Posted November 4, 2015

Public perceptions of expert disagreement:

Bias and incompetence or a complex and random world?

Nathan F Dieckmann[1,2], Branden B Johnson[2], Robin Gregory[2], Marcus Mayorga[2,3], Paul KJ Han[4], Paul Slovic[2,3]

Corresponding author: Nathan Dieckmann, (541) 485-2400, fax (541) 485-2403, dieckman@ohsu.edu

[1] Oregon Health & Science University
3375 SW Terwilliger Boulevard
Portland, OR 97239

[2] Decision Research
1201 Oak Street, Suite 200
Eugene, OR 97401

[3] University of Oregon
1585 E 13th Ave
Eugene, OR 97403

[4] Center for Outcomes Research and Evaluation
Maine Medical Center
39 Forest Avenue
Portland, ME 04105

http://pus.sagepub.com/content/early/2015/09/04/0963662515603271.abstract?rss=1

Abstract

Expert disputes can present laypeople with several challenges including trying to understand why such disputes occur. In an online survey of the U.S. public, we used a psychometric approach to elicit perceptions of expert disputes for 56 forecasts sampled from seven domains (climate change, crime, economics, environment, health, politics, terrorism). People with low education, or with low self-reported knowledge of the topic, were most likely to attribute expert disputes to expert incompetence. People with higher self-reported knowledge tended to attribute disputes to expert bias due to financial or ideological reasons. The more highly educated and cognitively able were most likely to attribute disputes to natural factors, such as the irreducible complexity and randomness of the phenomenon. We highlight several important implications of these results for scientists and risk managers and argue for further research on how people perceive and grapple with expert disputes.

Keywords:

Expert disagreement, forecasting, public beliefs, attribution

Public perceptions of expert disagreement:

Bias and incompetence or a complex and random world?

Expert disputes are common within many scientific domains. This is particularly true of expert forecasts regarding health, environmental, economic, and socio-political topics. Although experts may perceive these disagreements to be part of the normal scientific process, members of the lay public are likely to draw quite different inferences, particularly if scientific authority stems in part from its perceived consensus among experts (Collingridge & Reeve, 1986). It is important that we understand public reactions to publicized expert disputes to design better communication strategies among scientists and laypeople, to select appropriate risk management responses (Hoffman, Fischbeck, Krupnick, & McWilliams, 2007), and to more fully predict the relationship between scientific authority and perceptions of uncertainty in the public domain (e.g., Stilgoe, 2007). In this paper, we use a psychometric approach to examine public perceptions of expert disagreement across a diverse sample of forecasting topics.

Why do experts disagree?

Einhorn (1974) viewed consensus as a necessary feature of expertise itself. If several so-called experts present sharply conflicting forecasts on a topic this would be grounds for doubting whether they are experts at all. Expert disagreement from this traditional perspective is the result of incompetence (i.e., they are not experts) or either intentional or unintentional bias due to ideology, worldviews, or private interests (Hammond, 1996). However, there have been several alternative views of expert disagreement that focus on contributing structural and functional factors (Shanteau, 2000) and the fundamental limits of human judgment (Mumpower & Stewart, 1996). Shanteau (2000) argued that expert disagreement should be expected due to the ill-structured, complex, dynamic, uncertain, and evolving nature of real-world problems.

Scientists often report that they expect to disagree with each other and see this as a means by which important gaps in knowledge are identified. In this sense, disagreement is a part of the normal scientific process and it is expected that different experts will think about a problem differently, particularly problems that are complex and involve scientific uncertainty (Mumpower & Stewart, 1996). For instance, experts may have fundamental disagreements about the causes of phenomenona, the analytic methods that should be used, and the interpretation of existing empirical data. A study of expert views on foodborne illness in the United States, for example, distinguished between situations in which experts are highly confident in their own best estimates but disagree with each other and situations in which experts disagree with each other but also are uncertain about their own estimates (a not uncommon situation that decision makers would find especially problematic; Hoffman et al, 2007). From this perspective, disagreements may be considered to be completely natural among even the most competent and unbiased experts. Yet this general tolerance for disagreement may not be shared, at least to the same degree, among members of the public. And to the extent that disagreements are due to factors not related to science (e.g., bias or personal profit), scientists also may take a dim view of disputes.

Public perceptions of why experts disagree

The lay public is at a disadvantage in making sense of expert disputes. In many cases, they are only presented with public communications and have virtually no way of knowing the actual causes or magnitude of expert disagreements (Collins & Evans, 2007). However, just because the public does not have insider information does not mean that they will withhold judgment when confronted with conflicting expert predictions. Table 1 lists several different

causal inferences that members of the lay public may make when encountering expert disagreements.

