Theoretical/research paper

P | U | S

Public Understanding of Science
1–14

DOI: 10.1177/0963662515603271
pus.sagepub.com
SAGE

Public perceptions of expert disagreement: Bias and incompetence or a complex and random world?

Nathan F. Dieckmann
Oregon Health & Science University, USA; Decision Research, USA

Branden B. Johnson
Decision Research, USA

Robin Gregory
Decision Research, USA

Marcus Mayorga
Decision Research, USA; University of Oregon, USA

Paul K. J. Han
Maine Medical Center, USA

Paul Slovic
Decision Research, USA; University of Oregon, USA

Abstract
Expert disputes can present laypeople with several challenges including trying to understand why such disputes occur. In an online survey of the US public, we used a psychometric approach to elicit perceptions of expert disputes for 56 forecasts sampled from seven domains. People with low education, or with low self-reported topic knowledge, were most likely to attribute 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. Our results show that laypeople tend to use coherent—albeit potentially overly narrow—attributions to make sense of expert

Corresponding author:
Nathan F. Dieckmann, Oregon Health & Science University, 3375 SW Terwilliger Boulevard, Portland, OR 97239, USA.
Email: dieckman@ohsu.edu

disputes and that these explanations vary across different segments of the population. We highlight several important implications for scientists, risk managers, and decision makers.

Keywords
attribution, expert disagreement, forecasting, public beliefs

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 and 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 et al., 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 article, 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 alternative views of the sources of expert disagreement, citing the fundamental limits of human judgment (Mumpower and Stewart, 1996), or the complex, uncertain, and evolving nature of real-world problems that make disagreements among experts inevitable (Shanteau, 2000).

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 and Stewart, 1996). For instance, experts may have fundamental disagreements about the causes of phenomena, 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 two kinds of expert disagreements: those where experts are highly confident in their own best estimates versus those in which they 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

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

magnitude of expert disagreements (Collins and 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. These inferences are drawn from the literature (as noted above) but are not intended to constitute an exhaustive list of all possible inferences.

Inferences 1–3 describe different fundamental sources of uncertainty. Complexity and randomness both contribute to what Hacking (1975) called aleatory uncertainty, irreducible uncertainty that is a feature of the world itself. Laypeople making these inferences would regard the nature of the world as the primary problem, rather than 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 currently incomplete but theoretically attainable knowledge about phenomena (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. One possible inference is that uncertainty stems from either incompetence or bias on the part of the experts. Laypeople may also infer that apparent expert disagreement is artificially magnified due to experts being unwilling to admit uncertainty (inference no. 6). This inference could be classified as a type of linguistic uncertainty because it relates to how forecasts are communicated (Regan et al., 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 US 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 sample of the US public read a statement on expert disputes (“Often experts disagree over the size of an industrial health or safety risk”) and were asked to indicate the causes of such disputes. Overall, 55% 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 and 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 offer insight into some reasons for the range of public perceptions of disputes but are mainly drawn from interviews and limited to specific examples of expert disagreement concerning health and safety risks. In this 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. A first possible moderator is that individuals with more education and greater cognitive resources are likely to have richer mental models about the normal process of science (Rabinovich and 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 and Pirttilä-Backman (1999) provide evidence against this prediction in that the less educated respondents favored the lack-of-scientific-knowledge explanation and the more educated favored the expert self-interest explanation. In this study, we will examine this moderator hypothesis again in a quantitative fashion, using a robust measure of cognitive ability.

A second potential moderator is one's familiarity with or knowledge about the domain. Individuals who have more insight into how science and forecasting are 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. In this study, due to our interest in assessing lay perceptions of expert disagreement across a wide variety of topics and domains (i.e. across 56 forecast topics in seven different domains), assessing objective knowledge in each domain would have been prohibitive and we therefore examine self-reported knowledge as a potential moderator.

