Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Quality of Expertise

Edward D. Van Wesep

MANAGEMENT SCIENCE
Vol. 62, No. 10, October 2016, pp. 2937–2951
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

informs®
http://dx.doi.org/10.1287/mnsc.2015.2271

The Quality of Expertise

Edward D. Van Wesep
Leeds School of Business, University of Colorado Boulder, Boulder, Colorado 80309, edward.vanwesep@colorado.edu

Policy makers and managers often turn to experts when in need of information, but we should expect experts to be systematically biased. This is because the decision to research a question implies a belief that research will be fruitful. If priors about the impact of one's work are correct on average, then those who choose to research a question are optimistic about the quality of their work. The bias varies predictably with attributes of the question being studied. This fact has implications for a variety of mechanism design applications and yields predictions in accordance with a large literature in psychology.

Keywords: expertise; learning; prediction; overconfidence
History: Received March 31, 2014; accepted May 17, 2015, by Teck-Hua Ho, behavioral economics. Published online in *Articles in Advance* January 19, 2016.

1. Introduction

Policy makers and managers often need information to make a decision, and they turn to "experts" for that information. Experts have a deep knowledge of the theory and evidence concerning their subject of expertise, and they are trained with the skills to evaluate that evidence. A long theoretical literature in the field of economics, beginning with Crawford and Sobel (1982), asks the question of how much information policy makers are able to elicit from experts, given that experts may not agree with the objectives of the policy maker and may therefore attempt to convey false information. In this paper, I tackle a question one step primitive: Are experts even experts and, if so, under what circumstances? Stated another way, even if an expert agrees with the objectives of the policy maker, in what cases can the policy maker trust that the expert's information is correct? I show that experts are often not the best source of information, *even if they are the most informed individuals in the economy and genuinely wish to provide accurate information*. This fact is consistent with extensive empirical work (e.g., Tetlock 2006). I provide parameters that are, in principle at least, observable to a policy maker that help her determine what weight to put on expert opinion.

Given that there are likely several semantic issues that have already arisen in the reader's mind, I should define terms up front. For the bulk of this paper, I focus on what is, for practical purposes, academic research expertise. An *expert* is somebody who actively researches *questions* such as "does smoking cause lung cancer?," "does deficit-financed fiscal stimulus reduce the duration of recessions?," or "does a transaction tax reduce liquidity in financial markets?" These questions fall into *fields* such as medicine, economics, or finance.

I will compare the quality of expert opinion to the quality of nonexpert opinion. A *nonexpert* is somebody who is familiar with some past research on a topic but is not up-to-date. This may be another researcher in the field who studies different questions and has not seen the relevant literature since graduate school. It could be the expert herself, prior to becoming a researcher. It could also be a "dilettante" (in the words of Tetlock 2006), defined to be an individual with no particular expertise and a random, incomplete knowledge of past research. In each case, I show that the quality of expert opinion can be lower than the quality of nonexpert opinion.

The model works as follows: in each period, researchers enter fields (e.g., start their Ph.D.'s) and decide which questions to research (e.g., decide on a dissertation topic). In making their decisions, they observe past research on each question, the existence and quality of which is common knowledge, and form priors of the likely usefulness of future work on each question. These priors are, in expectation, correct.[1] Each researcher ultimately chooses a question that he or she considers important and for which he or she believes this work will be productive. After choosing a question, the researcher studies the question and updates his or her beliefs, taking into account past and current work. Once this cohort of researchers has formed opinions, a policy maker or manager interested in the answer to a question may solicit opinions from experts as well as from nonexperts.

At each step in the analysis, I make assumptions that are as charitable as possible toward experts.

[1] By this I mean that priors are drawn from a distribution whose expectation is the true value of the parameter.

