The Epistemic Contract: Fostering an Appropriate Level of Public Trust in Experts

Robert J. MacCoun*

Stanford University

Scientia nihil aliud est quam veritatis imago. (Science is but an image of the truth.)
Francis Bacon[1]

Back off man. I'm a scientist.
Bill Murray, playing Dr. Peter Venkman, in the movie *Ghostbusters*.[2]

These quotes nicely capture traditional views about the authority of science. Like Ghostbuster Dr. Peter Venkman, practicing scientists would like to wear the mantle of authority that the word "Science" conveys in post-Enlightenment culture. Like Bacon, we can try to justify that authority by invoking the idea that we are able to speak truth – or rather, that when we act as scientists, we are a pure lens that allows truth to shine *through* us.

In the postmodern era, this Baconian (or Mertonian [1938], or Venkmanian) view of science still has some currency (as we shall see), but it has lost some of its luster, and perhaps appropriately so. This essay might have been titled "Fostering Public Trust in Experts," but that wording implies that (a) trust in experts is lower than it

---

* Prepared for the Nebraska Symposium on Motivation. I thank Brian Bornstein, John Campbell, Dan Kahan, and Saul Perlmutter for helpful conversations. Address correspondence to maccoun@law.stanford.edu.

[1] Routledge Dictionary of Latin Quotations, 2004, p. 107.

[2] http://www.imdb.com/title/tt0087332/quotes

should be, and that (b) the public therefore needs to be persuaded to trust us more. Both points are debatable. Surely, the optimal level of public trust in experts is below 100 percent – perhaps well below 100 percent. Public trust in experts is a two-way street, an exchange relationship requiring something from each side if the potential benefits are to be achieved. Experts have to deserve trust.

There is a large theoretical and empirical literature on trust (see Kramer, 1999; Mayer, Davis, & Schoorman, 1995; McAllister, 1995; Rousseau, Sitkin, Burk, & Camerer, 1998; and the other chapters in this volume) that I will not attempt to review here. I will adopt the definition of trust offered by Rousseau and colleagues (1998):

> Trust is a psychological state comprising the intention to accept vulnerability based on positive expectations of the intentions or behavior of another.[3]

The Merriam-Webster dictionary defines an expert as someone "having, involving, or displaying special skill or knowledge derived from training or experience."[4] But I will mostly focus on the subset of experts who assume the mantle of special authority associated with "credentials" – an advanced degree and/or an affiliation with a university or professional organization. Note that this excludes many kinds of advisors; advisor-client relationships are important and interesting but raise

---

[3] This is quite similar to the definition given by Mayer, Davis, and Schoorman (1995, p. 712): "…the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party." But the latter phrase seems unnecessarily restrictive; the phrase "trust but verify" suggests that we often seek to monitor those we have entrusted with a task (see Williamson, 1993).

[4] http://www.merriam-webster.com/dictionary/expert

many separate complications (contractual agreements, principal-agent problems, etc.). My focus is on public responses to experts in empirical disciplines, including the natural sciences, engineering, and the social sciences, rather than the humanities. This does not imply that these expert opinions are necessarily based directly on empirical research; often they are not.

The classic "Yale School" approach (e.g., Hovland, Janis, & Kelley, 1953) maintained that source credibility is determined jointly by expertise and by trustworthiness, implying that expertise and trustworthiness are distinct constructs; i.e., credibility = f(expertise, trustworthiness). Yet Mayer, Davis, and Schoorman (1995) argue that ability is one of the determinants (along with benevolence and integrity) of trustworthiness, implying that trustworthiness = f(expertise). I think this illustrates a general slipperiness in the literature about whether the focal topic is trust, reliance, confidence, deference, support, or actual influence. It seems likely that these constructs, while distinct, feed into each other in a complex network of relationships. There may be value in systematically sorting out these constructs, but I prefer to sidestep definitional and measurement issues here except where I see obvious value in making distinctions.

French and Raven (1960) further broadened the analysis of influence with their distinctions among various foundations of "social power," including rewards, coercion, legitimacy, identity (affiliations, credentials), and expert knowledge. So there are many available cues for assessing expert credibility. In an important theoretical advance, "dual-process" theories of influence (e.g., Petty & Cacioppo, 1986) posit that such influence cues can operate in two different ways. If we are motivated and able, we can actively scrutinize the quality of a source's arguments. If we are either unmotivated or unable to do so, we can fall back on a more superficial reliance on "peripheral cues." Thus, "Ivy League professor" or "article in *Nature*" are cues that may indeed predict strong arguments from the source, but we also use them heuristically to infer that "she's probably correct" without knowing anything about those arguments. The very nature of the kinds of expertise examined in this

chapter implies a heavy reliance on this kind of heuristic processing; most of us lack the time, training, resources, and ability to fully scrutinize the details of expert statements – which is of course the whole point of having experts.

In the remainder of this chapter, I first provide a brief review of public opinion data on trust in scientists and other experts. These data, perhaps surprisingly, show that experts are largely viewed quite favorably. I then examine two research paradigms that highlight more nuanced aspects of our trust in experts, and argue that they offer converging evidence that, while citizens and experts bring both "inquisitorial" and "adversarial" motives to debates, the desire for truth carries real weight and is not simply given lip service. I close by articulating a normative *epistemic contract* for experts and their consumers, and I review recent developments that suggest ways of facilitating that contract's successful performance.

DO CITIZENS TRUST EXPERTS?

For much of the latter half of the 20th century, many scholars subscribed to a "deficit model" of the public's relationship to science (Allum et al., 2008; Retzbach & Maier, 2014) -- a view (a) that citizens distrust scientists, and (b) that the distrust is based on ignorance, so that better science education would improve citizen trust. Both premises are questionable.

First, overall, Americans have quite favorable opinions of science and of scientists. In a 2009 national survey (Pew Research, 2009), 84 percent of Americans felt that science's effect on society is "mostly positive"; only 6 percent felt it is "mostly negative." The same survey found that 70 percent believed that scientists contribute "a lot" to society's wellbeing – below the level for members of the military (84 percent) but well above the level for lawyers and business executives (23 and 21 percent, respectively). In 2012, 38 percent felt the US government is spending "too little" on scientific research, and only 12 percent said we are spending "too much" (NSF, 2014, Appendix table 7-24).

These views do not translate into an unconditional trust in scientists. Averaged over nearly 40 years, the General Social Survey finds that 40.8 percent of Americans have "a great deal" of confidence in the scientific community, but 46.2 percent have "only some" confidence, and 6.6 percent have "hardly any" confidence (Gauchat, 2012). In 2012, 42 percent of Americans agreed that "we believe too often in science, and not enough in feelings and faith" (NSF, 2014, Appendix table 7-19). And tellingly, surveys are more likely to find evidence of distrust in scientists when they ask about specific "hot-button" topics in science, such as global warming (e.g., Hmielowski, Feldman, Myers, Leiserowitz, & Maibach, 2013), stem cell research (Critchley, 2008), offshore oil drilling (Carlisle, Feezell, Michaud, Smith, & Smith, 2010), or especially expert testimony in adversarial legal proceedings (Cutler & Kovera, 2011). Still, views about science are predominantly favorable across different levels of political ideology and education, and are surprisingly similar among citizens who express doubts about global warming or evolution (Kahan, 2013; NSF, 2014; Pew Research, 2009). While public confidence is not absolute, it is difficult to identify anything that might be characterized as a crisis of confidence in experts.

Second, it is by no means clear that ignorance explains distrust of science, or that knowledge of science necessarily promotes trust. A meta-analysis of 193 different public opinion surveys (Allum et al., 2008) found a reliable positive association between science knowledge and trust in science. While this is consistent with the "deficit" model, it is notable that, controlling for other factors, science knowledge explained less than 1 percent of the variance in trust ratings. In some new areas of science like nanotechnology, ignorance is widespread, and yet people are optimistic about the technology (see Satterfield et al., 2009). Indeed, Kahan and colleagues have shown that the divergence in partisan views about technological risks can actually *increase* after exposure to factual information (Kahan, Peters, Wittlin, Slovic, Larrimore Ouellette, Braman, & Mandel, 2012), and that the divergence is

larger among those with *higher* levels of science literacy and numeracy (Kahan, Peters, Dawson, & Slovic, 2013).

