Scientists routinely apply statistical methods to assess the variability in their data. Those calculations add uncertainty to inferences about the data through the assumptions that they make (e.g., normal distributions, independent errors). Additional uncertainty arises from how the data are treated before analysis. For example, political polls may include or delete incomplete surveys or ones evoking negative answers to screening questions (e.g., "Are you following the lieutenant governor's race closely enough to answer questions about it?"). Methodological research can inform assessments of such uncertainties (29, 30). As elsewhere, their importance depends on the decision. Issues that matter for tight elections may be irrelevant for blowouts.

Statistical methods can also summarize the variability in the estimates produced by integrative models (e.g., simulations of climate, pandemics, or national economies). The assumptions that modelers make when creating and estimating their models add uncertainty to that already found in the data that they use. Scientific communities organized around models sometimes convene consensus-seeking bodies to characterize those uncertainties [e.g., the Intergovernmental Panel on Climate Change (IPCC), the Federal Reserve Bank, the Particle Data Group]. Needless misunderstandings may arise when they fail to communicate how they handle uncertainty.

Morgan and Keith's elicitation procedure (22) asked its experts to integrate uncertainty from all sources (22, 29). To that end, it included day-long elicitation sessions with individual experts. Each session began with detailed instructions, discussion of potential judgmental biases, and practice questions, familiarizing participants with the process. Its questions were formulated precisely enough to allow evaluating the judgments in terms of their consistency and accuracy. It used probes designed to help experts assess and express the limits to their knowledge.

A complementary approach asks experts to audit existing studies for their vulnerability to key sources of uncertainty. One such procedure is the Cochrane Collaboration's risk-of-bias assessment for medical clinical trials (32). Fig. 2 adapts that procedure to



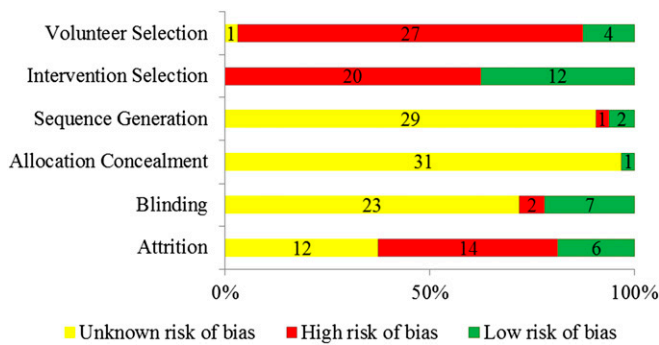


Fig. 2. Methodological flaws in field trials of interventions for reducing home electricity consumption. (Reproduced with permission from ref. 33.)

assess the uncertainty in field trials of programs designed to reduce residential electricity consumption. It shows that most trials had a high risk of biases likely to overestimate the effectiveness and underestimate the uncertainty of the programs being evaluated (33). For example, without blinding, participants' mere knowledge that they are being studied may affect their behavior. Although the possibility of such Hawthorne effects has long been known (34), their size is uncertain. A recent experiment reduced this uncertainty for electricity field trials, finding a 2.7% reduction in consumption during a month in which residents received weekly postcards saying that they were in a study. That Hawthorne effect was as large as the changes attributed to actual programs in reports on other trials (35).

**Conveying Uncertainty.** An audit like that in Fig. 2 creates a profile of the uncertainties surrounding scientific results. Unless scientists provide such summaries, observers must infer them. The well-known heuristics-and-biases research program captures some important patterns in such intuitive statistics (36, 37). Although known for its demonstrations of bias, that research predicts that performance will vary by task. The availability heuristic can produce good estimates when people receive and recall a representative sample of examples (38). The anchoring-and-adjustment heuristic can be useful when an informative value (or anchor) is salient. However, when appearances are deceiving, the same heuristics can produce unwittingly biased judgments, leaving people confidently wrong.

Effective communication is especially important when uncertainty arises from poor data quality (e.g., small samples, weak measurement) (39). Even when people recognize those problems, they may be insufficiently sensitive, as seen in biases such as the base-rate fallacy, belief in the law of small numbers, and insufficiently regressive predictions (37). When uncertainties arise from limits to the science, decision makers must rely on the community of scientists to discover and share problems, so as to preserve the commons of trust that it enjoys (40).