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 numerous studies focused on public perceptions of hazards (Slovic, 1987). This approach involves having participants rate a large number of targets (e.g. hazards and forecast topics) on a large number of rating scales and then using multivariate analyses to discover patterns in the responses. The approach has the advantage of allowing the study of many different topics and decreases the chances that participants are responding in socially desirable ways as its goals are not readily apparent. We tested the extent to which the lay public attributes expert disagreement to three main factors: (1) natural (aleatory) causes that center on irreducible complexity and randomness; (2) expert-based (epistemic) causes such as lack of knowledge, incompetence, and bias; or (3) communication factors that include experts' unwillingness to admit uncertainty.

We stratified our sample to examine perceptions of subgroups of individuals based on their 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 web panel subject pool. This panel comprises roughly 1500 members across the United States, with members recruited via online forums and Google Ads regarding paid survey opportunities, and paid at a rate of US$15 per hour, prorated for survey length. While the panel is not recruited to be representative of the US population, it is highly diverse in age, ethnicity, income, gender, political orientation, and worldviews. The panel is continually maintained for quality by a Web Survey Coordinator.

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 years). 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.

Forecast topics

We generated 56 forecast/prediction topics that laypeople might encounter in newspapers, online, or on television (see Appendix A, available at: http://pus.sagepub.com/content/by/supplemental-data). These general societal-level forecast topics (e.g. in contrast to personal forecasts in medicine or finance) were derived from seven different domains (health, politics, terrorism, climate change, economics, crime, and environment) with eight forecast topics per domain. We did not include "hard" science domains like physics or astronomy; results from our studies using these types of topics will be reported in future papers. Below are two forecast examples from the climate change and terrorism domains: "The average sea level rise along US coasts 15 years from now" and "Whether terrorists will succeed in downing a commercial airliner in the next 6 months." 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.

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 onto the next question. This procedure resulted in $n=39$–47 participants providing all ratings for each of the 56 forecast topics.

Measures

In a pilot study, we tested a range of potential survey items to measure our constructs of interest. Simple psychometric analyses (e.g. Cronbach's alpha) were used to identify the strongest items, which were then used in this study. Participants rated each forecast topic on seven different constructs (see Appendix B, available at: http://pus.sagepub.com/content/by/supplemental-data): 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 self-reported 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., Submitted). The fluid reasoning tests were modeled after the number and letter series completion tasks commonly used in intelligence quotient (IQ) tests. The number series test had eight items (e.g. What number comes in the blank: 2, 3, 5, 7, 11, 13, _______) and the letter series test had seven items of increasing difficulty (e.g. 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 (see section "Measures") across all of the participants who rated that 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 stepwise model selection. Problems with these approaches include the reliance on arbitrary thresholds for p-values and that backward and forward stepwise methods will not always result in the same final model (Venables and Ripley, 1997). Alternative approaches, not based on p-values, use information criterion (IC) to measure support for a given model. The IC approach allows all possible models from a given set of predictors to be fit and ranked by their relative support, providing multi-model inference as opposed to just focusing on the single "best" model and allowing assessment of predictor importance across all possible models (Buckland et al., 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 (Calgano and De Mazancourt, 2010; R Development Core Team, 2013). 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, and 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 and De Mazancourt, 2010). We also examined residual plots for each final model to confirm model adequacy.

These analyses 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 2 or less on the 5-point scale) and those who reported some to a lot of knowledge about a forecast domain (rating of 3 or higher on 5-point scale). Due to the need for aggregation at the topic level to account for each respondent rating only 7 of the 56 forecasts, such median splits were the only feasible option for examining moderator effects. Our goal for this study was to examine perceptions of expert disagreement in general, so we did not focus on any between-domain differences (e.g. terrorism and climate change).[2]

Results

Psychometrics

As suggested by our pilot work, we created composite construct scores by averaging individual items that measured the same construct and at least moderately correlated with each other. The goal of using composite scores is to create better multi-item measures of our constructs and to reduce the dimensionality of the data to facilitate analysis. Ratings on the three expert disagreement items were strongly correlated across forecast topics in all seven domains (Cronbach's alphas = .67–.77) and were averaged to create a single expert disagreement score. The expert knowledge (rs = .56–.67), expert credibility/competence (rs = .59–.77), bias (rs = .30–.43), and willingness to admit uncertainty or lack of knowledge (rs = .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, and .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 (high school 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 who 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% confidence interval (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, second panel).