Inferences 1–3 describe different fundamental sources of uncertainty. Complexity and randomness map onto what Hacking (1975) calls aleatory uncertainty, which describes irreducible uncertainty that is a feature of the world itself. Laypeople making these inferences regard the nature of the world as the primary problem and not the ability, competence or honesty of the experts. The third inference (experts lack knowledge) maps roughly to the concept of epistemic uncertainty, which is due to incomplete knowledge about phenomena that is theoretically attainable (Hacking, 1975). Thus, epistemic uncertainty can be reduced (e.g., through the scientific process) while aleatory uncertainty cannot. Inferences 4-6 are focused on the characteristics or actions of the experts. Inferring expert incompetence and/or bias from disagreements constitutes the traditional view discussed above. Laypeople may also infer that apparent expert disagreement is artificially magnified due to experts being unwilling to admit uncertainty (inference #6). Since this relates to how forecast are communicated, this inference could be classified as a type of linguistic uncertainty (Regan, Colyvan & Burman, 2002).

Outside of a few specific contexts, there have been surprisingly few empirical investigations of public perceptions of the causes of expert disagreements. Johnson and Slovic (1998) interviewed members of the U.S. public and asked for reactions after the presentation of a range of risk estimates for an environmental carcinogen. Participants most often responded that the expert disagreement was due to self-interest or expert incompetence. In another study, a U.S. sample read a statement about expert disagreement (e.g., “Often experts disagree over the size of an industrial health or safety risk”) and were asked about the causes of such disputes. Fifty-five percent of the sample selected all of the choices available (expert incompetence, self-interest,

and lack of scientific knowledge), with 18% choosing self-interest, and 14% choosing lack of scientific knowledge (Johnson, 2003). In a Finnish interview study on food additives, experts were deemed to disagree due to, in descending order, general difficulty in obtaining scientific knowledge, self-interest, and competence and knowledge differences across experts (Kajanne & Pirttilä-Backman, 1999). There was also a moderating effect of education, with less educated participants favoring the difficulty-in-obtaining-scientific-knowledge explanation and the more educated favoring the self-interest explanation.

These results are insightful concerning some of the reasons for the range of public perceptions of disputes but are mainly drawn from interviews and are limited to specific examples of expert disagreement concerning health and safety risks. In the present study, we focus on expert forecasts across a range of domains and examine other possible causal inferences from expert disagreements as well as several additional moderators of these perceptions.

Individual differences and public perceptions of expert disagreement

Perceptions of expert disagreements may also be moderated by the skills and knowledge of the individual evaluating the dispute. For instance, individuals with more education and greater cognitive resources likely have richer mental models about the normal process of science (Rabinovich & Morton, 2012) and the limitations of scientific knowledge and forecasting. Thus, we might expect that this subgroup of the public would be more likely to attribute expert disagreement to the fundamental challenges of making predictions in complex, real-world systems. However, the qualitative results of Kajanne & Pirttilä-Backman (1999), discussed above, provide evidence against this prediction in that it was the less educated that favored the lack-of-scientific-knowledge explanation and the more educated that favored the expert self-

interest explanation. In this present study, we will examine this moderator hypothesis again in a quantitative fashion using a robust measure of cognitive ability.

A second potential moderator is the familiarity with or knowledge about the domain. As above, we may expect that individuals who have more insight into how science and forecasting is conducted in a particular domain may be more likely to blame the world (complexity, randomness) and less likely to blame the expert (incompetence and/or bias). Members of the public with less knowledge about a domain may tend to attribute expert disagreement to the more traditional causes of incompetence and bias.

The present study

We used a psychometric survey approach to examine public perceptions of a range of forecast topics spanning seven broad domains. The psychometric approach has been used to great effect in several studies focused on public perceptions of hazards (Slovic, 1987). We tested the extent to which the lay public attributes expert disagreement to natural (aleatory) causes (irreducible complexity and randomness), expert (epistemic) causes (lack of knowledge, expert incompetence, expert bias) or to communication failures (expert unwillingness to admit uncertainty). We also stratified our sample to examine perceptions of subgroups of individuals based on cognitive ability/education and self-reported knowledge about the forecasting domain under study. The results from this study will expand our understanding of public reactions to expert disputes and provide insights into how scientific results and forecasts can be better communicated to different subgroups of the public.

Methods

Sample. Participants were randomly drawn from the Decision Research web panel subject pool. Forty participants completed the study in less than 10 minutes and were removed leaving *N*

= 342 (57% female). The mean age of the sample was 45 years (range = 22–76 yrs). Approximately 26% had a high school education or less, 31% attended some college or vocational school, 27% were college graduates, and 16% had advanced degrees.

Design and measures. Forecast topics. We generated 56 forecast/prediction topics that lay people might encounter in newspapers, online, or on television (see Appendix A). The forecast topics were derived from seven different domains (Health, Politics, Terrorism, Climate Change, Economics, Crime, & Environment) with eight forecast topics per domain. Within each domain, there were four binary topics (will event happen or not?) and four continuous topics (prediction of continuous quantity—e.g., gross domestic product). We also varied the time horizon of the forecast topics at four levels: Short (6 months), Medium (5 years), Long (15 years), and Very long (50 years). The goal of this design aspect was to have a roughly equal mix of time horizons across the domains. There were no significant effects of forecast type (binary or continuous) or time horizon on participant ratings and we do not discuss these manipulations further. In this study, we focused on general societal level forecasts (as opposed to personal forecasts in medicine or finance, for example) and did not include "hard" science domains like physics or astronomy. Results from our studies using these latter types of topics will be reported in future papers.