How do psychologists fare relative to other experts? Rotter and Stein (1971) asked two samples of university students, a sample of secretaries, and a sample of public school teachers to rate the truthfulness of 20 different occupations on a 4-point scale (where 1 = "can be counted on to tell the truth as they know it, almost all the time" and 4 = "lies more often than not"). Ratings were very similar across samples, with a mean of 1.59 for psychologists – behind physicians, clergymen, and dentists, tied with judges, and well ahead of labor union officials, politicians, and used car salesmen. Psychologists ranked 6th in competence and 3rd in altruism. Wood, Jones, and Benjamin (1986) found that 91 percent of citizens in four metropolitan areas had highly or somewhat favorable views of psychology. These earlier studies are admittedly somewhat difficult to interpret, because many people associate the term "psychologist" with a clinician offering one-on-one therapeutic services rather than a scientist conducting empirical research. More recently, however, the 2008 APA Benchmark survey (cited in Lilienfeld, 2012) identified widespread citizen doubts about the scientific rigor of psychology, with many seeing it as less rigorous than either medical or economic research. And in a recent cross-disciplinary undergraduate course entitled "Sense, Sensibility, and Science" (which I taught with physicist Saul Perlmutter and philosopher John Campbell), we specifically asked students to "think about your impressions of each scientific discipline presented below." As seen in Figure 1, they reported less trust in psychology and the social sciences than in the physical, environmental, or medical sciences.

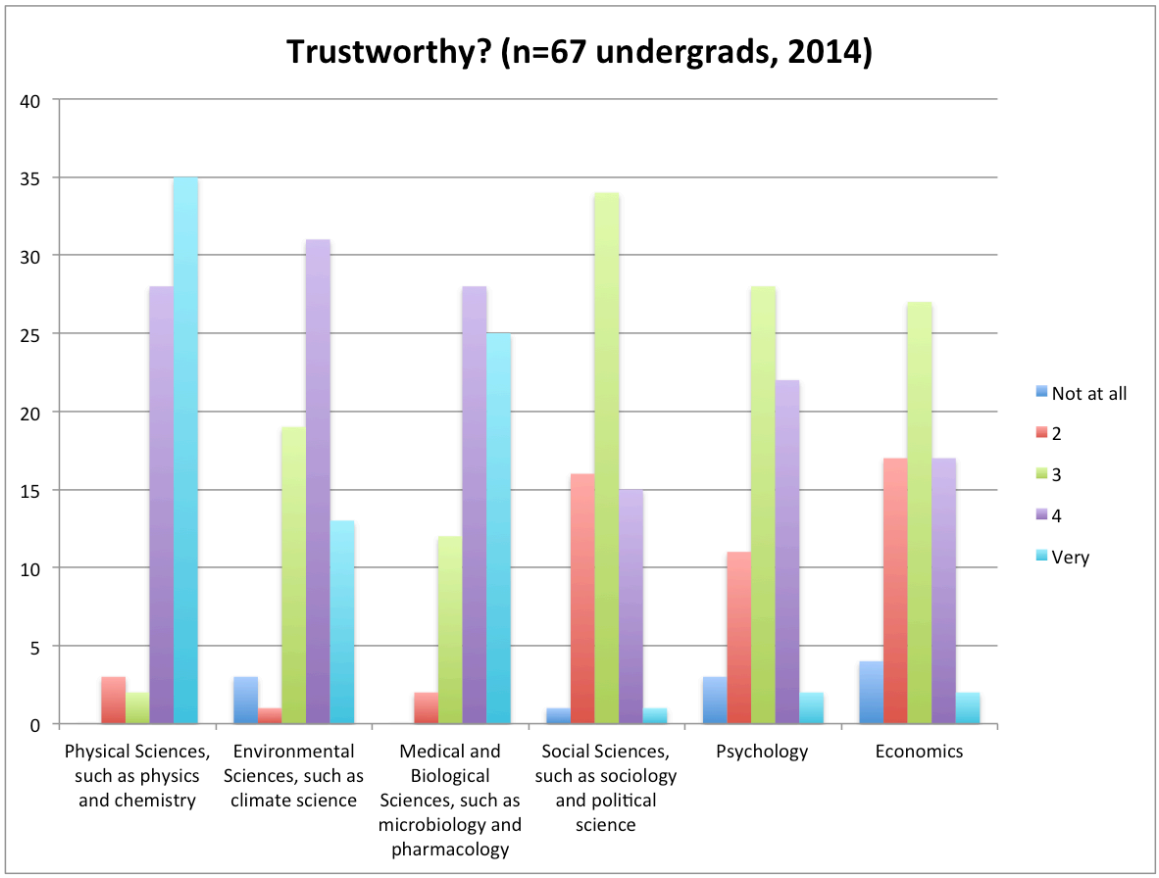

Figure 1. Data from "Sense, Sensibility, and Science" class at UC Berkeley, Spring 2014.

It is not clear whether these new results indicate a decline in psychology's reputation from earlier surveys, since the questions they asked are very different. It is possible that public views have changed in important ways since these older surveys were conducted. Twenge et al. (2014) document stark declines in public trust in others in various surveys between the early 1990s and 2012. Their analyses do not fully explain the trend but show persuasively that it is a period effect (changes in culture by year) rather than an age effect or a birth cohort effect. Thus, any reductions in confidence in experts might be part of a larger decline in trust overall.

But there are also reasons why trust in experts might specifically change over time. In recent years there have been new messages (both favorable and unfavorable)

about social science's credibility. On the one hand, Angrist and Pischke (2010) argue that economics is in the midst of a "credibility revolution" as it evolves from stylized rational choice analyses and econometric analysis of archival correlational data to more behaviorally realistic hypotheses tested using controlled experiments. This "revolution" is spreading into law schools, business schools, and public policy schools. On the other hand, behaviorally realistic hypotheses and controlled experiments have long been mainstays of empirical psychology, and yet it is undergoing a "crisis" of credibility (see Yost, 2012). Psychology's crisis results from a perfect storm of coinciding developments, including (a) some prominent cases of data fabrication (and subsequent article retractions), triggered in part by new statistical methods of forensic re-analysis of published results (see Simonsohn, 2013); (b) publicized failures to replicate various prominent research studies (see Pashler & Wagenmakers, 2012); and (c) emerging evidence that researchers in psychology (and other social and behavioral sciences) frequently engage in questionable practices designed to obtain or promote statistically significant results (Fanelli & Ioannidis, 2013; Ioannidis, 2005; John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011).[5] It is too early to say whether the general public is aware of these academic conversations, and/or whether they have changed public trust in science. It is not even clear what the direction of such changes might be. Learning about economics' "credibility revolution" might enhance trust – but it might also imply that we may have been giving economists way more authority than they deserved all these years. Learning about psychology's "credibility crisis" might impair trust – but it also conveys how a science can undergo constant self-scrutiny and self-improvement.

---

[5] Ioannidis (2008) famously argues that the percentage of reported findings in the literature that are true might be quite low (anywhere from 85 percent down to 0.15 percent in his simulations) depending on various assumptions about typical statistical power, the prior probability of our hypotheses, and the nature and direction of biases in our research methods.

WHAT DO CITIZENS WANT FROM EXPERTS?

It is easy, especially for an academic audience, to assume that "the more citizens trust experts, the better." But this assumption does not hold up to scrutiny. We should want citizens to trust experts…but only when the experts *should* be trusted. So trust should be contingent. On what factors should it be contingent? As noted earlier, there are many potential cues to credibility – an expert's professional training and pedigree, the expert's professional affiliations, whether the research has been published and the prestige of the publication outlet, etc. But these are "peripheral cues" (Petty & Cacioppo, 1986) – proxies for what the consumer really wants to know. What *does* the consumer really want from the expert? I will consider two major motives: An *inquisitorial* desire for truth, and an *adversarial* desire to win a dispute.[6]

These labels evoke two basic forms of legal proceedings – the adversarial approach used in the US and other common-law countries, and the inquisitorial approach used in Continental Europe and elsewhere. But my usage is closer to the more abstract usage in Thibaut and Walker's (1978) theory of procedure. For Thibaut and Walker, the key difference between inquisitorial and adversarial approaches involves *process control*; in adversarial procedures, parties to a dispute select and present their own preferred evidence, whereas in inquisitorial procedures, parties relinquish process control to an ostensibly disinterested third party who seeks out relevant evidence in a neutral fashion.