I assume that (i) only experts can observe current research on a topic, (ii) they want to convey the most accurate information possible to a querying policy maker, and (iii) their utilities depend only on their perceptions of their ultimate contributions to knowledge. This third assumption, though appearing innocuous, will generate a selection bias: the very fact that they have chosen to study a particular question implies that they may fundamentally disagree with others—who have chosen different questions—about the quality of research on that question. The key assumption driving the results is that experts do not understand the impact of this selection problem on their posteriors. When forming an opinion about the true answer to the question they study, they combine their prior for the quality of their work with information of its quality gleaned from doing the work, but they do not account for the selection problem generated by heterogeneous priors. In truth, although their priors are unbiased ex ante, they are not unbiased *conditional on having chosen a particular question*.[2]

The source of bias in this paper is identical to the source of "rational overoptimism" described in Van den Steen (2004), and this paper should be seen as an application of his more general model. I develop the implications of this bias for expertise and find several results that are consistent with empirical research. The first set concerns the decision to research a question. For example, I show that researchers are more likely to study a question if the question is newer, more important, and less well understood. This is because the marginal impact of research is higher when the quantity of past research is lower and when the question is more important. A researcher is more likely to choose a question if the data and/or techniques available for studying the question have recently improved. This is because the quality of new research, relative to old, is higher. She also is more likely to study a question if she believes that work will be fruitful. For example, a young economics Ph.D. candidate that sees behavioral theories as plausible is more likely to study the question of how people time-discount than a candidate who sees those theories as implausible. These results should conform closely to the reader's intuition about how young researchers choose their research topics, providing some comfort that the model accurately captures the process of becoming an expert.

[2] This ignorance of a selection problem is decidedly second order. It would perhaps be reasonable to allow for more significant bias among researchers—the psychological evidence appears to confirm that many experts suffer from first-order biases, such as overconfidence and narcissism. I show, however, that first-order biases are unnecessary to cast doubt on expertise. This does not mean that I dismiss first-order biases as untrue; it means only that results can be generated with or without them.

The second set of results concerns how properties of a question relate to the likely accuracy of expert and nonexpert beliefs. Experts put too much weight on their work, on average, but this is a more severe problem for some questions than for others. Specifically, questions that are less important, are more heavily studied historically, and do not feature novel research technologies are associated with less reliable experts.

Because the amount of published research is always weakly increasing, and because the content and quality of that research is common knowledge after its publication and dissemination, nonexpert knowledge about a question is always increasing over time. However, as this knowledge increases, the marginal benefit of additional research on that question decreases. The people who still choose to study it are therefore increasingly biased over time. Expert opinion is better than nonexpert opinion when a question is new: there is little published work for nonexperts to observe and little selection into becoming an expert. The rate of improvement in expert opinion decreases over time, as the amount of new research relative to past research decreases and as selection into the question becomes more severe. At some point, the level of selection into a question is so significant that expert opinion is less accurate than nonexpert opinion. In fact, an expert's opinion will become less accurate as she performs more research.

The fact that nonexperts can have more accurate views than experts may seem absurd, but it is consistent with the data. Tetlock (2006) compiles considerable data from two decades of surveys of experts, laymen, and dilettantes (smart people with solid educations but no specific expertise) in which they were asked to forecast for a wide variety of variables in a variety of fields. For example, respondents were asked questions such as "what is GDP [gross domestic product] growth likely to be next year" or "what is the likelihood that the United States will send ground troops to Iraq?" Tetlock finds that it is typical for experts to be no better—and often worse—at forecasting these outcomes than dilettantes and that experts were often worse at forecasting within their fields than out. They were certainly worse than some simple rules that they should have known to use—for example, "for numeric questions, always predict no change." These findings are consistent with the implications of the model, where experts will generally be more accurate than laymen but sometimes less accurate than dilettantes.

The third set of results concerns how the properties of the field relate to the likely accuracy of expert and nonexpert beliefs. Growing fields offer many opportunities for novel and important research, implying less accurate expert opinions for older questions in these fields. Static fields offer few interesting alternative questions, so researchers studying an old question need not be particularly biased.