Procedure. The study was conducted online. Participants were first presented with a general introduction: "In this study, we are interested in your perceptions of different kinds of forecasts or predictions that are made by experts in a given field. You may encounter these types of forecasts on television, in newspaper articles or other written materials, or online." Each participant was then presented with seven different forecast topics, one pseudo-randomly drawn from each of the seven domains. The order of the seven selected forecasts was also randomized.

Participants made each rating on all seven forecast topics before moving on to the next question. This procedure resulted in approximately n = 40 participants providing all ratings for each of the 56 forecast topics.

Measures. Participants rated each forecast topic on seven different constructs (see Appendix B): Expert disagreement (three items), Complexity (one item), Irreducible Randomness (one item), Expert knowledge (two items), Expert credibility/competence (two items), Expert bias (affected by private interests or personal ideology; two items), and Expert willingness to admit uncertainty or lack of knowledge (two items). They also indicated their level of general knowledge in the domain for each of the forecasts they rated.

In an earlier study session, participants completed an eight-item numeracy measure (Weller et al., 2013) and two fluid reasoning tests (Dieckmann et al., 2014). These fluid reasoning tests were modeled after the number and letter series completion tasks commonly used in IQ tests. The number series test had eight items (ex. What number comes in the blank: 2, 3, 5, 7, 11, 13, _____) and the letter series test had seven items of increasing difficulty (ex. What letter comes in the blank: A, E, I, M, Q, U, _____).

Analytic approach. For each of the 56 forecasts, we calculated the mean rating on each of the seven model constructs across all of the participants that rated each forecast. The analysis was focused at the level of these mean forecast ratings (N = 56). We used multiple linear regression to identify which of the six predictors (randomness, complexity, expert knowledge, credibility/competence, bias, admit uncertainty) were most strongly related to ratings of expert disagreement across the set of forecasts. One inherent difficulty of this approach is deciding which predictors should remain in a given model and which should be removed as relatively unimportant. The traditional solution to this problem is to use forward or backward step-wise

model selection. Problems with these approaches include the reliance on arbitrary thresholds for p-values and that backward and forward step-wise methods will not always result in the same final model (Venables & Ripley, 1997). Alternative approaches, not based on p-values, use information criterion (IC) as measures of support for a given model. For instance, all possible models from a given set of predictors can be fit and ranked by their relative support (IC). In addition, this approach of fitting all possible models allows for multi-model inference as opposed to just focusing on the single "best" model. This approach also allows the assessment of predictor importance across all possible models (Buckland, Burnham, & Augustin, 1997).

Model tests were conducted with an automated model selection and multi-model inference approach implemented in the glmulti package for the R statistical computing environment (R Development Core Team, 2013; Calgano & Mazancourt, 2010). The Bayesian Information Criterion (BIC) was used to assess fit and compare alternative models, where smaller BICs indicate superior fit after accounting for sample size and model complexity (Raftery, 1995). All possible explanatory models were fit and the "best" model was determined by the smallest BIC.[1] The relative support of this "best" model was then compared to the other plausible models. Raftery (1995) reports the following rules of thumb for comparing models with respect to the BIC: Weak evidence = BIC diff 0–2; Positive evidence = BIC diff 2–6; Strong evidence = BIC diff 6–10; Very strong evidence = BIC diff > 10. Thus, any model that has less than a two unit BIC difference as compared to the best model (approximately) should be considered plausible. We also assessed predictor importance in a multi-model fashion by examining the importance (or support) for each predictor averaged across all possible models (for details see Calgano & Mazancourt, 2010).

[1] We report the results for all possible models that do not include any multiplicative (or interaction) effects. Additional analyses were conducting allowing multiplicative effects but none of the multiplicative terms received strong support. Thus, we report the results that do not include such terms for simplicity.

The analyses above were conducted for the full sample as well as stratified by cognitive ability and self-reported knowledge in the forecast domain. A median split was used to separate participants higher and lower in cognitive ability. Mean ratings on each construct for each forecast were then created separately for each group. Self-reported knowledge in the forecast domain was rated by each participant separately for each forecast presented. Again, a median split was used to separate the mean level forecast data for people who self-reported very little to little knowledge (rating of two or less on the five-point scale) and those who reported some to a lot of knowledge about a forecast domain (rating of three or higher on five-point scale).

Results

Psychometrics. Ratings on the three expert disagreement items were strongly correlated across forecast topics in all seven domains (αs = .67–.77) and were averaged to create a single expert disagreement score. The expert knowledge (*r*s = .56–.67), expert credibility/competence (*r*s = .59–.77), bias (*r*s = .30–.43), and willingness to admit uncertainty or lack of knowledge (*r*s = .56–.64) items were all moderately to strongly correlated across forecast topics in all seven domains and were averaged to create single construct scores.