Thibaut and Walker (1978) argued that inquisitorial procedures are best for *truth conflicts.* Truth conflicts are disputes about the best possible inferences about some true state of affairs. In an idealized truth conflict, the participants may differ in their reading of theory and evidence, but they share a common and overriding epistemic

---

[6] The adversarial motive is non-epistemic, but experts can have non-epistemic motives (e.g., to make money) without caring whether they win. In such cases, it is often their sponsor who wants to win.

motive of reaching the truth of the matter. Thibaut and Walker argued that adversarial procedures are preferable for *conflicts of interest.* In a conflict of interest, the participants have additional, non-epistemic motives (for justice, for retribution, for material gain, for ideological supremacy, for fame) that may distort their interpretation of evidence. It is easy to lionize truth conflicts as the more noble pursuit, but to do so is to adopt the inquisitorial perspective, and to forget that there are many noble pursuits for which truth-seeking is of secondary concern – or even a dangerous distraction from urgent action.

Thibaut and Walker's (1978) analysis is insightful and justly influential, but overly simplistic. Descriptively, topics like gun control, climate control, and birth control involve both factual disputes (truth conflicts) and value tradeoffs (conflicts of interest). And normatively, there is not one but many component decisions to make, some of which might be best handled in an inquisitorial procedure and some of which may be better handled in an adversarial procedure – or by voting, or even by simply agreeing to disagree if no urgent actions are required. MacCoun (1998) lists a number of important disanalogies between litigation in a courtroom and scientific or policy analytic disputes, suggesting that adversarial procedures are more suitable for legal cases than for scientific problem solving. For example, litigation tends to represent both sides of dichotomous questions ("Did he do it?" or "was she negligent?"), but in science the investigation might lead to options not even considered at the outset. Also, litigation requires closure -- a decision needs to be made – whereas science strives to avoid premature commitment to firm conclusions. Finally, in litigation, everyone understands that attorneys are acting as advocated biased toward their clients; in science, an investigator who is being biased is violating social role expectations.

But in any case, I am adapting Thibaut and Walker's (1978) terminology in a way they might not have intended, to refer to the principal motives that consumers and experts bring to a topic. As seen in Table 1, what a citizen wants from experts will be influenced by the citizen's own motives and perception of the experts' motives. If

the citizen is motivated to seek the truth and believes the experts are as well, then expertise should be judged by how believable each expert is – how likely it is that they are correct. If the citizen's motivation is inquisitorial but they perceive the expert(s) to be acting out of adversarial motives (trying to win rather than to be correct), the citizen's task orientation may shift from judging argument validity to assessing the experts' relative honesty. (In the terminology of Eagly, Wood, and Chaiken [1978], this is a shift from looking for a *knowledge bias* to looking for a *reporting bias*.) This is likely to have a corrosive effect on trust in experts, leading to a cynical view about their value. But if the citizen cares more about winning than about finding the truth, the perceived motives of the experts seem less likely to matter: Such a citizen should assess experts by whether they provide useful ammunition for the dispute.

In fact, citizens demonstrably *do* care about experts' motives (e.g., Critchley 2008; Eagly et al., 1978; Walster et al., 1966), and I take this as one of several lines of evidence (see below) that most citizens do care about truth, not just about winning. Indeed, there is an ecological sense in which the adversarial use of experts (to win) is parasitic on the widespread existence of an inquisitorial motive to use experts (to find truth). If no one believed that truth mattered (or, more narrowly, that experts have a higher propensity for finding truth), and everyone knew that, there would be no reason to consult or cite experts. In a world of Machiavellians who know they are all Machiavellians, expertise would have little currency in disputes.

Table 1. Citizens' motives and their perceptions of expert motives.

		Perceived expert motive	
		Inquisitorial	*Adversarial*
Citizen's motives	*Inquisitorial*	Who is more likely to be correct?	Who is more honest?
	Adversarial	Does this help or hurt our side?	

Journalists also play a role in evoking different shared scripts or frames for thinking about expertise. The "inquisitorial frame" is the script of the noble and dedicated scientist, working tirelessly to get to the bottom of some puzzle, emerging with insights that enlighten all and perhaps enhance human welfare. The "adversarial frame" evokes many different scripts – high school debating competitions, Presidential debates, and Sunday morning "dueling expert" news shows – but surely the canonical image is of "hired guns" retained by each side in legal trials (see Cutler & Kovera, 2011). Journalists encourage the adversarial frame because it makes for more compelling stories, but also because of a "balance norm" that journalistic fairness requires giving voice to both sides of a dispute (see Boykoff & Boykoff, 2004). Survey researchers also promote one or the other frame. As noted above, very broad and abstract survey questions about "trust in science" evoke an inquisitorial frame rather than an adversarial frame because they do not even mention any disputed issues; surveys on specific politicized topics are much more likely to uncover skeptical views about experts.

ACCURACY, CONFIDENCE, AND CALIBRATION

If my motive for consulting expert views is inquisitorial, I want to know which experts are most accurate. But even if my motives are adversarial, I may want to find experts that are thought to be accurate – so long as they support my position.

Of course, if we had a way of knowing the true answer to our question, we could readily vet expert accuracy – but at that point, who still needs an expert? So instead, we look for good *proxies* for accuracy – variables that are not accuracy in the instant case, but that seem likely to predict it. One approach is to assess the expert's track record of *past accuracy*. This is often surprisingly difficult, requiring (a) unambiguous predictive or diagnostic statements by the expert, for (b) some meaningful reference class or population of cases, and (c) unequivocal data on the correct outcomes, which either occur later or were not yet revealed to the expert. These requirements are easy to meet for some domains (e.g., meteorology, sports handicapping), partially obtainable in other domains (e.g., radiology, economic

forecasting), and very difficult to obtain in still others (e.g., lawyering, psychiatry, paleontology).

With effort and ingenuity, researchers have studied accuracy rates in many domains. Because it is so hard to define the reference class or population of such predictions, it is impossible to offer a blanket assessment of expert accuracy – there is a risk of oversampling difficult cases where errors are most likely to occur. For example, expert radiologists are not consulted for extremely routine cases but only for the more difficult, high-risk cases, and the top national specialists may work almost exclusively on cases that other experts have been unable to resolve. In any case, the track record for prediction and control in the human sciences -- business and economics (Makridakis, Hogarth, & Gaba, 2009), general medicine (Fink, Lipatov, & Konitzer, 2009), psychiatry and clinical psychology (Dawes, Faust, & Meehl, 1989), law (Goodman-Delahunty et al., 2010), and political forecasting (Tetlock, 2005) – is fairly discouraging, showing that significant errors are far from rare. In all these domains, data-based statistical models consistently outperform intuitive "clinical" expert judgments (Dawes et al., 1989), and these quantitative models are themselves a form of expertise. But even our data-based models often fare badly.

Why? Expert incompetence is surely part of the story (see Gilovich, Griffin, & Kahneman, 2002), as are biases and conflicts of interest that can afflict even the most competent experts (see Moore, Cain, Loewenstein, and Bazerman, 2005). But a big part of the problem is the inherent noisiness and complexity of the open systems of variables that jointly influence human behavior. Choices, behaviors, and their outcomes have multiple determinants, and the relevant parameters change dynamically with context. The variables are endogenously related such that our actions are both a cause of and a response to changing circumstances, and other actors who influence us may be adapting their choices to their expectations of our actions.

Does this mean that relying on experts is a wild gamble? Yes, and no, depending on what one means by "rely." We should not blindly believe everything experts tell us, but we can look to experts to *tell us how much we should believe them* – i.e., their confidence in their opinions. Unfortunately, many lines of evidence suggest that experts tend to be *overconfident*, that is, more confident in their opinions than is warranted by the evidence (Lichtenstein, Fischhoff, & Phillips, 1982; Lin & Bier, 2008).