The fourth set of results concerns how the properties of the experts themselves relate to the quality of their opinions. When significant new research technologies become available, work on a question is likely to be more fruitful, attracting less biased researchers. For example, the great recession beginning in 2007 provided a wealth of new data regarding the business cycle and financial crises. Researchers choosing to study questions in these areas post-2007 are likely to be less biased than those choosing to study these questions prior to 2007. Researchers choosing questions in these periods form what I call "golden cohorts," in that they have more accurate opinions than older researchers when they are young and more accurate opinions than younger researchers when they are old. *For their entire careers*, they are the "most expert." One could argue that these cohorts have appeared, in practice, surrounding innovations in research opportunities (e.g., in economics, the discovery/invention of game theory, the marginal revolution, or the capital asset pricing model).[3]

I also argue that extrinsically motivated experts' beliefs can be more accurate than those of intrinsically motivated experts. Motivation matters because it determines how a question is chosen. Intrinsically motivated experts choose questions based on where they believe they are likely to have an impact, but a belief of impact can be due to either actual impact or a bias. Extrinsically motivated experts prefer questions where they expect to have impact, but they also prefer questions where past work was more precise, because high-quality past work improves current prediction. These preferences imply less selection into questions and therefore less bias.

The fifth set of results concerns the relationship between experts' self-assessed and actual accuracies. Because past work is common knowledge, differences in experts' self-assessments arise from differences in priors regarding the quality of their work. Researchers who are pessimistic, in the sense that they believe their work is worse than it actually is, will report lower confidence in their beliefs than researchers with correct priors. Because they underweight their work in forming a posterior, they also have less precise posteriors: in the *pessimistic domain*, self-assessed and actual accuracy positively correlate. Researchers that are optimistic will report higher confidence in their beliefs but will overweight their work and therefore have less accurate posteriors than researchers with correct priors: in the *optimistic domain*, self-assessed and actual accuracy negatively correlate. If there were no selection into expertise, the empirical correlation between self-assessed and actual accuracy would be approximately zero. However, because pessimistic researchers will choose alternative areas of expertise, *the empirically observed correlation within the set of experts will be negative*. This relationship has been observed in dozens of studies, and this model provides an endogenous basis for it.

[3] Naturally, there are other reasons that golden cohorts may arise. Notably, upon the introduction of a new research technology, young researchers have a greater incentive to invest the time required to master the technology and will therefore be more adept at employing it than older researchers. Assuming that the technology has diminishing returns, later researchers will turn to newer technologies, leaving the earlier cohorts superior at all stages of their careers. These explanations are not mutually exclusive.

2. Review of the Literature

This model relates to four literatures in economics and psychology: (i) the causes and consequences of overconfidence, (ii) how best to elicit expertise, (iii) the effect of heterogeneous priors on organizations, and (iv) the quality of real-world expertise.

Extensive work in psychology and behavioral economics has established that overconfidence appears to be common. Although some theorists have responded to these studies by rejecting their designs (e.g., Benoit and Dubra 2011), most have tried to understand whence the overconfidence arises. For example, Compte and Postlewaite (2004) argue that because confidence improves performance, some overconfidence is optimal. Zabojnik (2004) finds that overconfidence can arise as a result of the process of acquiring information about one's own abilities. Santos-Pinto and Sobel (2005) find that if people possess attributes on several dimensions and have heterogeneous value functions over those attributes, they will acquire skills on the dimensions that they value most and thence perceive themselves as superior to the average. For example, in self-assessing driving skill, some people believe that driving slowly is safe. They therefore drive slowly and pat themselves on the back for being better-than-average drivers. Much like this paper, heterogeneous priors plus selection drives an important empirical observation. Most closely related to the mechanism in this paper, Van den Steen (2004) shows that overconfidence is inherent to activities in which there is a participation choice. I focus in this paper on applying his result to a specific participation choice—the choice to become an expert—and evaluating the impact of this bias on research opinion.