The precision in the summaries that scientists produce for one another provides the foundation for communications that reduce others' need to rely on intuition. As mentioned, meteorologists' precise probability-of-precipitation forecasts are readily understood, once the meaning of "precipitation" is clear (41). It has also been found that most people can extract needed information from drug facts boxes with numerical summaries of risks and benefits and narrative summaries of threats to validity (e.g., a clinical trial's size, duration, and inclusion criteria). With imprecise summaries, however, lay observers are left guessing. For example, there is wide variation in how laypeople interpret the expressions of uncertainty improvised by the IPCC, in hopes of helping nonscientists (42).

### Decisions About Potential Options: What Is Possible?

Some decision makers are neither waiting for a signal to deploy a preselected option nor choosing from fixed options, but, rather,

trying to create ones. Such people need to understand the science relevant to how their world works in order to devise ways to deal with it. Uncertainty is part of what they need to know. Greater uncertainty may prompt them to act sooner (to reduce it) or later (hoping that things become more predictable). When they choose to act, they may wish to create options with more certain outcomes in order to know better what they will get, or less certain ones in order to confuse rivals.

**Characterizing Uncertainty.** Fig. 3 reflects one way to organize the science relevant to creating options, for decisions about drinking water safety (43). In this influence diagram (Fig. 3), the nodes are variables and the links are relationships (44–46); an arrow means that knowing the value of the variable at its tail should influence predictions of the variable at its head. Such models make predictions by simulation. Each model run samples a value for each variable, and then uses them to predict outcomes. The distribution of those predictions is the computed uncertainty about the outcomes.

Although consumers in developed countries may give drinking water safety little thought, they still must monitor their environment for decision points (e.g., "Is my tap water brownish just because of turbidity? "Should I take that 'boil water' notice seriously?"). When such decisions arise, they must create and evaluate response options (e.g., home testing, bottled water). Officials, too, must also create and evaluate options both to remedy problems (e.g., stopping contamination) and to inform consumers. Their confidence in the science will help them to determine when and where to look for better options.

Like scientific theories, models have uncertainties in both their variables and their relationships. For example, predicted storm surges at one location might be based on observations at another point along a coast with varying microclimates and off-shore geomorphology. Those predictions might reflect historical relationships from periods with weaker storms and lower seas (47). Additional uncertainty arises when models omit variables or relationships, whether because they seem unimportant or because the model's formalisms cannot accommodate them. For example, energy models often neglect social factors (e.g., how quickly people adopt new technologies that they oppose). It takes insight and humility to identify these uncertainties. It takes research to assess their impact.

**Assessing Uncertainty.** Assessing uncertainty typically involves running a model with values sampled from probability distributions over the possibilities and seeing how sensitive its predictions are to those uncertainties. If the predicted outcomes are all unattractive, then adding factors may allow for the creation of additional, and

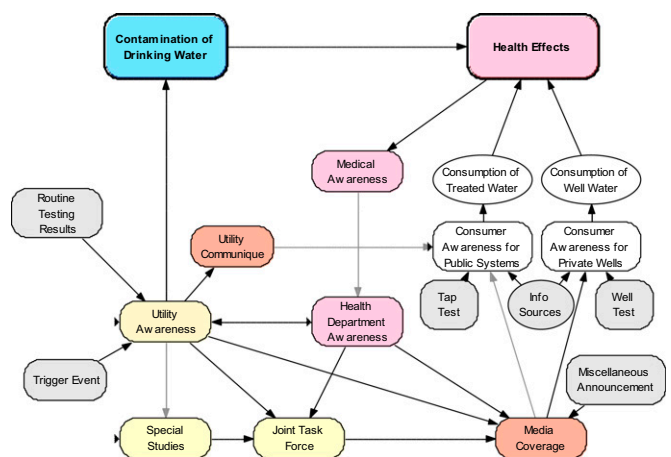


Fig. 3. Influence diagram showing the expertise needed for systematic assessment of the uncertainty in responses to drinking water contamination. (Reproduced with permission from ref. 43.)

better, options. For example, simulations using the model in Fig. 3 found that even the best communications could not reduce the health effects of cryptosporidium contamination because the pathogen was so difficult to detect. However, better outcomes were possible by expanding the model to include options for protecting vulnerable populations (e.g., providing bottled water to people with AIDS). Adding other factors could create options for developing countries with endemic cryptosporidiosis (48).