The numeracy, number completion, and letter completion tasks all showed acceptable reliability (Kuder-Richardson αs = .70, .80, .68, respectively). The average inter-scale correlation between numeracy, number completion and letter completion was $r = .65$ (range .58–.68). Treating each of the scales as an item of a single cognitive ability construct resulted in acceptable internal consistency ($\alpha = .84$; scale range 0–23). Scores from these three scales were summed to create a broad cognitive ability scale.

Moderator characteristics. As expected, there was a significant linear relation between level of education and cognitive ability scores (HS or less $M = 9.05$; some college $M = 10.69$;

College $M = 12.04$; Graduate degree $M = 14.16$; $F(3, 330) = 12.23$, $p < .001$). We examined whether there was a relation between cognitive ability and self-report ratings of knowledge in the domain for the participants that rated each forecast. Cognitive ability was not significantly related to ratings of self-reported knowledge in any of the seven forecast domains, suggesting that examining these moderators separately may result in unique information about the perceptions of these different subpopulations.

Predictors of perceived expert disagreement. Analyzing the mean level ratings from the full sample resulted in 13 different candidate models within 0.2–2 BIC units from each other. Thus, we had weak evidence for the superiority of any given model and there was a high amount of uncertainty regarding variable importance. This may be due to the mix of subpopulations within the full sample. For this reason, we focused on the stratified subgroup models below which resulted in more interpretable findings.

Higher cognitive ability. The best predictive model of expert disagreement for the higher cognitive ability group included complexity (beta = .21, 95% CI = .05, .37), randomness (beta = .18, 95% CI = .05, .32), and expert bias (beta = .20, 95% CI = .05, .36) as unique, significant predictors (Adjusted $R^2 = .46$, BIC = –58.30). Higher levels of perceived expert disagreement were associated with higher perceived complexity and randomness in the domain and higher perceived expert bias due to ideology and/or private interests. The BIC difference between this model and the next best model (with complexity, bias, and expert knowledge as predictors) was 1.9, suggesting very close to positive evidence for the superiority of the best model. Figure 1 shows the model-averaged importance of terms. The bar graph shows the estimated importance of each predictor as the summed IC weight from all models in which the term appears. Values closer to 1.0 indicate more importance or weight for a predictor.

Lower cognitive ability. The best predictive model of perceived expert disagreement for the lower cognitive ability group included only expert credibility/competence (beta = –.35, 95% CI = –.61, –.10; Adjusted R^2 = .11, BIC = –10.02). Higher perceived expert disagreement was associated with lower belief in expert credibility/competence. The BIC difference between this model and the next best model was 2.20 (positive evidence for the superiority of the best model; see Figure 1).

Higher self-reported knowledge in forecast domain. The best predictive model of expert disagreement for the higher self-reported knowledge in forecast domain group included only expert bias as a significant predictor (beta = .37, 95% CI = .20, .54; Adjusted R^2 = .25, BIC = –59.64). Higher levels of perceived expert disagreement were associated with higher perceived expert bias due to ideology and/or private interests. The BIC difference between this model and the next best model (with bias and randomness as predictors) was 1.8, suggesting very close to positive evidence for the superiority of the best model (see Figure 3).

Lower self-reported knowledge in forecast domain. The best predictive model of expert disagreement for the lower self-reported knowledge in forecast domain group included only expert credibility/competence as a significant predictor (beta = –.41, 95% CI = –.61, –.19; Adjusted R^2 = .20, BIC = –30.05). As with those of lower cognitive ability (above), higher perceived expert disagreement was associated with lower belief in expert credibility/competence. There were two models with a BIC difference of ~1.0 from this best model. The first model included expert knowledge as a predictor in addition to expert credibility/competence and the second included randomness in addition to expert credibility/competence. This suggests that we have only weak evidence for the superiority of the best model, although expert

credibility/competence is the strongest predictor in each of these alternative models (see Figure 4).

Discussion

Our results suggest that most members of the public will attribute expert disputes to expert incompetence and/or expert bias and self-interest. Expert incompetence was the strongest predictor of perceived expert disagreement for participants with less education/cognitive ability and those that self-reported having little to no knowledge about a forecasting domain. These results are in contrast to earlier, mostly interview, studies that found the lack-of-scientific-knowledge attribution to be the leading explanation for expert disagreement among participants with lower education levels.

Contrary to expectations, participants who reported having moderate to high general knowledge about a domain were not more likely to attribute expert disagreement to natural (aleatory) causes (irreducible complexity and randomness). The strongest predictor was expert bias or self-interest. Only the more educated/high cognitive ability group attributed expert disagreement to irreducible complexity and randomness, although expert bias was still a strong predictor for this group as well. This differed from the Finnish results (Kajanne & Pirttilä-Backman, 1999), in which the more educated stressed the bias explanation for expert disputes on food additives.

There are several potential reasons for the divergence of results between our study and previous studies on this topic. The first is that the psychometric approach is a more indirect and expressly quantitative method for examining the relations between public perceptions. Unlike interview and direct survey questions, the participants were not aware of the goals of this study, making it less likely that they would respond in a socially desirable way or that they would

respond affirmatively to all possible causes even when in doubt. We see this as a strength of our study and a good complement to prior work that used more direct questions. However, because the current analyses are all correlational, we cannot determine the direction of the causal arrow. Additional experimental work should be conducted to follow up on this and prior results.