Research on the quality of elicited expert information began with Crawford and Sobel (1982) and has become quite extensive (see, e.g., Krishna and Morgan 2001, 2004; Aumann and Hart 2003). It is typically assumed that the expert providing the information has different preferences from the decision maker receiving it, providing an incentive for the sender to muddle the information and making the inference problem difficult. However, a small but growing number of papers allows the expert and decision maker to share identical preferences. Most notably, Che and Kartik (2009) allow shared preferences but heterogeneous priors, and they find that the decision maker may prefer to hire an expert that does not fully share her priors. The reason is that an expert who shares the decision maker's priors has little incentive to acquire information and provide accurate advice. A "biased" expert (relative to the decision maker) can only be convincing by providing quality information, making the acquisition of that information more valuable. Van Wesep (2015) considers the setting of customer reviews, in which somebody with knowledge of a product provides that information to somebody considering acquiring the product. He finds that reviews will be, on average, polite when it is most important to distinguish the bad from the worst and will be rude when it is most important to distinguish the good from the best. In this paper, I also assume that the one providing the information genuinely wishes to be accurate, but ask when we should expect her to succeed. This is one step primitive to the standard question of how to elicit and interpret her information.

A key assumption of this paper is that different potential experts have differing priors for the quality of their work, and that this affects the quality of expert opinion. This is an example of a much broader literature on the effect of heterogeneous priors on organizational form.[4] Van den Steen (2010a, b) shows that several common aspects of corporations, such as interpersonal control, low-powered managerial incentives, and corporate culture, all arise naturally from difficulties generated by heterogeneous priors. Sethi and Yildiz (2012) ask why different social groups hold such divergent beliefs, even when that divergence is known. They find that if priors are heterogeneous and unobservable, communication cannot aggregate all disperse private information into public beliefs. If priors are observable, communication is sufficient to aggregate information. I find a similar result here. If the expert's prior is observable, then the manager or policy maker soliciting her opinion can infer her information from her posterior. If her prior is unobservable, then inference is not possible, and it may be best to solicit alternative opinions.

Finally, there is a growing empirical literature in psychology concerning the specific question of this paper: Are experts even experts, and, if so, when? Repeated studies have found examples in which expert opinion can be less accurate than nonexpert opinion and can even be less accurate than simple models of which the expert should be aware (see Camerer and Johnson 1991 for an early review).[5] There is significant work showing that information often reduces the precision of predictions.[6] Information also serves to make those making predictions more confident of their predictions, implying a negative relationship between confidence and accuracy.

Once it is clear that many experts are not expert when it comes to prediction and forecasting, a natural question is what makes some better than others. Tetlock (2006), for example, separates experts that form opinions using a single, dominant worldview from those that look at each situation on a case-by-case basis and finds that the latter are both more accurate and less confident. This literature is quite large and complementary to the model that I present in this paper. I show that even if people/experts are not inherently overconfident, they will tend to be overconfident in their area of expertise as a result of the endogenous sorting of experts into areas of expertise. Further, the more confident they are, the less accurate they will be. Rather than explore how properties of the individual correlate with accuracy, as much of the work in psychology has done, I explore how properties of questions and fields should correlate with accuracy. To my knowledge, this question is novel, and the method of analysis I use is more theoretical. Further, I seek endogenous, rather than exogenous, explanations for the failure of expert predictions.

3. The One-Period Model

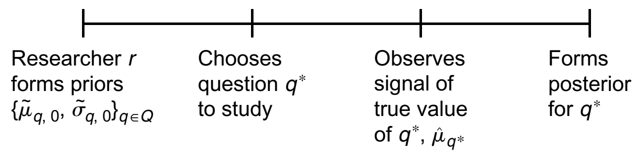
I begin with a one-period model that captures the major forces involved and follow with an infinite-period model in the next section. There is one researcher, denoted r, who must choose a question $q \in Q$ to study, where Q is a finite set of available questions. Each question has an importance V_q, which is society's agreed-upon value of answering the question, and it is defined by the number μ_q, which represents "the truth" about q. For example, if the question is "what is the elasticity of labor demand?," then the truth might be $\mu_q = -0.05$. The truth is unknown to the researcher, who has a prior for each value of μ_q, denoted by $\tilde{\mu}_q$, which is distributed normally with mean μ_q and variance $\sigma^2_{q,0} > 0$. The researcher also has a prior as to how precise her research will be for each question, $\tilde{\sigma}_{q,1}$, defined in greater detail below.