For example, Tetlock (1998) studied the predictions made by professional foreign policy experts. Importantly, he took great care to ask them to make concrete yes-no predictions about future events – something you are less likely to hear on, say, the *PBS News Hour* – and that enabled him to later assess their accuracy rates once the time period stated in the question had elapsed. His results were discouraging in two different ways. First, the experts did only slightly better than one would have done by tossing a fair coin, suggesting concrete predictions are not something foreign policy experts are good at. But at least as troubling is that the experts who turned out to be incorrect provided confidence ratings that were about as high as those of the experts who turned out to be correct. So knowing whether they know they are good at predictions is something else they were not good at.

In hindsight, it may seem unsurprising that experts are confident when they speak out. The competition for positions and grants and students and publications tends to select for confident people. Scientists are socialized to avoid "going public" with results until they are confident that they are correct. Journalists and policy makers selectively choose and reward experts for being confident rather than wishy-washy. But these factors are not the whole story, and maybe not even the main part of the story, because we see similar overconfidence in lay judgments.

One way to assess overconfidence is to assess *calibration*. Roughly, sources are calibrated to the extent that they are *X* percent accurate for items where they were *X* percent confident. If I have no idea whether the Nile River is shorter or longer than

the Amazon River, I should express a confidence level of 50 percent (i.e., "I'm just guessing"). If you see me as a potential expert and I tell you that I am only 50 percent confident, I am essentially telling you "don't ask me – ask someone else, or just flip a coin." If I turn out to be correct on, say, 65 percent of the occasions where I say I am just guessing about dichotomous items (such as true-false questions), it shows that I am miscalibrated – specifically, I am *underconfident*. At the other extreme, if I say "oh, oh, I know this one, I'm sure of it," I am telling you to trust me. In this case, I am miscalibrated if I am only correct, say, 85 percent of the occasions where I say I am 100 percent confident – I am *overconfident*.[7] According to McKenzie, Liersch, and Yaniv (2008, p. 179), "[i]f people were well calibrated, 90% of their 90% confidence intervals would contain the true value. However, true values typically fall within such intervals between 30% and 60% of the time, indicating extreme overconfidence."

Figure 2 shows the calibration data for 322 UC Berkeley students (graduate and undergraduate) answering general knowledge questions (e.g., "Which is longer? Panama Canal or Suez Canal?"). The qualitative pattern is typical for calibration research using dichotomous yes/no predictions. Students who stated no confidence (.5) were correct more often than chance, suggesting slight underconfidence. But students who were fairly confident (.7 to 1.0) were wrong much more often than they expected – the classic overconfidence finding.

---

[7] A technicality: At the bottom of the confidence scale, one can only be underconfident, and at the top, one can only be overconfident. Also, the set of questions chosen to assess collaboration can be unrepresentative of the domain they assess (e.g., geographic knowledge). These factors can produce overconfidence as a statistical artifact rather than a psychological phenomenon – but from the consumer's standpoint, the source is still overconfident. And overconfidence can be observed in datasets that are not vulnerable to these artifacts (e.g., Brenner, Koehler, Liberman, & Tversky, 1996).

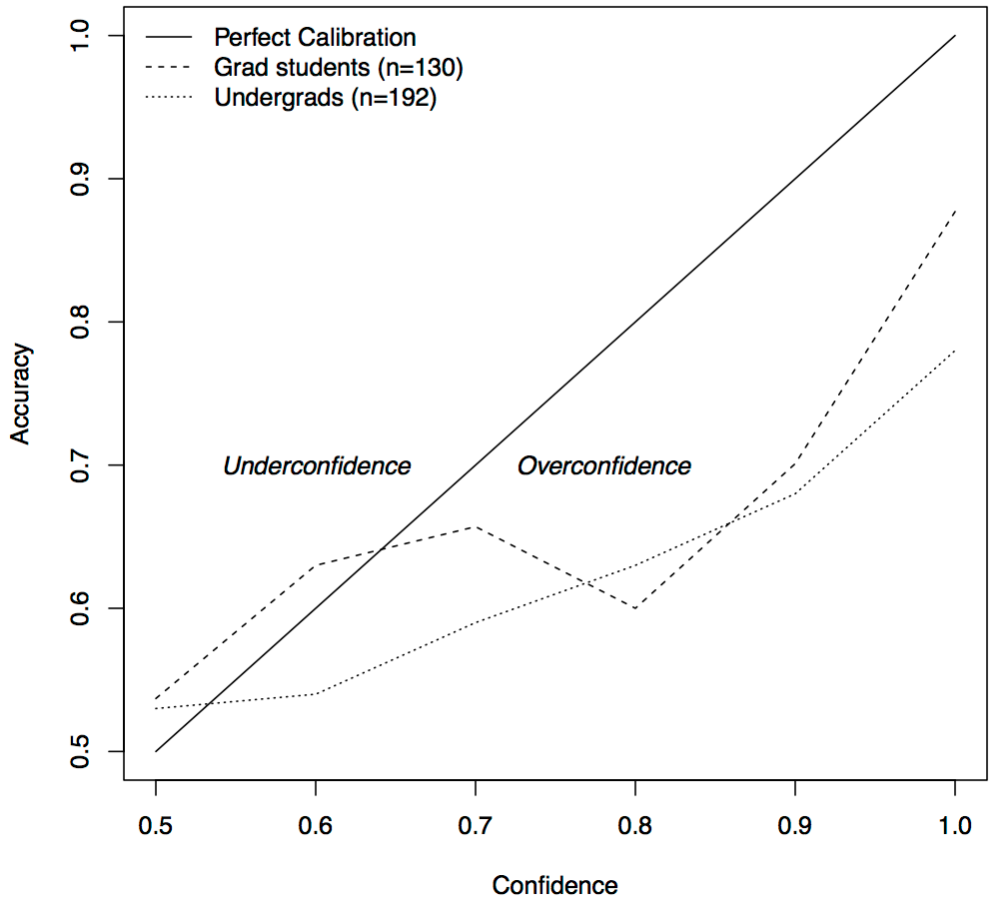

Figure 2. Calibration data for UC Berkeley students answering true/false factual knowledge questions (source: author's unpublished data).

Does it matter that these students were non-experts? Not really; similar results have been found for experts in a variety of professional domains, such as the estimation of physical constants by physicists, or stock-market forecasting by financial consultants (e.g., Braun & Yaniv, 1992; Henrion, & Fischhoff, 1986; Lichtenstein, Fischhoff, & Phillips, 1982; Lin & Bier, 2008). McKenzie, Liersch, and Yaniv (2008) directly compared the calibration of information technology professionals and students for questions about the information industry or about the campus; in either domain, the domain experts (professionals and students, respectively) were more accurate, but similar in overconfidence.

One well-known exception to the general pattern of expert overconfidence is that meteorologists tend to be extremely well-calibrated (Murphy & Winkler, 1977). This says less about their intelligence or their character than about their task. They make many thousands of predictions for a living, and their predictions are quickly, unambiguously, and very publically verifiable by the weather that occurs. Again, the claim is not that they are unusually accurate compared to other experts – indeed, their accuracy falls off very rapidly for events that occur more than a few days in the future. What is notable is that they are calibrated – they know what they know. If a meteorologist says there's a 50 percent chance of precipitation tomorrow, they have not at all told you what will happen – but you can bank on the fact that on half of such days, it will rain.

Other apparent exceptions to overconfidence occur when experts actively strive to "manufacture uncertainty" (Michaels & Monforton, 2005) or act as "merchants of doubt" (Oreskes, & Conway, 2010). This is usually portrayed as a cynical effort to undermine public activism against, say, global warming or tighter tobacco regulations, but it can be an important role for experts, particularly in areas where it is the public or political figures who are overconfident.[8] In any case, this is uncertainty (a prediction near 50 percent) rather than calibration or underconfidence; indeed the experts in question might be quite confident that the outcomes are unpredictable.

A remarkable example of expert calibration involves an expert who is not even human: IBM's Watson computer, who beat the world's top human players on the game show *Jeopardy*. According to Nick Wakeman (20110, Watson "builds an

[8] When my colleagues and I published a study demonstrating why the effects of marijuana legalization on use and revenues were extremely uncertain (Kilmer et al., 2010), we were denounced on various websites for being either useless or cowardly.

evidence profile to determine what are the most likely correct answers. ...For some questions, one answer will have a high confidence level. This is when Watson is most likely to buzz in. For other questions, none of the answers will have a high confidence level and Watson will not buzz in. As IBM Vice President Dave McQueeney told Wakeman: "That's the interesting thing. The machine knows when it doesn't know the answer."