A second potential reason for divergence from prior results is the broad range of topics we examined using this approach. Previous research focused on specific topics (e.g., risks of food additives), potentially limiting the generalizability of the results. Third, we focus specifically on expert *forecasting* at varying time horizons, as opposed to statements about disputed facts (e.g., food additives cause cancer). Members of the public may have fundamentally different perceptions of whether or how experts answer what-will-happen questions as compared to what-is questions.

Implications. From a theoretical perspective, these findings suggest that it may be useful to explore the role of a positivist view of science (e.g., Steel, List, Lach and Shindler, 2004; Rabinovich & Morton, 2012) in the response to expert disputes, particularly among those with lower education, cognitive ability and self-reported knowledge. For this group of participants, attribution of expert disputes to expert incompetence may be the result of a belief that science is objective and certain, and therefore, any disagreement must be an indication of faulty experts.

The attribution of expert disputes to bias—a strong explanation for those who see themselves as having moderate to high knowledge of the topic, and somewhat less so for the high education/cognitive ability group—implies a more sophisticated view of science as being socially constructed and thus (for better or worse) subject to influence from financial or ideological interests. However, it is unclear from these results whether this viewpoint implies a crude, deliberate skewing of forecasts to fit one's biases, or a more subtle model of the bias in

expert forecasts. Further research is again needed to identify and understand the importance of relevant moderators. For example, given that the two factors are not correlated, perhaps those who believe themselves to be knowledgeable are more prone to believe in deliberate bias in scientific estimates, while those with high cognitive ability might be more prone to believe in unconscious bias.

Only the most educated, cognitively able people in our sample selected irreducible complexity and randomness of the topic area as an explanation for expert disputes, outweighing even their co-attribution to expert bias. As this is an explanation also used by scholars and experts themselves (see Introduction), it is tempting to attribute this judgment to this group having the most sophisticated view of science and expertise generally. We do not dispute this interpretation, but note that such sophistication need not rule out other attributions: besides the co-attribution to expert bias found here, we observe that scientists such as gravitation wave physicists use a diverse set of both crude (e.g., nationality) and subtle cues to judge the credibility of other researchers in their field (e.g., Collins & Evans, 2007, pp. 50–51, note 10). Research to assess whether lay attributions for expert disputes over forecasts are similar to their attributions for expert disputes over descriptions of current conditions or causal relationships also is needed. Such convergence seems likely but data to confirm that hypothesis are warranted.

As for practical implications, under a deficit model of public understanding of science (e.g., Wynne, 1991) the findings that most people underweight the role of complexity and randomness in making predictions, and that even relatively sophisticated lay observers favor a bias explanation for expert disputes, will foster suggestions for communication and education strategies to counteract these perceptions. For example, one potential strategy would be to embed simple epistemological education within communications to reinforce concepts like randomness,

complexity and limitations in our ability to know. The findings may suggest a need for audience segmentation—i.e., use of different interventions for different segments of lay society—but otherwise do not alter the deficit-based interpretation.

A more "public engagement"-focused model would suggest interactions between experts and laypeople to broaden the range of explanations both "upward" and "downward" in sophistication, rather than emphasize only the natural explanations. For example, the less-educated might learn to conceive of positivist science as characterizing (at most) "settled science" and add the notion of bias to their repertoire, while the highly (self)-knowledgeable might add irreducible complexity and randomness to their repertoire without abandoning the bias attribution. Experts might learn to think of bias as an at least plausible explanation for expert disputes beyond "it explains my opponents' positions," and thus not an unwarranted stance for laypeople to take when observing expert disputes.

Regardless of which, if either, "educational" model is pursued, a larger question for the public is how to assess the credibility of contending scientists. This study has identified a range of lay attributions as to when and why experts might disagree over forecasts in diverse fields, but whether and how these beliefs affect which of the disputing parties seems most credible is yet to be studied systematically. It is likely that diverse cues to credibility will be deployed (Collins & Evans, 2007), but neither their relative use nor potential moderators (e.g., cognitive ability) of such use have been examined.

Conclusion. Multiple difficulties face the lay public in trying to evaluate or engage with uncertain information about the state of the world or forecasted events and actions. Unlike the relatively straightforward questions about how people might interpret or use ranges of risk estimates or forecasts to make choices (Dieckmann, Slovic, & Peters, 2009; Dieckmann, Peters,

Gregory, & Tusler, 2012), understanding expert disputes or even divergences in forecasts poses several serious challenges. Is this dispute relevant to my life? Who am I to believe? What does it mean, to me and/or to society, that consensus cannot be reached on this topic? Our study shows, across a diverse set of domains, that laypeople tend to use coherent—albeit overly narrow—attributions to make sense of why expert disputes occur and that these explanations are likely to vary predictably across different segments of the population. Further research can expand our understanding of how people cope with the challenges posed by such disputes, the implications of their responses for risk managers, and the degree to which public explanations, credibility cues, and other interpretations of disputes serve to enhance or hinder relations between science and society.