[4] See Morris (1995) for a discussion of the philosophical underpinnings of the heterogeneous prior assumption.

[5] For example, Dawes et al. (1989) show that doctors are less accurate in predicting patient outcomes than a simple statistical model. (Dawes and Corrigan 1974, p. 97) write that "the statistical analysis was thought to provide a floor to which the judgment of the experienced clinician could be compared. The floor turned out to be a ceiling."

[6] For example, Goldstein and Gigerenzer (2002) show that Americans are better at guessing the larger of two German cities than guessing the larger of two American cities, about which they presumably know more.

Figure 1 Timeline of the One-Period Model



Note. In the one-period model, the researcher is born and forms priors for the truth of each question and the precision of her future research. Using this information, she chooses a question to study, observes a signal of the truth for that question, and then forms a posterior.

Importantly, she does not *know* at the outset of her career how good her research will be on any question, but she does have a prior.

The researcher begins by observing the available questions and forming priors regarding (i) the truth and (ii) the likely quality of her own research on each question. Next, she chooses a question to study, denoted by q^*. She then observes a signal of the truth for q^*, denoted by $\hat{\mu}_{q^*}$, which is a normal random variable with mean μ_{q^*} and *unknown* variance $\sigma^2_{q^*,1}$. Finally, she forms a posterior for q^*, which is a weighted average of her prior $\tilde{\mu}_{q^*}$ and her signal $\hat{\mu}_{q^*}$. This sequence of events is pictured in Figure 1. Throughout the paper, a tilde will always represent a prior, and a hat will represent a signal.

Her prior for the quality of her research about q is a prior for the variance of the signal of the truth μ_q that she will observe if she chooses to study q. That is, $\tilde{\sigma}^2_{q,1}$ is a prior belief of the value of $\sigma^2_{q,1}$, which is unknown. This prior is distributed according to $\tilde{\sigma}^2_{q,1} \sim G_q(\cdot)$, where G_q is increasing and differentiable over its entire support $[0, \bar{G}]$, with $G_q(0) = 0$ and $G_q(\bar{G}) = 1$. Note that G_q is unknown to the researcher, but priors are unbiased: $E_{G_q}(\tilde{\sigma}^2_{q,1}) = \sigma^2_{q,1}$.

In forming a posterior, the researcher assigns Bayesian weights to her prior and her signal, under the assumption that her signal has variance $\tilde{\sigma}^2_{q,1}$. Because $\tilde{\sigma}^2_{q,1}$ is an unbiased estimate of $\sigma^2_{q,1}$, and she has no other information upon which to base a weight, this weighting seems reasonable. Indeed, as I show below, there is no alternative weighting rule that is superior.[7]

The remaining assumptions describe the researcher's preferences and, therefore, how she decides what to study. I assume that the researcher prefers to study questions that are more important and those upon which she believes that she can have greater impact. Two variables define a researcher's expected impact. First, her expected impact studying q is increasing in the variance of her prior about q: if little is already known, her marginal impact will be greater. For example, if we already know the relationship between cigarettes and lung cancer, then there is little incremental benefit that we could gain by studying it. Second, her expected impact studying q is decreasing in her expectation of the variance of her own research. For example, if the researcher believes that she is especially skilled at acquiring proprietary data, then she expects her work on questions requiring proprietary data to be more fruitful. These assumptions are equivalent to her choosing a question q to maximize her utility $u(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1})$, subject to the following conditions:

A1. Utility is increasing in the importance of the question: $u_1(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1}) > 0$.

A2. Utility is increasing in the variance of her prior of μ_q: $u_2(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1}) > 0$.