Expert calibration is a trait that is easily overlooked, but one that we should actively cultivate during graduate training. An expert who is highly calibrated is a good judge of his or her knowledge. In essence, an expert who is calibrated and states low confidence is telling us "you don't have much basis for believing me." But citizens and policy makers who prefer a particular course of action have some reason to welcome such news. In essence, the difference between perfect certainty and the expert's stated certainty (100% - subjective confidence) defines *a region of extraevidentiary discretion*, where the decision maker has freedom to base his or her decisions on non-evidentiary considerations.[9] "We don't know whether assault weapon bans actually save lives? Then fine, my supporters can keep their Uzis."

CONFIDENCE AND CALIBRATION AS PERSUASIVE CUES

Expert overconfidence might not be a problem if people simply anticipated it and discounted it accordingly. Unfortunately, they do not. For example, eyewitness confidence is known to be a poor guide to eyewitness accuracy, yet it is one of the cues jurors rely on most heavily in deciding whether to believe an eyewitness identification (see Bradfield & Wells, 2000). Price and Stone (2004) reviewed evidence for this reliance on source confidence, labeling it the "confidence heuristic." A heuristic is a cognitive shortcut; sometimes a rule of thumb but often a proxy variable that provides a rough substitute for something more difficult to

---

[9] This also implies that an overconfident expert is unfairly restricting the decision maker's zone of discretion.

observe. In this case, Price and Stone suggested that confidence serves as a rough proxy for accuracy.

That people rely on a confidence heuristic is troubling, because confidence is an unreliable proxy for accuracy, but even more so because it is a *biased* proxy for accuracy – people are not just randomly miscalibrated, they are usually (except at the upper end of the probability scale) *over*confident. It is also distasteful to think that people can be overconfident and not only "get away with it" but even get rewarded for it. And in fact, there are good reasons to question whether this really works. A basic principle in both evolutionary biology and economic game theory is that signals that are easily faked have little value (Zahavi & Zahavi, 1997). Cosmides and Tooby (1992) even argue that evolution has provided us with a hard-wired cognitive module for "cheater detection."And indeed, in a series of experiments, my colleagues and I (Tenney, MacCoun, Spellman & Hastie, 2007; Tenney, Spellman, & MacCoun, 2008) demonstrated that overconfidence can backfire.

In two experiments, Tenney et al. (2007) varied eyewitness confidence (high vs. low) as well as whether the eyewitness was shown to have made an error during testimony. In both studies, learning about a single error hurt the credibility and impact of the high-confidence witness, but not the low-confidence witness. (Indeed, the error actually enhanced the credibility of the low-confidence witness.) In two additional experiments, Tenney et al. (2008) replicated and extended these results, showing that (a) they were not simply due to jurors preferring cautiousness or modesty in a witness, and (b) the confident witness is not discredited when the error is one no reasonable person could have anticipated.

Tenney, Spellman and I (2008) argued that people use an implicit "presumption of calibration" principle: Our default assumption is that people who say they are confident are calibrated, so we are willing to rely on confidence as a cue or proxy for accuracy. But this assumption is fragile, and readily dropped when the source is found to be in error, even about a peripheral detail. I depict this logic in Figure 2.

The presumption of calibration hypothesis is reminiscent of linguistic theories of conversational pragmatics. For example, Grice (1989, p. 27) offers two "maxims of quality" that he believes to be implicitly assumed in our conversations: "1. Do not say what you believe to be false. 2. Do not say that for which you lack adequate evidence."

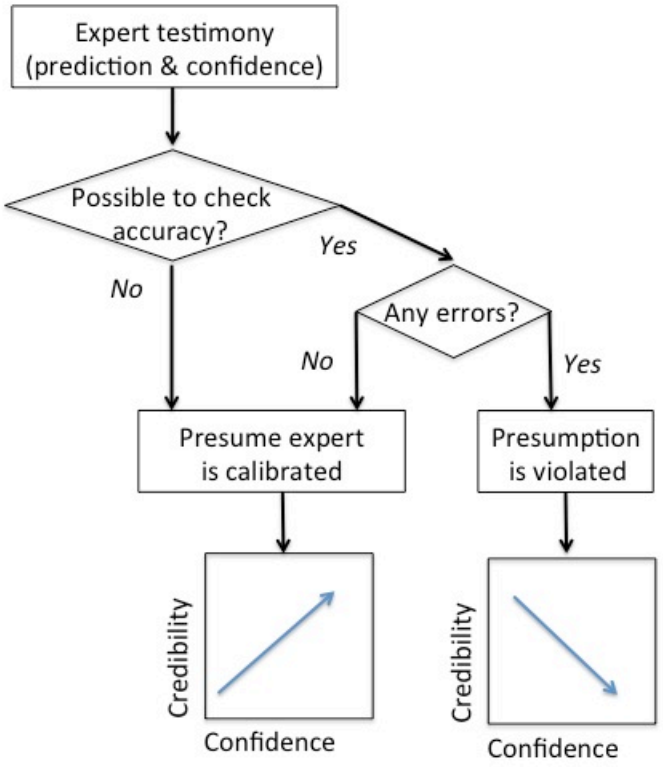

Figure 3. The presumption of calibration.

It is gratifying to know that listeners, at least in some contexts, hold speakers accountable for unwarranted confidence. Unfortunately, Sah, Moore, and I (2013) recently established an important boundary condition on the discrediting effect of overconfidence. Importantly, our results are consistent with the presumption-of-calibration logic shown in Figure 3.

In a first experiment, we conceptually replicated the basic Tenney et al. results in an advisor paradigm using physical judgments of weight and confidence intervals

rather than categorical confidence statements. But a key step in the "presumption of calibration" logic is that consumers need to discover the source's accuracy in order to judge the source's calibration. Thus, in a second experiment, we varied the availability of feedback on the advisor's performance. In one condition, feedback was freely available, and the results replicated the Tenney et al. pattern; that is, demonstrable overconfidence hurts credibility. But in a second condition, there was no feedback available, and in this case, we replicated the Price and Stone (2004) "confidence heuristic" pattern: Confidence was rewarded, irrespective of whether the experts were giving increasingly better or worse advice over time. In a third condition, feedback was available, but participants had to purchase it by investing some experimentally allocated credits toward a prize. Despite the fact that the feedback price was carefully designed to be a good investment, most respondents opted out of buying the feedback. As a result, most resorted to the confidence heuristic.

Regrettably, all too many advising situations are analogous to our "costly feedback" (or even "no feedback") condition. Barriers to obtaining feedback on expert performance may include the need to wait for outcomes to occur, the ambiguity of interpreting outcomes, conflicting data, conflicting interpretations of data, proprietary or classified data, and the ambiguity of parsing what the expert really predicted. As noted earlier, for some professions, this is relatively easy; for others it is very difficult. I return to this point and consider some possible solutions at the end of the chapter.

The evidence for the presumption-of-calibration hypothesis provides potential reason to believe that citizens are motivated, at least in part, by an inquisitorial search for truth. If they were not, why should a source's error be discrediting, and why should we care about calibration? Still, this is not a conclusive argument; it could be that the source is discredited simply because we think the source's tainted testimony makes him or her less effective in advancing any adversarial goals we might have.

BIASED ASSIMILATION, NAÏVE REALISM, AND ATTITUDE ATTRIBUTION

There are two potentially important limitations of most of the calibration-confidence studies (and indeed, many of the studies in the "advisor paradigm"). First, most (but not all; see Price & Stone, 2004) of these studies present participants with the advice of a single expert, whereas we routinely encounter situations where multiple experts disagree. Of course, the mere existence of disagreement does not necessarily imply the adversarial frame. Consider the recent scientific debate about the causes of "colony collapse disorder" – the rapid disappearance of large numbers of honeybees. Experts disagreed about potential explanations – mites, bacteria, viruses, fungi, antibiotics, electromagnetic radiation – but I doubt this disagreement hurt the credibility of the scientists or the scientific process, because the disagreements were voiced in good faith in the context of a shared pursuit of the truth. It is not disagreement that evokes the adversarial frame; it is the possibility of conflicts of interest, ideology, and ego.