-

Acknowledgments

We would like to acknowledge the generous support of the National Science Foundation: This material is based upon work supported by NSF under Grant Nos. #1231231 (Robin Gregory, PI; Nathan Dieckmann co-PI) and #0925008 (Nathan Dieckmann, PI) to Decision Research. All views expressed in this paper are those of the authors alone.

References

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics, 53*, 603–618.

Calgano, V., & de Mazancourt, C. (2010). glmulti: An R package for each automated model selection with (generalized) linear models. *Journal of Statistical Software, 32*(12), 1–29.

Collingridge, D., & Reeve, C. (1986). *Science speaks to power: The role of experts in policy-making*. London: Frances Pinter.

Collins, H., & Evans, R. (2007). *Rethinking expertise*. IL: University of Chicago Press.

Dieckmann, N. F., Slovic, P., & Peters, E. (2009). The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. *Risk Analysis, 29*, 1473–1488. doi: 10.1111/j.1539-6924.2009.01279.x

Dieckmann, N. F., Peters, E., Gregory, R., & Tusler, M. (2012). Making sense of uncertainty: Advantages and disadvantages of providing an evaluative structure. *Journal of Risk Research, 15*, 717–735. doi: 10.1080/13669877.2012.666760

Dieckmann, N. F., Gregory, R., & Peters, E. (2014). *Seeing what you want to see: How imprecise uncertainty ranges enhance motivated cognition*. Manuscript submitted for publication.

Einhorn, J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology, 59*, 562–571.

Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. UK: Cambridge University Press.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Hoffman, S., Fischbeck, P., Krupnick, A., & McWilliams, M. (2007). Elicitation from large, heterogeneous expert panels: Using multiple uncertainty measures to characterize information quality for decision analysis. *Decision Analysis, 4*, 91–109.

Johnson, B. B., & Slovic, P. (1998). Lay views on uncertainty in environmental health risk assessment. *Journal of Risk Research, 1*, 261–279.

Johnson, B. B. (2003). Further notes on public response to uncertainty in risks and science. *Risk Analysis, 23*, 781–789.

Kajanne, A., & Pirttilä-Backman, A.-M. (1999). Laypeople's viewpoints about the reasons for expert controversy regarding food additives. *Public Understanding of Science, 8*, 303–315.

R Development Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rabinovich, A., & Morton, T. A. (2012). Unquestioned answers or unanswered questions: Beliefs about science guide responses to uncertainty in climate change in risk communication. *Risk Analysis, 32*, 992–1002.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.

Shanteau, J. (2000). Why do experts disagree? In B. Green, R. Cressy, F. Delmar, T. Eisenberg, B. Howcraft, M. Lewis, D. Schoenmaker, J. Shanteau & R. Vivian (Eds.), *Risk behaviour and risk management in business life* (pp. 186–196). Dordrecht, The Netherlands: Kluwer Academic Press.

Slovic, P. (1987). Perception of risk. *Science, 236*, 280–285.

Steel, B., List, P., Lach, D., & Shindler, B. (2004). The role of scientists in the environmental policy process: A case study from the American West. *Environmental Science and Policy, 7*, 1–13.

Stilgoe, J. (2007). The (co-)production of public uncertainty: UK scientific advice on mobile phone health risks. *Public Understanding of Science, 16*, 45–61.

Venables, W. N., & Ripley, B. D. (1997). *Modern applied statistics with S-PLUS* (3rd ed.). New York, NY: Springer-Verlag.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making, 26*, 198–212. doi: 10.1002/bdm.1751