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 1, third panel).

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 1, fourth panel).

Discussion

Our results suggest that most members of the public within the subgroups that we studied will attribute the existence of expert disputes primarily to expert incompetence and/or expert bias and self-interest. Expert incompetence was the strongest predictor of perceived expert disagreement for

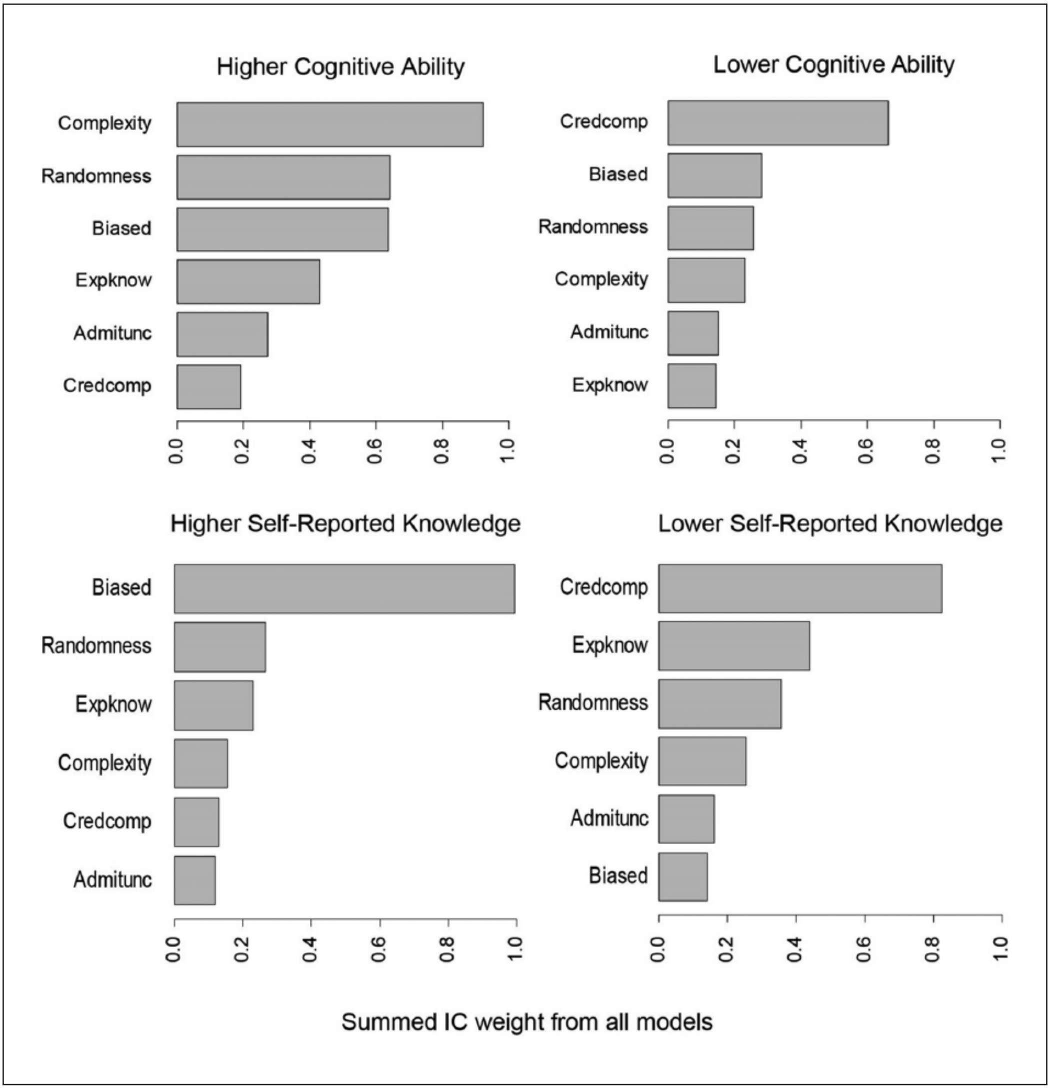

Figure 1. Model-averaged importance of predictors in each subgroup model.