And that highlights a second limitation of many of these studies. Most (but not all; see Tenney et al., 2007, 2008) of these studies use tasks that are mostly truth conflicts rather than conflicts of interest. How do confidence and calibration work when there are two opposing experts in a more adversarial situation? It is one thing to believe an expert with high confidence. But if there are *two* experts, and *both* are confident, and they *disagree*, surely something is amiss.[10]

---

[10] Given the complexity and stochastic nature of many causal systems, I think it is probably theoretically possible for two experts, each well-calibrated in the past, to be fairly confident in opposing predictions, but only under rare circumstances. In the three-dimensional space of confidence, calibration, and disagreement, that corner mostly will be empty.

But there is another research literature that routinely studies conflicting expert sources (albeit with high confidence either implicitly or explicitly held constant): the *biased assimilation* literature.[11]

Unlike many of the general opinion surveys, biased assimilation studies tend to explicitly juxtapose two or more experts, which all but guarantees that at least one of the two experts is wrong – or at least more wrong than the other expert. Following the basic design of the classic study in the paradigm (Lord, Lepper, & Ross, 1979), people are exposed to research evidence that either supports or contradicts their own personal beliefs about an issue. Across a wide range of topics, the consistent finding is that people find expert findings more credible and plausible when the research is congenial with the consumer's own beliefs, even when researchers hold the methodology constant. Note that the term "biased assimilation" does not mean this phenomenon is necessarily irrational or unjustified, though in some cases it clearly is. (For reviews of this literature and discussion of alternative normative interpretations, see MacCoun, 1998; MacCoun & Paletz, 2011). Similar results are found in studies by Kahan and his colleagues (e.g., Kahan, Braman, Slovic, Gastil, & Cohen, 2009; Kahan, Jenkins-Smith, & Braman, 2011; Kahan, Peters, Wittlin, Slovic, Larrimore Oullette, Braman, & Mandel, 2012). They show that consumers selectively interpret research evidence on technological risks in a manner that protects their personal commitments to cultural values like egalitarianism or individualism.

The most cynical view of these biased assimilation results is that people simply choose to believe whatever they want to believe. This is probably not the case; Kunda (1990) reviews evidence that most people's beliefs are at least partially constrained by the available evidence, even when it is uncongenial. A less cynical view is that people are "naïve realists" – they believe that there is a clear reality out

[11] Another relevant literature looks at expert testimony at trial (Cutler & Kovera, 2011).

there that directly determines what we perceive (Pronin, Gilovich, & Ross, 2004). As such, sources that seem to agree with one's views must simply be perceiving what is real. Thus, experts who disagree with one's view pose a puzzle, but the naïve realist can readily solve it by attributing the discordant expert's testimony to some form of bias – e.g., the expert's political ideology or an economic conflict of interest. Revealingly, even in the adversarial setting, experts seem to display naïve realism – they often fail to recognize the very real potential for bias that comes from being retained by one side of a dispute (Commons, Miller, & Gutheil, 2004; Murrie, Boccaccini, Guarnera, & Rufino, 2013).

A recent study by MacCoun and Paletz (2011) provides some evidence supporting this naïve realism account. We presented approximately 1000 California adults with new research findings regarding public policies that are more popular with liberals (gun control, medical marijuana) or with conservatives (capital punishment, school vouchers) – except that we varied the direction of the finding so that each policy was either found to be effective or had no effect. As in previous studies, there was a reliable biased assimilation effect such that people were more inclined to believe results congenial with their own political views. We then asked citizens to speculate about the political ideology of the social scientists who conducted the target studies. Consistent with naïve realism, when the findings were congenial, most people chose not to speculate about the researcher's politics. But when the findings were uncongenial, citizens were increasingly likely to speculate that the investigator was liberal (for a liberal finding) or conservative (for a conservative finding) – what social psychologists call an *attitude attribution* effect (Jones & Harris, 1967), but a selective one. Figure 4 outlines the process.

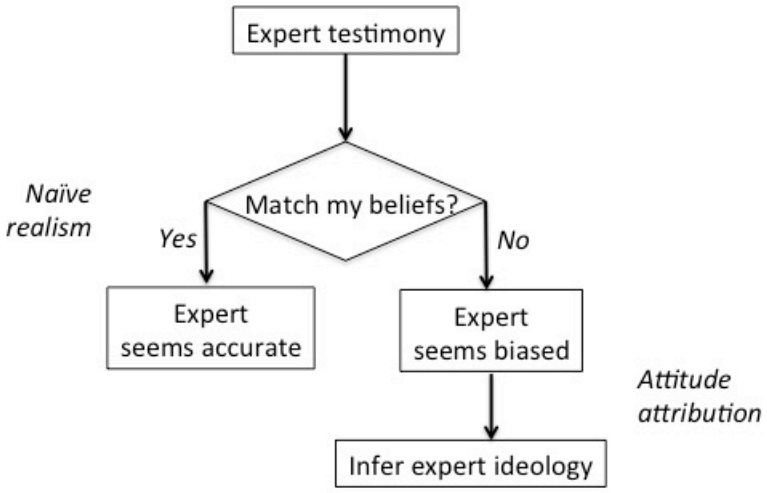

Figure 4. Naïve realism and attitude attribution.

Overall, this tendency to "explain away" uncongenial results by invoking ideology was somewhat stronger for conservative respondents. One interpretation is that conservatives are more skeptical of science in general, or less willing to be constrained by evidence (see Mooney, 2006, 2012). There is some evidence for this interpretation; for example, using the General Social Survey, Gauchat (2012) shows that in the 1970s, self-identified conservatives reported the highest trust in science, but by 2010 they expressed less trust than either liberals or moderates. But Kahan and colleagues (Kahan, Braman, Slovic, Gastil, & Cohen, 2012; Kahan, Jenkins-Smith, & Braman, 2011) show that the divergence in partisan views about technological risks can actually increase after exposure to factual information, or with increasing science literacy and numeracy. Indeed, in the case of climate change, views were *more,* rather than less, polarized among the most technically knowledgeable

respondents. Also, the fact of the matter is that academic researchers *are* disproportionately likely to hold political views to the left of center (e.g., Pew Research, 2009), and of course conservatives do not hold a monopoly on the selective use or rejection of scientific research (Berezow & Campbell, 2012).

The calibration and confidence paradigm and the biased assimilation paradigm examine different dimensions of the broader topic of trust in experts, but I think there is some continuity in their results. I would not argue that the presumption-of-calibration process (Figure 3) and naïve realism in the biased assimilation paradigm (Figure 4) are descriptions of the same phenomenon, but I think it is reasonable to conjecture that they are closely linked. Naïve realism enables us to presume that others are calibrated (until proven otherwise). In both paradigms, people seem willing to give expert sources the benefit of the doubt. I take this as further support for the proposition that the inquisitorial motive is more basic than the adversarial motive, at least in the domains examined in the research reviewed here.

A NORMATIVE PERSPECTIVE: THE EPISTEMIC CONTRACT

I am not certain that my interpretation is correct as an empirical matter – no overconfidence here! But I do want to offer a normative, aspirational account of the appropriate trust relationship between experts and their consumers. My account is "epistemic," meaning that it pertains to the validity of some claim of knowledge. And it is "contractual," in the sense that it outlines an agreement setting out expectations regarding the obligations of each agreeing party. Of course, this is a metaphorical contract, a rhetorical device in the tradition of Rousseau or Locke. I think these expectations are implicit and hence poorly articulated when experts claim expertise and when citizens consult experts. To make the expectations more explicit and more concrete, I offer what I will call "the Epistemic Contract":

1. *The Expert:*
    a. *If the expert wants to claim the mantle of authority for a topic, and be granted special consideration (above and beyond an ordinary citizen) in a debate,*

   b. *then the expert should strive to be calibrated, clearly delineating the strengths and limitations of his or her knowledge.*

2. *The Consumer:*
   a. *If the consumer sincerely wants to make informed decisions, and to claim expert support for his or her views,*
   b. *Then the consumer's opinions should be constrained by, and susceptible to revision in light of, available expert opinions,*
   c. *but only to the extent that the expert's sincere confidence dictates.*

The epistemic contract is an aspirational model, but I do not think it is a naïve one. It allows both the expert and the consumer to hold a wide range of additional motives. The expert may also want material or social reward or political influence. The consumer may want to use the expert's testimony to gain material or social reward or political influence. These motives do not violate the epistemic contract, so long as the expert strives to be an honest and calibrated broker of information, and the consumer agrees to either be constrained by that information – or to forgo the use of expert support in a debate. By "constrained," I have in mind belief revision that is at least qualitatively (directionally) consistent with what Bayesian updating might dictate.