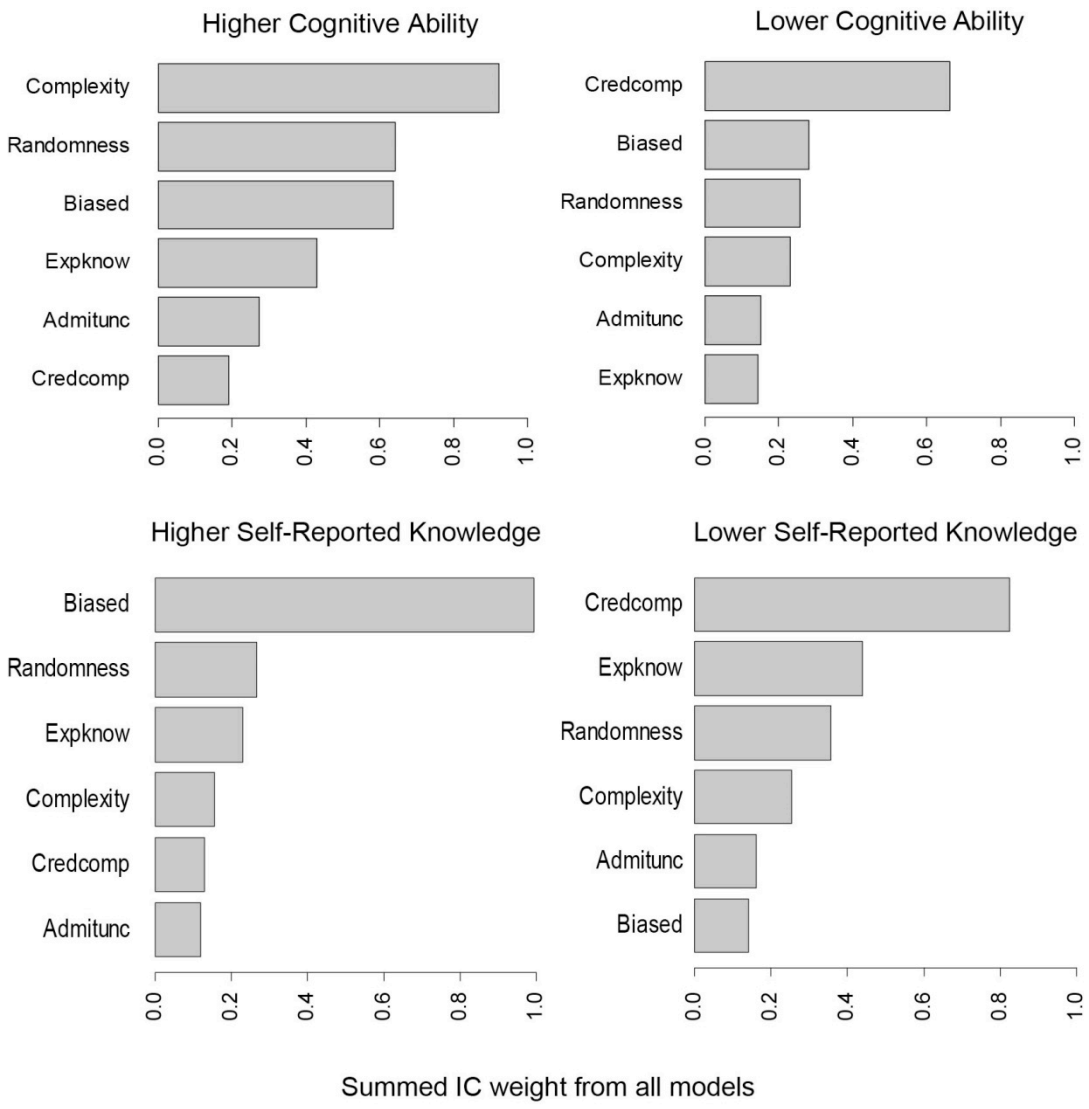
Wynne, B. (1991). Knowledges in context. *Science, Technology and Human Values, 16*, 111–121.

Table 1

Possible Inferences of the Lay Public About Expert Disagreements

Causal Inference	***Description***: Experts disagree because . . .
1. Too much complexity in domain	. . . making predictions is very difficult in complex, chaotic systems with a large number of diverse interrelated components.
2. Too much randomness in domain	. . . making predictions is very difficult in domains where events have a lot of fundamental unpredictability or "randomness".
3. Experts lack knowledge	. . . they have not yet acquired enough scientific knowledge about the causes of the event.
4. Experts are incompetent	. . . they are incompetent and are not really "experts" at all.
5. Experts are biased	. . . one or more experts are intentionally or unintentionally biasing their conclusions due to ideology, worldviews, or private interests.
6. Experts are unwilling to admit uncertainty	. . . they are not willing to admit uncertainty and are providing simplistic overly precise forecasts.

Figure 1. Model averaged importance of predictors in each subgroup model.



Appendix A. Forecasts

HEALTH

1.The number of confirmed human cases of bird influenza 6 months from now.

2.Whether the cost of U.S. health care per person will be lower 6 months from now.

3.Whether the world ranking of the U.S. with respect to life expectancy will be lower 5 years from now.

4.The number of adults in the United States who will be clinically obese 5 years from now.

5.Whether heart disease will still be the leading cause of death 15 years from now.

6.The suicide rate in the U.S. 15 years from now.

7.Whether there will be a cure for most types of cancer 50 years from now.

8.The relative world ranking with regards to public health of the U.S. 50 years from now.

POLITICS

1.Whether Immigration laws in the U.S. will be revised significantly within the next 6 months.

2.The likelihood that the Republican party will change its primary election rules to exclude Tea Party candidates in the next 6 months.

3.Whether the Affordable Care Act (Obamacare) will still be law 5 years from now.

4.The number of U.S. states that will permit concealed handgun carry in 5 years.

5.Whether Puerto Rico will become a U.S. state within the next 15 years.

6.The percentage of eligible Americans who vote in the Presidential elections 15 years from now.

7.Whether Israel will cease to exist as a separate country within the next 50 years.

8.The number of countries in the world that will have democratic governments 50 years from now.

TERRORISM

1.Whether terrorists will succeed in downing a commercial airliner in the next 6 months.

2.The number of terrorist bombings in Afghanistan in the next 6 months.

3.Whether Israel will launch a military assault against terrorist cells in Syria in the next 5 years.

4.The number of people living in refugee camps in Somalia 5 years from now.

5.Whether terrorists will succeed in poisoning a drinking water reservoir in the United States in the next 15 years.