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, if one assumes that self-reported knowledge tracks objective knowledge, participants who reported 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 for this group 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 results obtained in Finland (Kajanne and Pirttilä-Backman, 1999), in which the more educated participants emphasized a 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 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 approach 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 and field work should be conducted to follow-up on these issues.

A second potential reason for divergence from prior results is the broad range of topics we examined using this approach. Previous research has 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.

Other limits include uncertainty about the extent to which our survey results would generalize to public attributions of expert disputes "in the wild" and the possibility that we did not include all possible inferences for participants to choose from. Our measure of self-reported knowledge was also very general such that we could not distinguish between the types of knowledge people felt they had (e.g. lots of scientific facts vs social knowledge about scientific institutions). Future work should include more detailed measures of self-reported scientific knowledge and, to the extent possible, complement self-reports with more objective knowledge measures.

Implications

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, that any disagreement among experts must be an indication of faulty experts. From a theoretical perspective, these findings suggest that it may be useful to explore the role of a lay positivist view of science in the response to expert disputes, particularly among those with lower education, cognitive ability, and self-reported knowledge (e.g. Rabinovich and Morton, 2012; Steel et al., 2004).

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 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 role of bias in expert forecasts. 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. Further research is suggested to identify and understand the importance of relevant moderators.

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 often used by scholars and experts themselves, it is tempting to attribute this judgment to this group having a more sophisticated view of science and expertise generally. We do not dispute this interpretation but note that it 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 and Evans, 2007: 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.

As for practical implications, under a deficit model of public understanding of science (e.g. Wynne, 1991), the finding that most of our subgroups underweight the role of complexity and randomness in making predictions, and that even relatively sophisticated lay observers include 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 such as randomness, complexity, and inherent limitations in knowledge. The findings also may suggest a need for audience segmentation—that is, the use of different interventions for different segments of lay society—but otherwise do not alter the deficit-based interpretation.

A model focused more on "public engagement" might suggest interactions between experts and laypeople to broaden the range of explanations both "upward" and "downward" in sophistication, rather than emphasizing only the natural explanations. For example, from this perspective, the less educated participants 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.

Critics of the deficit model might argue against any such educational effort on the grounds that laypeople may have situation-specific knowledge or justified skepticism about experts' willingness to acknowledge (or even recognize) uncertainty that leads them to doubt that experts will agree on domain forecasts.[3] In other words, it might be misleading to assume that we need to improve "scientific literacy" about randomness, complexity, and the rest of the inferences about scientific disputes that we have explored here. We do not dispute these assertions, but suggest that they are beside the point for the specific issue we have probed, which is not whether laypeople are always right about the degree of disagreement among experts or on which (if any) of the disputing experts are correct.

Given the wide variety of reasons for disputes among experts—who in specific cases can be incompetent, biased, struggling with the randomness and complexity of the universe, or all three—laypeople who only emphasize one or two reasons for experts to disagree will often be wrong to some degree. This means that at best they will have a misleading view of the world and may engage in non-optimal behaviors as a result of this view. Certainly, if laypeople disagree among themselves as to how to explain expert disputes, they may be less likely to work together in addressing any related challenges, thus reducing their power vis-à-vis these same experts.

Regardless of which "educational" model is proposed, a larger question for members of 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 are yet to be studied systematically. It is likely that diverse cues to credibility will be deployed (Collins and 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 et al., 2009, 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 potentially 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.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. We report the results for all possible models that do not include any multiplicative (or interaction) effects. Additional analyses were conducted 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.
2. Preliminary analyses suggested that there were, in fact, few domain-level differences in these perceptions; these will be reported as part of later publications.
3. In this context, we note that whether experts will acknowledge uncertainty was *not* a major explanation of disputes for any of our subgroups.