Just as modelers must assess the uncertainty created by omitting factors from their models, so must scientists assess the uncertainty created by factors that their science typically ignores or takes for granted. Table 1 identifies four common properties of decision-making experiments that could affect the performance observed in them. Where everyday life has those properties, then the research results could be robust. Otherwise, the research might produce “orchids”—exotic, theoretically informative phenomena that are rarely observed in the wild, hence an uncertain foundation for predicting behavior there.

Any discipline could summarize its generic uncertainties. For example, fields that use computational models could assess the uncertainty created by omitting hard-to-quantify factors. Fields that rely on qualitative analysis could assess the uncertainty from such expert judgment. The Numerical Unit Spread Assessment Pedigree (NUSAP) protocol offers a general approach, characterizing any field by its pedigree, giving higher scores to ones that use direct outcome measures, rather than surrogates; have experimental evidence, rather than just statistical relationships; and use widely accepted methods, rather than investigator-specific ones (49, 50).

When research communities exclude relevant disciplines, uncertainty may be seriously underestimated. For example, the landmark Reactor Safety Study (51) assessed physical threats to nuclear power plants but neglected human factors, such as design flaws, lax regulatory regimes, unfriendly user interfaces, and punishing work schedules—all seen at Three Mile Island, Chernobyl, Fukushima, Browns Ferry, and other incidents. Conversely, aviation has reduced uncertainty by addressing human factor problems in instrument design (52) and cockpit team dynamics (53). Decision makers need to know which factors a field neglects and what uncertainty that creates.

**Conveying Uncertainty.** To create options, people need to know the science about how things work. A common formulation for that knowledge is having a mental model of the domain (54–56). Mental models have been studied for topics as diverse as logical relationships, spatial patterns, mechanical systems, human physiology, and risky technologies. The research typically contrasts lay models with expert ones, such as the rules of syllogistic reasoning, schematics of biological system, maps, and analyses like Fig. 3. The contrast shows where people need help and which facts go without saying. Lay beliefs can go astray in each element of a model. People can misunderstand or ignore factors or relationships; they can know current science but not the uncertainty surrounding it; and they can integrate the pieces poorly when the processes are unintuitive (e.g., dynamic, nonlinear) or demand too much mental computation.

Communications designed to improve mental models target specific problems, in order to avoid overwhelming recipients by saying too much or boring (and even insulting) them by repeating

things that they know already. Messages addressing specific uncertainties include “Future storms are so hard to predict that you should be ready for anything,” “Experience with these treatments is so limited that you might wait until we know more,” and “Today’s financial markets move too fast to rely on previously successful trading strategies.”

When the processes are unintuitive, people may need explanations, such as “if you’ve never experienced a rip tide, you can’t imagine its power” (57) or “young women’s breast tissue is too thick for mammography to be useful.” (58). If the processes require difficult mental arithmetic, then the communication may need to “run the numbers” (e.g., showing how small risks accumulate over repeated exposures) (59). If the kind of science is unfamiliar, then science communication may require some science education (e.g., how models work). If the discourse of science is bewildering, then communications may need to explain the role of controversy in revealing and resolving uncertainties, and how passionate theoretical disagreements might have limited practical importance. If scientists tend to overstate their results, then a warning may be needed. As elsewhere, the test of success is empirical: Do people grasp the science well enough to create and evaluate options (60)?

### Eliciting Uncertainty

Science communication is driven by what audiences need to know, not by what scientists want to say. Thus, it is unlike communicating with students, who must accept curricular definitions of what matters. However, relevance poses challenges. Mastering the uncertainties of many decisions forces laypeople to become experts in everything. Addressing the needs of many decision makers forces scientists to become experts in everyone.