Nevertheless, I recognize that the epistemic contract is not easily enforceable. As Sah, Moore, and I (2013) have suggested, for experts and consumers to be bound by evidentiary considerations, they each need an accurate assessment of *calibration,* which in turn requires an assessment of how the expert's confidence tracks his or her accuracy or validity. That's a tall order.

FACILITATING AN APPROPRIATE LEVEL OF TRUST IN EXPERTS

A healthy trust relationship between experts and their consumers requires an effort from both sides. Experts have to earn trust, and consumers need to learn that experts can be trusted. As I have argued, a powerful way for both sides to eachive this is through calibration data linking confidence to accuracy. How can we improve

expert calibration and consumer access to calibration data? There are a variety of different approaches to this problem, some old, some new, and some still purely hypothetical:

Organized skepticism (Merton, 1938). Scientists use peer review and replication to scrutinize new claims. These methods do not always work, but they are gradually improving (see Bornmann & Mungra, 2011). And it is all too easy to forget that the perfect is the enemy of the good: Even flawed systems of peer review and replication are preferable to none at all, and there are many domains of expertise that are not presently subjected to such scrutiny, and should be. For example, legal scholarship is increasingly empirical, and often draws on lines of argument from probability theory, economics, or other conceptual frameworks. Yet much of this work is published in law reviews without any formal peer review process. And financial consultants provide extremely consequential advice to their clients with little or no oversight from their firms or from regulators.

Open science. In response to the replicability crisis in psychology, there are active efforts to increase transparency by encouraging (or even requiring) researchers to register and/or openly post their hypotheses, their pilot data, their instruments and methods, their data, and their analyses (see Miguel et al., 2014). These efforts now include collaborative multi-institution attempts to audit and replicate published research findings (e.g., Alogna et al., 2014).

Blind analysis. Open science makes it easier to detect bias in research. But it is possible to *prevent* many biases by blinding the investigator in ways that enforce objectivity. Double-blind studies have long been used to blind investigators in the collection of data, but physicists have developed methods of perturbing data (with noise or bias) so that investigators do not know which hypothesis their results favor until the analysis is already complete, and these methods can be adapted for use in psychology and the social sciences (see MacCoun & Perlmutter, under review).

Aggregation. Meta-analysis was an important advance in aggregating evidence across experts, and through moderator analysis, it is enabling us to see which features of what expert sources and methods influence results (e.g., Braver, Thoemmes, & Rosenthal, 2014). More recently, Nate Silver's *FiveThirtyEight* blog demonstrated that aggregating across pre-election polls produces a forecast that is more accurate than its component parts (Silver, 2012). Bayesian model-averaging methods are another approach in the same spirit.

Prediction markets. Prediction markets allow participants to buy and sell shares in outcomes – to "put some skin in the game." For example, at the Iowa Electronic Markets, run by the Tipper College of Business at the University of Iowa, participants were able to buy and sell futures contracts on the success of the Republican Party at retaking the Senate in November 2014. The market gave the Republicans a 70 percent chance of winning, which is what indeed happened.[12] Prediction markets shift the focus from individual experts to the collective expertise of a community of opinionated people. Which may seem crazy, except that it seems to work. Though initially controversial, the evidence to date suggests that prediction markets perform at least as well as opinion polls, and often better, though the key ingredients of the recipe (monetary stakes, information pooling, sample selection biases) are still under investigation (see Arrow et al., 2008). A variation on prediction markets is the kind of public wagers that prominent scientists sometimes make with each other (see Giles, 2002).

Forecasting tournaments. As noted earlier, Tetlock's (2005) work on foreign policy experts suggested that they were neither accurate nor well calibrated. But since then, he and his colleagues have demonstrated that properly designed forecasting tournaments can not only identify accurate and calibrated predictors, but can also promote constant improvements in accuracy and calibration (Mellers et al., 2014;

---

[12] The IEM can be found at https://tippie.uiowa.edu/iem/ ; the 2014 Senate trading is summarized at https://tippie.uiowa.edu/iem/media/story.cfm?ID=3389

Tetlock et al., 2014). For several years, Tetlock and colleagues have solicited probabilistic forecasts on world events from hundreds of professional and amateur forecasters. These new tournaments have produced much more successful, and better calibrated, forecasting than Tetlock found in his earlier work. These tournaments have a recipe that includes requiring *testable predictions, explicit articulation of uncertainty, opportunities for revision, clear metrics for accuracy and calibration (Brier scores), training, effective group process,* and *publically observable performance data.* Intriguingly, their approach does not involve an accredited guild with restrictive membership, and some of their most accurate "superforecasters" are lay citizens without specialized training, credentials, or access to classified or proprietary data.

Reputation markets. Amazon, Yelp, Facebook, Twitter, Airbnb, and other contemporary web-based services rely heavily on organized systems for collecting and disseminating reputational data, and they have already worked out many of the glitches involved in setting up reliable and fair feedback. Citation counts serve as one traditional reputational metric for experts, and new variations are being developed to take into account the expert's career length, discipline, research topic, and so on (Cronin & Sugimoto, 2014).

CONCLUSION

Citizens often trust experts, but their trust is contingent. Unfortunately, it is more likely to be contingent on *fidelity* (correspondence to citizens' preferences) rather than on *validity* (correspondence to empirical truth). Still, citizens are far from impervious to validity, and they appear to want experts to be accurate. Citizens appear to assume experts are unbiased unless their testimony is unexpected (naïve realism), and they seem to assume that an expert's confidence is warranted unless the evidence shows otherwise (the presumption of calibration).

To foster an appropriate level of trust in experts, we need better systems for promoting and assessing expert accuracy and calibration. Fortunately, the kinds of

innovative methods described above suggest that it may become easier for consumers of expert opinion to decide what and whom to believe, and when. By highlighting the flaws and foibles of expert judgment, these approaches may not produce a net increase in trust in experts, but they will promote an *appropriate* level of trust in experts. And they will encourage and enable experts to earn that trust.

REFERENCES

Allum, N., Sturgis, P., Tabourazi, D. & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: a meta-analysis. *Public Understanding of Science, 17,* 35–54.

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, S., Birch, S., Birt, A. R., ... Zwaan, R. A. (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science, 9,* 556–578.

Angrist, J. D., & Pischke, J. S. (2010). The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*, 3–30.

Arrow, K. J., et al. (2008). The promise of prediction markets. *Science, 320,* 877-878.

Berezow, A. B., & Campbell, H. (2012). *Science left behind: Feel-good fallacies and the rise of the anti-scientific left.* New York: PublicAffairs.

Bornmann, L., & Mungra, P. (2011). Improving peer review in scholarly journals. *European Science Editing, 37,* 41-43.

Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: global warming and the US prestige press. *Global Environmental Change, 14,* 125–136.

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five *Biggers* criteria. *Law and Human Behavior, 24,* 581–594.

Braun, P. A., & Yaniv, I. (1992). A case study of expert judgment: Economists' probabilities versus base-rate model forecasts. *Journal of Behavioral Decision Making, 5,* 217-231.

Braver, S. L., Thoemmes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science, 9,* 333-342.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes, 65*, 212-219.

Carlisle, J. E., Feezell, J. T., Michaud, K. E., Smith, E. R., & Smith, L. (2010). The public's trust in scientific claims regarding offshore oil drilling. *Public Understanding of Science, 19,* 514-527.

Commons, M. L., Miller, P. M. & Gutheil, T. G. (2004). Expert witness perceptions of bias in experts. *Journal of the American Academy of Psychiatry and the Law, 32,* 70-75.