A3. Utility is decreasing in the variance of her prior of the quality of her own research: $u_3(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1}) < 0$.

These assumptions are designed to accurately reflect what drives most people to research a question and capture a variety of reasonable choices of utility function. For example, a researcher may care specifically about how much her work reduces the variance in society's estimate of q. That is, she may care specifically about the difference in variance between her prior and posterior for μ_q and choose the question that offers the largest potential variance reduction. Then her utility function would be written as $u_q = [\sigma^2_{q,0} - 1/(1/\tilde{\sigma}^2_{q,1} + 1/\sigma^2_{q,0})] \times V_q = [\sigma^2_{q,0}/(1 + \tilde{\sigma}^2_{q,1}/\sigma^2_{q,0})]V_q$. This clearly satisfies assumptions A1–A3.

3.1. The Precision of Posteriors vs. Priors

The researcher's posterior for q^* is assumed to be the weighted average of her prior for μ_{q^*}, and her signal of it. Because she is Bayesian, the weights should be the relative precisions of her prior and her signal. Because she does not know the precision of her signal, she uses her prior for that precision:

$$\mu_{q^*\text{posterior}} = \frac{1/\sigma^2_{q^*,0}}{1/\sigma^2_{q^*,0} + 1/\tilde{\sigma}^2_{q^*,1}}\tilde{\mu}_{q^*} + \frac{1/\tilde{\sigma}^2_{q^*}}{1/\sigma^2_{q^*,0} + 1/\tilde{\sigma}^2_{q^*,1}}\hat{\mu}_{q^*}.$$

If her prior for the precision of her signal were correct, then the precision of her posterior would simply be the sum of the precisions of her prior and her signal. However, her prior for the precision of her signal does not equal the true precision, so the variance of her posterior is given by

$$\begin{aligned} &\mathrm{Var}\left(\frac{1/\sigma^2_{q^*,0}}{1/\sigma^2_{q^*,0} + 1/\tilde{\sigma}^2_{q^*,1}}\tilde{\mu}_{q^*} + \frac{1/\tilde{\sigma}^2_{q^*}}{1/\sigma^2_{q^*,0} + 1/\tilde{\sigma}^2_{q^*,1}}\hat{\mu}_{q^*}\right) \\ &\quad = \frac{1/\sigma^2_{q^*,0} + \sigma^2_{q^*,1}/(\tilde{\sigma}^2_{q^*,1})^2}{(1/\sigma^2_{q^*,0} + 1/\tilde{\sigma}^2_{q^*,1})^2}. \end{aligned}$$

[7] The model could be adjusted so that she gets a signal of $\sigma^2_{q,1}$ after she observes $\hat{\mu}_{q^*}$. So long as the eventual weight of $\hat{\mu}_{q^*}$ in the posterior depends at least slightly on her prior $\tilde{\sigma}^2_{q,1}$, the qualitative results would stand.

Most of the results in this paper concern what makes expert posteriors more or less accurate and therefore involve comparative statics involving the expectation of the variance calculated above. It is a useful check to note that accurate beliefs of the quality of one's work provide for the most accurate posteriors.

Lemma 1. *A researcher's belief is most accurate when her prior of the quality of her signal is correct, i.e., when* $\tilde{\sigma}^2_{q^*,1} = \sigma^2_{q^*,1}$.

The researcher will choose to research the question that yields the highest utility. That is, she chooses q^* so that $q^* = \arg\max_q u(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1})$. Suppose that we wish to know what priors will lead to a researcher studying question q^*. Higher values of V_{q^*} and $\sigma^2_{q^*,0}$ and lower values of $\tilde{\sigma}^2_{q^*,1}$ will make q^* more attractive and therefore more likely to be chosen. Similarly, lower values of V_q and $\sigma^2_{q,0}$, and higher values of $\tilde{\sigma}^2_{q,1}$ for all of the other potential questions $q \in Q \backslash q^*$ will make it more likely that q^* is chosen.