One way to reduce that load is by adopting standard reporting formats, which experts can learn to create and decision makers can learn to use. Subjective probability distributions are one such format, with seemingly growing acceptance. Signal detection theory could be another, were the research conducted to make its quantitative estimates ( $d'$ ,  $\beta$ ) as clear as its concepts (discrimination ability, caution). Influence diagrams offer a representation with the precision needed for analyzing a problem and a graphic format for conveying its gist (18).

Table 2 offers a standard protocol for eliciting and reporting expert assessments of scientific uncertainty. It is designed to require little additional effort from scientists, by making explicit concerns that are already on their minds, in terms that should be familiar.

**Variability.** All measurement has variability, arising from variations in procedure (e.g., how long a thermometer is left in), phenomena (e.g., how body temperature changes diurnally), and measured individuals (e.g., how much they run “hot” or “cold”). Scientists routinely estimate variability. Routinely sharing those estimates should cost little, while freeing decision makers from having to guess them. Numerical estimates are needed because verbal quantifiers (e.g., “stable measurement,” “widely varying”) communicate poorly to people unfamiliar with scientists’ conventional usage of the terms (61–63). The  $\pm X\%$  format used in survey research, and required by its professional organizations (64), is one example.

**Internal Validity.** Scientists ask common questions when evaluating studies. The protocol reports their assessments for the uncertainties

**Table 1. Four common properties of decision-making experiments and their potential effects on participants’ performance**

i)	The tasks are clearly described: can produce better decisions, if it removes the clutter of everyday life, or worse decisions, if it removes vital context, such as the choices that other people are making.
ii)	The tasks have low stakes: can produce better decisions, if it reduces stress, or worse decisions, if it reduces motivation.
iii)	The tasks are approved by university ethics committees: can produce better decisions, if it reduces participants’ worry about being deceived, or worse decisions, if it leads to artificiality.
iv)	The tasks focus on researchers’ interests: can produce better decisions, if researchers seek decision makers’ secrets of success, or worse decisions, if researchers are committed to documenting biases.

**Table 2. A protocol for summarizing scientific uncertainty, illustrated in the context of medical clinical trials**

Step <i>i</i>	Identify key outcomes for decision makers (e.g., stroke) and how to measure them (e.g., annual probability).
Step <i>ii</i>	Summarize variability
Step <i>iii</i>	Summarize internal validity
Selection bias	Do the initial groups differ from randomly assigned ones? Were the groups drawn from same population, over same time periods, and with the same inclusion and exclusion criteria?
Attrition bias	Do the final groups differ from the initial ones? Did the groups differ as a result of which participants dropped out (e.g., because the treatment did not seem to be working or their lives were too disorderly to continue) or were excluded from analyses (e.g., for incomplete data or seemingly anomalous responses)?
Administration	Was the study conducted as intended? Were instructions followed in administering the treatment and analyzing the results?
Performance bias	Does the manipulation have unintended effects? Were participants affected by knowing that they were in the study (or in a study), perhaps trying to satisfy researchers' (real or perceived) expectations?
Step <i>iv</i>	Summarize external validity
Population bias	Do treatment groups differ from the general population? Might they be relatively sensitive to positive effects or to unintended side effects?
Intervention bias	Are treatments administered differently in different conditions? Might they be applied less consistently, intensively, or obviously?
Control group bias	Do untreated groups differ in other ways? Might they receive more (or less) of other treatments with more (or less) supervision?
Scenario bias	Do other conditions differ from those of the study? Might other factors diminish (or enhance) the treatment's effect? Might the world have changed?
Step <i>v</i>	Summarize the strength of the basic science
Directness	How well do a field's measures capture key outcomes? The strongest sciences measure outcomes directly rather than relying on proxy measures (e.g., biomarkers that appear related to health states).
Empirical basis	How strong are the best available estimates? The strongest sciences base their theories on large, well-controlled experiments rather than on datasets that are small or collected under variable conditions (e.g., dose-response relationships derived from epidemiological data).
Methodological rigor	How strong are the best methods? The strongest sciences have methods with well-understood strengths and weaknesses, and extensive experience in their application.
Validation	How well are theoretical results confirmed? The strongest sciences have foundations (theories, hypotheses, relationships) that are strongly confirmed by evidence from multiple, independent sources.
Step <i>vi</i>	Summarize uncertainty (e.g., 95% credible interval) Statements of the form, "Considering the variability of the evidence (step <i>ii</i> ) and my assessments of the internal validity of the studies that collected it (step <i>iii</i> ), their relevance to the decision-making domain (step <i>iv</i> ), and the strength of the underlying science (step <i>v</i> ), I am 95% certain that the true value of the critical outcome (step <i>i</i> ) is between Y and Z."