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Critchley, C. R. (2008). Public opinion and trust in scientists: The role of the research context, and the perceived motivation of stem cell researchers. *Public Understanding of Science, 17,* 309-327.

Cronin, E. B., & Sugimoto, C. R. (2014). *Beyond bibliometrics: multidimensional indicators of scholarly impact.* MIT Press.

Cutler, B. L., & Kovera, M. B. (2011). Expert psychological testimony. *Current Directions in Psychological Science, 20,* 53-57.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668-1674.

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology, 36,* 424-435.

Fanelli, D., & Ioannidis, J. P. A. (2013). US studies may overestimate effect sizes in softer research. *Proceedings of the National Academy of Science, 110,* 15031-15036.

Fink, W., Lipatov, V., & Konitzer, M. (2009). Diagnoses by general practitioners: Accuracy and reliability. *International Journal of Forecasting, 25,* 784-793.

French, J. P. R. Jr., and Raven, B. (1960). The bases of social power. In D. Cartwright and A. Zander (eds.), Group dynamics (pp. 607-623). New York: Harper and Row.

Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review, 77,* 167-187.

Giles, J. (2002). Scientific wagers: Wanna bet? *Nature, 420,* 354-355.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge University Press.

Goodman-Delahunty, J., Granhag, P. A., Hartwig, M., & Loftus, E. F. (2010). Insightful or wishful: Lawyers' ability to predict case outcomes. *Psychology, Public Policy, and Law, 16,* 133-157.

Grice, P. (1989). *Studies in the ways of words.* Harvard University Press.

Hmielowski, J. D., Feldman, L., Myers, T. A., Leiserowitz, A., & Maibach, E. (2013). An attack on science? Media use, trust in scientists, and perceptions of global warming. *Public Understanding of Science, XX, xxx-xxx.* [Print version not out yet.]

Henrion, M., & Fischhoff, B. (1986). Assessing uncertainty in physical constants. *American Journal of Physics, 54,* 791-798.

Hovland, C. I., Janis, I. L., & Kelley, J. J. (1953). *Communication and persuasion.* New Haven: Yale University Press.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2,* 696-701.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524-532.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Psychology, 3*, 1–24.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making, 8,* 407-424.

Kahan, D. M., Braman, D., Slovic, P., Gastil, J., & Cohen, G. (2009). Cultural cognition of the risks and benefits of nanotechnology. *Nature Nanotechnology, 4,* 87-90.

Kahan, D. M., Jenkins-Smith, H., Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research, 14,* 147-174.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2013, September 3). Motivated numeracy and enlightened self-government. Yale Law School, Public Law Working Paper No. 307. Available at http://ssrn.com/abstract=2319992

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Larrimore Oullette, L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change. *Nature Climate Change, 2,* 732-735.

Kilmer, B., Caulkins, J. P., Pacula, R. L., MacCoun, R. J., & Reuter, P. H. (2010). *Altered state? Assessing how marijuana legalization in California could influence marijuana consumption and public budgets.* Santa Monica, RAND.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*, 216-247.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50,* 69-98.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480-498.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *The Academy of Management Review, 23,* 438-458.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: heuristics and biases. Cambridge University Press.

Lillienfeld, S. O. (2012). Public skepticism of psychology: Why many people perceive the study of human behavior as unscientific. *American Psychologist, 67,* 111-129.

Lin, S. W., & Bier, V. M. (2008). A study of expert overconfidence. *Reliability Engineering & System Safety, 93,* 711-721.

Lord, C. G., Ross, L., Lepper, M. R. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098-2109.

MacCoun, R. (1998). Biases in the interpretation and use of research results, *Annual Review of Psychology, 49*, 259-287.

MacCoun, R. J., & Paletz, S. (2009). Citizens' perceptions of ideological bias in research on public policy controversies. *Political Psychology, 30,* 43-65.

MacCoun, R. J., & Perlmutter, S. (under review). Blind analysis as a correction for confirmatory bias in physics and in psychology. Prepared for Lilienfeld, S. O., & Waldman, I. (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions*. John Wiley and Sons.

Makridakis, S., Hogarth, R. M., & Gaba, A. (2009). Forecasting and uncertainty in the economic arena. *International Journal of Forecasting, 25,* 794-812.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20,* 709-734.

McKenzie, C. R. M., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes, 107,* 179-191.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S., Moore, D., Atanasov, P., Swift, S., Murray, T, Stone, E., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science, 25,* 1106-1115.

Merton, R. K. (1938). Science and the social order. *Philosophy of Science, 5,* 321-337.

Michaels D., & Monforton, C. (2005). Manufacturing uncertainty: Contested science and the protection of the public's health and environment. *American Journal of Public Health, 95,* S39-S48.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., Laitin, D., Madon, T., Nelson, L., Nosek, B. A., Peterson, M., Sedlmayr, R., Simmons, J. P., Simonsohn, U., & Van der Laan, M. (2014). Promoting transparency in social science research. *Science, 343,* 30-31.

Moore, D. A., Cain, D. M., Loewenstein, G. and Bazerman, M. (eds.), *Conflicts of interest: Problems and solutions from law, medicine and organizational settings*. London: Cambridge University Press.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press.

Mooney, C. (2006). *The Republican war on science.* New York: Basic Books.

Mooney, C. (2012). *The Republican brain: The science of why they deny science—and reality.* New York: Wiley.

Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Journal of the Royal Statistical Society: Applied Statistics, 26,* 41-47.

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science, XXX,* xxx-xxx. [Print version not out yet.]

NSF (2014). *Science and engineering indicators 2014.* Washington, D.C.: National Science Foundation.

Pasher, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7,* 528-530.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Pew Research (2009, July 9). *Public praises science; scientists fault public, media.* Pew Research Center for the People and the Press.

Price, P. C., & Stone, E. R. (2004). I ntuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17,* 39–57.

Retzbach, A., & Maier, M. (2014). Communicating scientific uncertainty: Media effects on public engagement with science. *Communication Research, XX, xxx-xxx.* [print version not out yet]

Rotter, J. B., & Stein, D. K. (1971). Public attitudes toward the trustworthiness, competence, and altruism of twenty selected occupations. *Journal of Applied Social Psychology, 1,* 334-343.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23,* 393-404.

Sah, S., Moore, D., & MacCoun, R. (2013). Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organizational Behavior and Human Decision Processes, 121*, 246-255.

Satterfield, T., Kandlikar, M., Beaudrie, C. E. H., Conti, J., & Harthorn, B. H. (2009). Anticipating the perceived risk of nanotechnologies. *Nature Nanotechnology, 4,* 752-758.

Silver, N. (2012). *The signal and the noise*. New York: Penguin.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359-1366.

Simonsohn, U. (2013). Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science, 24,* 1875-1888.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18,* 46-50.

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): Fact-finders rely on others who are well calibrated. *Journal of Experimental Social Psychology, 44,* 1368-1375.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Tetlock, P. E., Mellers, B. A., Rohrbaugh, N., & Chen, E. (2014). Forecasting tournaments: Tools for increasing transparency and improving the quality of debate. *Current Directions in Psychological Science, 23,* 290-295.

Thibaut, J., & Walker, L. (1978). A theory of procedure. *California Law Review, 66,* 541-566.

Twenge, J. M., Campbell, W. K., & Carter, N. T. (2014). Declines in trust in others and confidence in institutions among American adults and late adolescents, 1972–2012**.** *Psychological Science,* 25, 1914-1923.

Wakeman, N. (2011, February 11). IBM's 'Jeopardy!' match more than game playing: Big Blue sees Watson driving strides in analytics. *Washington Technology.*

Walster, E., Aronson, E., & Abrahams, D. (1966). On increasing the persuasiveness of a low prestige communicator. *Journal of Experimental Social Psychology, 2,* 325-342.

Williamson, O. E. (1993). Calculativeness, trust, and economic organization. *Journal of Law and Economics, 36,* 453-486.

Wood, W., Jones, M., & Benjamin, L. T. (1986). Surveying psychology's public image. *American Psychologist, 41,* 947-953.

Yong, E. (2012). Bad copy. *Nature, 485,* 298-300.

Zahavi, A and Zahavi, A (1997). *The Handicap Principle*. Oxford University Press.