References

Buckland ST, Burnham KP and Augustin NH (1997) Model selection: An integral part of inference. *Biometrics* 53: 603–618.

Calgano V and 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 and Reeve C (1986) *Science Speaks to Power: The Role of Experts in Policy-Making*. London: Frances Pinter.

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

Dieckmann NF, Gregory R and Peters E (Submitted) Seeing what you want to see: How imprecise uncertainty ranges enhance motivated cognition.

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

Dieckmann NF, Slovic P and Peters E (2009) The use of narrative evidence and explicit likelihood by decision makers varying in numeracy. *Risk Analysis* 29: 1473–1488.

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*. Cambridge: Cambridge University Press.
Hammond KR (1996) *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York, NY: Oxford University Press.
Hoffman S, Fischbeck P, Krupnick A and 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 BB (2003) Further notes on public response to uncertainty in risks and science. *Risk Analysis* 23: 781–789.
Johnson BB and Slovic P (1998) Lay views on uncertainty in environmental health risk assessment. *Journal of Risk Research* 1: 261–279.
Kajanne A and Pirttilä-Backman A-M (1999) Laypeople's viewpoints about the reasons for expert controversy regarding food additives. *Public Understanding of Science* 8: 303–315.
Mumpower JL and Stewart TR (1996) Expert judgement and expert disagreement. *Thinking and Reasoning* 2: 191–211.
R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Rabinovich A and Morton TA (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 AE (1995) Bayesian model selection in social research. *Sociological Methodology* 25: 111–163.
Regan HM, Colyvan M and Burgman MA (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12: 618–628.
Shanteau J (2000) Why do experts disagree? In: Green B, Cressy R, Delmar F, Eisenberg T, Howcraft B, Lewis M, et al. (eds) *Risk Behaviour and Risk Management in Business Life*. Dordrecht: Kluwer Academic Press, pp. 186–196.
Slovic P (1987) Perception of risk. *Science* 236: 280–285.
Steel B, List P, Lach D and 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 WN and Ripley BD (1997) *Modern Applied Statistics with S-PLUS*, 3rd edn. New York, NY: Springer-Verlag.
Weller JA, Dieckmann NF, Tusler M, Mertz CK, Burns WJ and Peters E (2013) Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making* 26: 198–212.
Wynne B (1991) Knowledges in context. *Science, Technology & Human Values* 16: 111–121.

Author biographies

Nathan F. Dieckmann is an Assistant Professor at the Oregon Health & Science University and a Research Scientist at Decision Research. He conducts basic and applied research in a range of fields including decision making, risk communication, and statistical methodology. His current work is focused on the development of decision aids and methods for the effective presentation of uncertainty in a variety of domains.

Branden B. Johnson (PhD, Geography, Clark University) specializes in risk perception and risk communication research related to technological, natural and social hazards, and environmental issues. Current research includes how laypeople interpret and respond to disputes among scientists; Americans' beliefs, attitudes, and behaviors regarding Ebola; and whether and how stereotypes of institutions (government, business, nonprofits) affect public trust in specific organizations managing hazards.

Robin Gregory works on applied problems of stakeholder consultation, environmental risk management, judgment under uncertainty, and negotiated decision making using methods drawn from decision analysis and behavioral psychology.

Marcus Mayorga (MS) is a Doctoral Student at the University of Oregon and Research Coordinator at Decision Research. His main research field is charitable decision making and altruism. He also works in diverse fields such as risk, emotion, individual differences, and moral psychology. His current work is focused on how charitable motivations and decisions are affected by individual differences in emotion regulation and universalism.

Paul K. J. Han (MD, MA, MPH) is a Health Services Researcher and Physician, and Director of the Center for Outcomes Research and Evaluation at the Maine Medical Center Research Institute. His main research interests are in risk communication and shared decision making, and his work focuses on understanding and improving the communication and management of uncertainty in health care.

Paul Slovic is a President of Decision Research and a Professor of psychology at the University of Oregon. He studies human judgment, decision making, and risk perception and has published extensively on these topics. His most recent work examines "psychic numbing" and the failure to respond to mass human tragedies.