Steps *iii* and *iv* are based on CONSORT criteria for evaluating medical clinical trials (31, 62). Step *v* is based on the NUSAP criteria for evaluating the strength of sciences (47) Ref. 9 summarizes research regarding step *vi*.

that have been found to pose the greatest threats to the validity of medical clinical trials (32, 65–67). Although expressed in terms of experimental research, these threats have analogs in observational studies. For example, selection bias arises when individuals are not randomly assigned to experimental groups. In observational studies, the equivalent is ignoring confounds (e.g., the effects of cohort differences on correlations between age and behavior).

**External Validity.** Decision makers need to know how confidently they can extrapolate results from the contexts that scientists have studied to the ones that interest them. Sometimes scientists have evidence that can limit the uncertainty (e.g., "although the study involved women, gender differences are rare with such behaviors"). Sometimes, they can say little (e.g., "no one has ever studied men"). The protocol in Table 2 proposes reporting on four such threats.

**Strength of Science.** The protocol adopts the NUSAP framework (49) for four aspects of a field's pedigree. The best fields have

strong theoretical foundations supported by robust experimental methods and converging results. The weakest ones have poorly understood observational data. The weaker a field, the greater its uncertainty.

**Credible Intervals.** Quantitative summaries of uncertainty take the form, "I am XX% certain that the true value is between Y and Z." A credible interval is the same size as the confidence interval when scientists consider only observed variability. It is wider when scientists question those observations or the strength of the underlying science. It is narrower when scientists have enough confidence in the basic science to discount anomalous observations. It can be higher or lower if scientists perceive bias (e.g., clinical trials designed to get the maximum effect; climate change estimates made conservatively).

Unless there are negligible threats to the internal validity, external validity, and pedigree of scientific results, credible intervals should differ in size from confidence intervals. Nonetheless, many scientists are uncomfortable providing them. Table 3



**Table 3. Frequently asked questions addressing four concerns of scientists reluctant to express their uncertainty in credible-interval form**

Concern 1	If I give credible intervals, people will misinterpret them, inferring greater precision than I intended.
Response	Behavioral research has found that most people (i) like receiving explicit quantitative expressions of uncertainty (such as credible intervals), (ii) can interpret them well enough to extract their main message, and (iii) misinterpret verbal expressions of uncertainty (e.g., “good” evidence, “rare” side effect). For audiences that receive the reports created with the protocol (Table 2), understanding should be greater if they receive credible intervals than if they have to infer them (63).
Concern 2	People cannot use probabilities.
Response	Behavioral research has found that laypeople can often provide reasonably consistent probability judgments if asked clear questions and extract needed information if provided with well-designed displays (41, 60, 74). Whether they do so well enough to satisfy their decision-making needs is an empirical question, which should be answered with evidence rather than speculation.
Concern 3	My credible intervals will be used unfairly in performance evaluations.
Response	Such judgments can protect experts from unfair evaluations, unjustly accusing them of having conveyed too much or too little confidence, especially when supported by the rationale for those judgments. The protocol provides such protection—if the experts’ management stands behind it.
Concern 4	People do not need such judgments.
Response	Decision makers must act with some degree of uncertainty. Not helping them means preferring the risk of having them guess incorrectly over the risk of expressing oneself poorly.

lists four common reasons and potential responses (68). Decision makers who receive a full report (Table 2, steps *i-v*) might be able to infer the credible interval that it implies. Whether they can is an empirical question.