6.The annual number of deaths worldwide due to terrorist attacks 15 years from now.

7.Whether there will be fewer international terrorist events 50 years from now.

8.The proportion of the U.S. budget directed toward homeland security 50 years from now.

CLIMATE CHANGE

1.Whether progress will be made on an agreement to reduce total carbon emissions across the globe within the next 6 months.

2.The percentage of U.S. ski resorts forced to close over the next 6 months due to lower snowfalls and higher winter temperatures.

3.Whether there will be an increase in the number of extreme weather events 5 years from now.

4.The number of deaths in the U.S. from heat stroke in 5 years.

5.Whether climate change will improve U.S. economic growth in 15 years.

6.The average sea level rise along U.S. coasts 15 years from now.

7.Whether the city of Miami, Florida will disappear entirely under water within 50 years.

8.The percentage of current animal species that will become extinct due to climate change within 50 years.

ECONOMICS

Predicting …

1.Whether sales in the U.S. housing market will be higher 6 months from now.

2.The U.S. unemployment rate 6 months from now.

3.Whether the value of the Dow Jones stock market index will be above 20,000 5 years from now.

4.The average mortgage interest rate in the U.S. in 5 years.

5.Whether Microsoft will overtake Apple in sales of tablets and smartphones in 15 years.

6.The ratio of the Gross Domestic Product (GDP) of India to that of the United States in 15 years.

7.Whether China will have a higher standard of living, measured as average yearly income, than Canada 50 years from now.

8.The percentage change in the energy efficiency (Gross Domestic Product [GDP] divided by total energy use) of the U.S. in 50 years.

CRIME

Predicting …

1.Whether "stop-and-frisk" policing will still be legal in New York City 6 months from now.

2.The average number of bicycle thefts in a major metropolitan city 6 months from now.

3.Whether the use of hand-held cell phones while driving will be banned in all states 5 years from now.

4.The average annual number of murders in the United States in 5 years.

5.Whether drug-related murders in Mexico will be lower in 15 years.

6.The number of U.S. states that will have legalized marijuana in 15 years.

7.Whether the U.S. demand for illegal drugs will be lower 50 years from now.

8.The violent crime rate (per 1000 citizens) in the U.S. 50 years from now.

ENVIRONMENT

Predicting…

1.Whether laws protecting endangered species in the United States will be significantly weakened by Congress within the next 6 months.

2.The percentage of total U.S. energy supplied by natural gas produced by fracking in 6 months.

3.Whether removing dams from the Columbia River will result in larger salmon populations 5 years from now.

4.The total amount of money spent by businesses to comply with the U.S. Environmental Protection Agency's enforcement of stream protection regulations in 5 years.

5.Whether tough laws to halt cutting of the Amazonian rainforest will be passed within the next 15 years.

6.China's overall levels of air pollution 15 years from now.

7.Whether the Bengal tiger will become extinct in the next 50 years.

8.The number of Bison (American buffalo) present on the western prairies 50 years from now.

Appendix B. Study Measures

Construct	Item	Scale
Expert disagreement	How often do you think experts disagree about the causes of each of these events?	1 = *very rare*– 5 = *very frequent*
	How often do you think experts disagree about what methods should be used to forecast each of these events?	1 = *very rare*– 5 = *very frequent*
	How often do you think experts disagree about how to interpret the results from forecasts?	1 = *very rare*– 5 = *very frequent*
Complexity	How much complexity do you think there is in the processes that generate each of these events? A more complex system is one with a large number of diverse interrelated components that exhibits behavior that cannot be predicted by just looking at the individual components.	1 = *very little*– 5 = *A lot*
Randomness	How much randomness do you think there is in the processes that generate each of these events? A system with more "randomness" is one that is more fundamentally unpredictable by its very nature. This fundamental unpredictability cannot be reduced by gaining more knowledge about how the system works.	1 = *very little*– 5 = *A lot*
Expert knowledge	How much do you think experts understand about the causes of each of these events?	1 = *very poor*– 5 = *very good*
	How much do you think experts know about the interactions among all of the factors that contribute to making each of these events occur?	1 = *very poor*– 5 = *very good*
Credibility/competence	How credible do you think the experts are that typically make these types of forecasts?	1 = *not at all*– 5 = *very credible*
	How competent do you think the experts are that typically make these types of forecasts?	1 = *not at all*– 5 = *very competent*

Construct	Item	Scale
Willing to admit uncertainty	To what extent do you think that the experts that typically make these types of forecasts are willing to admit uncertainty?	1 = *not willing*– 5 = *very willing*
	To what extent do you think that the experts that typically make these types of forecasts are willing to admit lack of knowledge?	1 = *not willing*– 5 = *very willing*
Bias	To what extent do you think that the experts that typically make these types of forecasts are affected by private interests?	1 = *not affected*– 5 = *very affected*
	To what extent do you think that the experts that typically make these types of forecasts are affected by their own ideology and beliefs?	1 = *not affected*– 5 = *very affected*
Self-reported knowledge	How much general knowledge do you have about each of these fields?	1 = *very little*– 5 = *A lot*