Lemma 2. *A researcher is weakly more likely to study question* q^* *if*

(1) q^* *is more important relative to other questions*: V_{q^*} *is higher or any element of* $\{V_q\}_{q \in Q\backslash q^*}$ *is lower.*

(2) *The prior for* q^* *is less precise relative to other questions*: $\sigma^2_{q^*,0}$ *is higher or any element of* $\{\sigma^2_{q,0}\}_{q \in Q\backslash q^*}$ *is lower.*

(3) *The prior for the quality of her own work on* q^* *is higher relative to other questions*: $\tilde{\sigma}^2_{q^*,1}$ *is lower or any element of* $\{\tilde{\sigma}^2_{q,1}\}_{q \in Q\backslash q^*}$ *is higher.*

These three results form sources of selection bias to which expert beliefs will be subject.

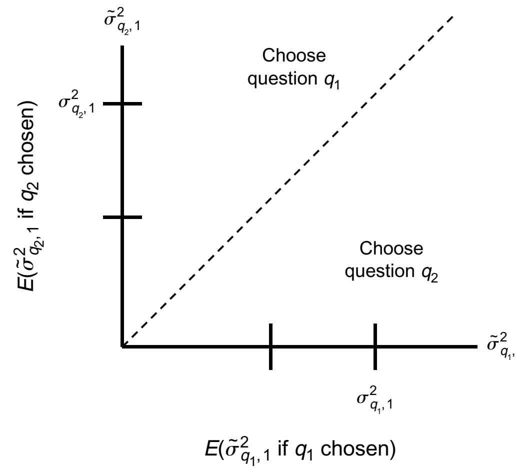
In the domain where an expert believes that her signal is more accurate than it actually is, as her belief of the quality of her work increases, the precision of her posterior decreases. This should be intuitive: the more precise she believes her signal to be, the more weight she will put on it. Because she is already putting too much weight on her signal, this increased weight makes her less accurate.

Corollary 1. *If a researcher believes that her signal is more precise than it actually is, then the precision of her posterior is decreasing in that belief.*

Because of the selection into a chosen question, most experts will believe that their signals are more precise than they actually are and will therefore have suboptimally precise posteriors.

For any set of parameters and priors about questions other than q^*, $\{V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1}\}_{q \in Q\backslash q^*}$, and any value and prior precision for q^*, $\{V_{q^*}, \sigma^2_{q^*,0}\}$, there is a threshold value $\bar{\sigma}^2_{q^*}$ such that a researcher selects question q^* over the other potential questions if $\tilde{\sigma}^2_{q^*,1} < \bar{\sigma}^2_{q^*,1}$ and

Figure 2 **Graphical Representation of Researcher Bias**



Notes. The set of questions is $Q_t = q_1, q_2$. Each question has the same value and same prior, so a researcher will choose to study the one for which she expects her research to be more precise—i.e., the one for which her prior for the variance of the signal is lower. Because priors are correct, in expectation, the expected value of the prior, conditional on a question being chosen, is less than the unconditional expectation, which is accurate.

chooses an alternative question if $\tilde{\sigma}^2_{q^*,1} > \bar{\sigma}^2_{q^*,1}$.[8] This means that *the expected prior for the variance of research on* q^*, *conditional on* q^* *being chosen, is less than the unconditional expectation of that variance*. Because priors are correct, on average, priors will be biased conditional on choosing q^*, and a researcher will consequently put too much weight on her work.

This result can be seen graphically in Figure 2. I plot an example in which there are two questions, q_1 and q_2. They have identical values and equally accurate priors for the truth: $V_{q_1} = V_{q_2}$ and $\sigma^2_{q_1,0} = \sigma^2_{q_2,0}$. Therefore, a researcher will choose question q_1 over q_2 if $\tilde{\sigma}^2_{q_1,1} < \tilde{\sigma}^2_{q_2,1}$, and vice versa. The researcher is more likely to choose question q_1 if she underestimates the variance of her signal for q_1. This implies that the expected value of her prior of the variance of her signal of q_1, conditional on q_1 being chosen, is less than the true variance.