### Conclusion

Communicating uncertainty requires identifying the facts relevant to recipients’ decisions, characterizing the relevant uncertainties, assessing their magnitude, drafting possible messages, and evaluating their success. Performing these tasks demands commitment from scientists and from their institutions. It also demands resources for the direct costs of analysis, elicitation, and message development, and for the opportunity costs of having scientists spend time communicating uncertainty rather than reducing it (through their research). Making this investment means treating communication as part of scientists’ professional responsibility and rewarding them for strengthening the public goodwill that science needs (40).

This investment would be more attractive if it advanced science, as well as making it more useful. A standard reporting protocol might help do that. Although assessing uncertainty is at the heart of all science, such protocols are not (67, 68). Indeed, the Consolidated Standards of Reporting Trials (CONSORT) scheme, reflected in Table 2 and Fig. 3, was a response to the inconsistent reporting of medical clinical trials. The FDA has adopted a standard format for summarizing the evidence and uncertainties underlying its drug approval decisions (70). The

drug facts box uses a similar strategy (41). The open-data movement’s reporting standards seek to record scientific uncertainty and reduce the file-drawer problem, whereby scientists report results that affirm their hypotheses while finding reasons to discard inconsistent ones. Such practices add inestimable uncertainty to published science by obscuring its capitalization on chance (71–73).

For scientists, uncertainty obscures theoretical questions. For people who rely on science, uncertainty obscures choices. Those awaiting a signal for action need to know whether the evidence is certain enough to pass the threshold defined by their decision rule. Those choosing among fixed options need to know how far to trust predictions of valued outcomes. Those creating options need to know how well the processes shaping their outcomes are understood.

Table 2 offers a protocol for summarizing those uncertainties. For scientists, it should require little additional work, merely asking them to report, in standard form, judgments that they make already. For decision makers, it should make uncertainties easier to grasp by putting them in a common format. For communicators, it should allow economies of scope in creating ways to address recurrent issues. The result should be better science and better decisions.

**ACKNOWLEDGMENTS.** Preparation of this manuscript was supported by the National Science Foundation (SES-0949710).

- Fischhoff B, Kadavy J (2011) *Risk: A Very Short Introduction* (Oxford Univ Press, Oxford).
- Lichtenstein S, Slovic P, eds (2006) *The Construction of Preferences* (Cambridge Univ Press, New York).
- von Winterfeldt D (2013) Bridging the gap between science and decision making. *Proc Natl Acad Sci USA* 110(Suppl 3):14055–14061.
- Baron J, Hershey JC (1988) Outcome bias in decision evaluation. *J Pers Soc Psychol* 54(4):569–579.
- Fischhoff B (1975) Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *J Exp Psych Human Perc Perf* 1(3):288–299.
- Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics* (Wiley, New York).
- Mohan D, et al. (2012) Sources of non-compliance with clinical practice guidelines in trauma triage: A decision science study. *Implement Sci* 7:103.
- American College of Surgeons (2006) *Resources for Optimal Care of the Injured Patient* (ACS, Chicago).
- O’Hagan A, et al. (2006) *Uncertain Judgements: Eliciting Expert Probabilities* (Wiley, Chichester, UK).
- Ericsson A, Simon HA (1993) *Verbal Protocols as Data* (MIT Press, Cambridge, MA).
- Murphy AH, Lichtenstein S, Fischhoff B, Winkler RL (1980) Misinterpretations of precipitation probability forecasts. *Bull Am Meteorol Soc* 61(7):695–701.
- National Research Council (1982) *Survey Measure of Subjective Phenomena* (National Academy Press, Washington).
- Macintyre S, West P (1993) ‘What does the phrase “safer sex” mean to you?’ Understanding among Glaswegian 18-year-olds in 1990. *AIDS* 7(1):121–125.
- National Oceanic and Atmospheric Administration (2013) *Hurricane/Post-Tropical Cyclone Sandy October 22–29, 2012* (NOAA, Washington).
- Kahneman D (2011) *Thinking, Fast and Slow* (Farrar Giroux & Strauss, New York).
- Fischhoff B (1992) Giving advice. Decision theory perspectives on sexual assault. *Am Psychol* 47(4):577–588.
- Fischhoff B (2007) Nonpersuasive communication about matters of greatest urgency: Climate change. *Environ Sci Technol* 41(21):7204–7208.
- Reyna VF (2012) A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgm Decis Mak* 7:332–339.
- National Institutes of Health (1997) Breast cancer screening for women ages 40–49. Consensus Statement Online January 21–23, 1997 15(1):1–35.
- US Preventive Services Task Force (2009) Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 151(10):716–726.
- Woloshin S, et al. (2000) Women’s understanding of the mammography screening debate. *Arch Intern Med* 160(10):1434–1440.
- Raiffa H (1968) *Decision Analysis* (Addison-Wesley, Reading, MA).
- Howson C, Urbach P (1989) *Scientific Reasoning: The Bayesian Approach* (Open Court, Chicago).
- Morgan MG, Keith DW (1995) Subjective judgments by climate experts. *Environ Sci Technol* 29(10):468A–476A.

25. Gärdenfors P, Sahlin N (1982) Unreliable probabilities, risk taking, and decision making. *Synthese* 53(3):361–386.
26. Edwards W, Lindman H, Savage J (1963) Bayesian statistical inference for psychological research. *Psychol Rev* 70(3):193–242.
27. Gelman A, Shalizi CR (2013) Philosophy and the practice of Bayesian statistics. *Br J Math Stat Psychol* 66(1):8–38.
28. Seidenfeld T (1979) Why I am not an objective Bayesian: Some reflections prompted by Rosenkrantz. *Theory Decis* 11(4):413–440.
29. Armstrong JS (2010) *Persuasive Advertising* (Palgrave Macmillan, New York).
30. Poulton EC (1994) *Behavioral Decision Making* (Lawrence Erlbaum, Hillsdale, NJ).
31. Morgan MG (2014) Use (and abuse) of expert elicitation in support of decision making for public policy. *Proc Natl Acad Sci USA* 111(20):7176–7184.
32. Higgins JPT, Altman DG, Sterne JAC, eds (2011) Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Updated March 2011)*, The Cochrane Collaboration 2011, eds Higgins JPT, Green S, Chap 8. Available at [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Accessed May 4, 2014.
33. Davis AL, Krishnamurti T, Fischhoff B, Bruine de Bruin W (2013) Setting a standard for electricity pilot studies. *Energy Policy* 62:401–409.
34. Roethlisberger FJ, Dickson WD (1939) *Management and the Worker* (Harvard Univ Press, Cambridge, MA).
35. Schwartz D, Fischhoff B, Krishnamurti T, Sowell F (2013) The Hawthorne effect and energy awareness. *Proc Natl Acad Sci USA* 110(38):15242–15246.
36. Kahneman D, Slovic P, Tversky A, eds (1982) *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge Univ Press, New York).
37. Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.
38. Tversky A, Kahneman D (1973) Availability: A heuristic for judging frequency and probability. *Cognit Psychol* 5(2):207–232.
39. Kahneman D, Tversky A (1973) On the psychology of prediction. *Psychol Rev* 80: 237–251.
40. Fiske ST, Dupree C (2014) Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proc Natl Acad Sci USA* 111:13593–13597.
41. Schwartz LM, Woloshin S (2013) The Drug Facts Box: Improving the communication of prescription drug information. *Proc Natl Acad Sci USA* 110(Suppl 3):14069–14074.
42. Budescu DV, Broomell SB, Por HH (2009) Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychol Sci* 20(3): 299–308.
43. Casman EA, Fischhoff B, Palmgren C, Small MJ, Wu F (2000) An integrated risk model of a drinking-water-borne cryptosporidiosis outbreak. *Risk Anal* 20(4):495–511.
44. Burns WJ, Clemen RT (1993) Covariance structure models and influence diagrams. *Manage Sci* 39(7):816–834.
45. Fischhoff B, Bruine de Bruin W, Guvenc U, Caruso D, Brilliant L (2006) Analyzing disaster risks and plans: An avian flu example. *J Risk Uncertain* 33(1):133–151.
46. Howard RA (1989) Knowledge maps. *Manage Sci* 35(8):903–922.
47. Wong-Parodi G, Strauss BH (2014) Team science for science communication. *Proc Natl Acad Sci USA* 111:13658–13663.
48. Casman E, et al. (2001) Climate change and cryptosporidiosis: A qualitative analysis. *Clim Change* 50(1-2):219–249.
49. Funtowicz SO, Ravetz J (1990) *Uncertainty and Quality in Science for Policy* (Kluwer, London).
50. Bammer G (2008) *Uncertainty and Risk*, ed Smithson M (Earthscan, London).
51. Nuclear Regulatory Commission (1974) *Reactor Safety Study* (NRC, Washington), WASH-1400.
52. Wickens CD, McCarley JS (2008) *Applied Attention Theory* (CRC Press, Boca Raton, FL).
53. Salas E, Burke CS, Bowers CA, Wilson KA (2001) Team training in the skies: Does crew resource management (CRM) training work? *Hum Factors* 43(4):641–674.
54. Bruine de Bruin W, Bostrom A (2013) Assessing what to address in science communication. *Proc Natl Acad Sci USA* 110(Suppl 3):14062–14068.
55. Johnson-Laird PN (2006) *How We Reason* (Oxford Univ Press, New York).
56. Morgan MG, Fischhoff B, Bostrom A, Atman C (2001) *Risk Communication: The Mental Models Approach* (Cambridge Univ Press, New York).
57. Webster GD, Agdas D, Masters FJ, Cook CL, Gesselman AN (2013) Prior storm experience moderates water surge perception and risk. *PLoS ONE* 8(5):e62477.
58. Silverman E, et al. (2001) Women's views on breast cancer risk and screening mammography: A qualitative interview study. *Med Decis Making* 21(3):231–240.
59. Cohen J, Chesnick EI, Haran D (1971) Evaluation of compound probabilities in sequential choice. *Nature* 232(5310):414–416.
60. Fischhoff B (2013) The sciences of science communication. *Proc Natl Acad Sci USA* 110 (Suppl 3):14033–14039.
61. Cohen J, Dearnley EJ, Hansel CEM (1958) A quantitative study of meaning. *Br J Educ Psychol* 28:141–148.
62. Fillenbaum S, Rapoport A (1971) *Structures in the Subjective Lexicon* (Academic, New York).
63. Erev I, Cohen BL (1990) Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organ Behav Hum Decis Process* 45(1):1–18.
64. American Association of Public Opinion Research (2010) *Standards and Ethics*. Available at [www.aapor.org/Standards\\_and\\_Ethics/5102.htm](http://www.aapor.org/Standards_and_Ethics/5102.htm). Accessed May 4, 2014.
65. Mahon WA, Daniel EE (1964) A method for the assessment of reports of drug trials. *Can Med Assoc J* 90(9):565–569.
66. Mosteller F, Gilbert JP, McPeck B (1980) Reporting standards and research strategies for controlled trials: Agenda for the editor. *Control Clin Trials* 1(1):37–58.
67. Moher D, et al. (2010) CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ* 340:c869.
68. Fischhoff B (2012) Communicating uncertainty: Fulfilling the duty to inform. *Issues Sci Technol* 28(4):63–70.
69. PBL Netherlands Environmental Assessment Agency (2012) *Guidance for Uncertainty Assessment* (PBL, Amsterdam), 2nd Ed.
70. Food and Drug Administration (2013) *Structured Approach to Benefit-Risk Assessment for Drug Regulatory Decision Making* (FDA, Washington).
71. Cohen J (1962) The statistical power of abnormal-social psychological research: A review. *J Abnorm Soc Psychol* 65(1):145–153.
72. Nosek B, Bar-Anan Y (2012) Scientific utopia: Opening scientific communication. *Psych Inquiry* 23(3):217–243.
73. Davis AL, Fischhoff B (2014) Communicating uncertain experimental evidence. *J Exp Psychol Learn Mem Cogn* 40(1):261–274.
74. Fischhoff B, Brewer N, Downs JS, eds (2011) *Communicating Risks and Benefits: An Evidence-Based User's Guide* (Food and Drug Administration, Washington, DC).