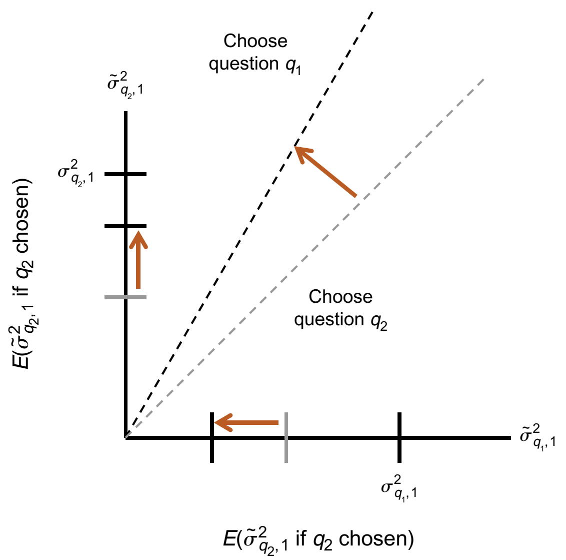
We therefore arrive at the result that selection into studying a question causes a researcher to be biased in favor of her own work, even though her priors are unbiased ex ante.

Lemma 3. *A researcher is, on average, biased in favor of her own research. The weight that she puts on her work is higher, on average, than the optimal weight.*

The fact that researchers are biased in favor of their own work arises endogenously, and naturally, from the fact that they do not actually know the quality of their work but choose topics to research where they

[8] The case of equality occurs with probability zero and may be ignored.

Figure 3 (Color online) Graphical Representation of Researcher Bias: Variation



Notes. As in Figure 2, the set of questions is $Q_t = q_1, q_2$. Initially, each question has the same value and same prior, so a researcher will choose to study the one for which she expects her research to be more precise—i.e., the one for which her prior for the variance of the signal is lower. If the value of question q_2 goes up relative to q_1, or the prior knowledge about q_1 goes up relative to q_2, then the researcher is willing to choose question q_2 even with a prior that her signal regarding q_2 will not be as accurate. This causes the conditional expectation of the accuracy of her signal to approach the unconditional expectation if she chooses question q_2. In the limit, as the value of question q_2 increases, she will always choose it and the unconditional and conditional expectations will equalize. For question q_1, however, the conditional and unconditional expectations further diverge. This means that experts are more accurate for more important questions and for questions where less is known.

expect to have impact. Bias is not assumed, but it arises nonetheless.

These lemmas lead to the primary results of the paper.

Proposition 1. *If parameters are such that question q^* is sometimes chosen, then the expected precision of an expert's posterior for q^* is higher, ceterus paribus, if*

(1) *the prior for question q^* is weak: $\sigma^2_{q^*,0}$ is lower; and*
(2) *the question is more important: V_{q^*} is higher.*

In Figure 3, I consider the same assumptions on parameters as in Figure 2 and show what happens if the value of question q_2 increases relative to q_1 or the precision of the prior for μ_{q_2} decreases relative to the precision of the prior for μ_{q_1}. In either case, the researcher is more willing to choose question q_2. The situations in which q_1 is chosen and $\tilde{\sigma}^2_{q_1,1} < \sigma^2_{q_1,1}$ become even more rare, relative to the situations in which $\tilde{\sigma}^2_{q_1,1} > \sigma^2_{q_1,1}$, so the expected bias of a researcher choosing question q_1 increases. In essence, the loss of researchers from question q_1 was not drawn evenly from the entire group. Instead, the lost researchers were marginal, and marginal researchers are less biased.

It turns out that the selection problem can be so severe that experts, despite having greater knowledge than nonexperts, actually have less precise posteriors. They may even have posteriors less precise than their priors.

Corollary 2. *An expert's posterior can be less precise, on average, than her prior.*

Experts can have less accurate posteriors than priors: by studying question q^*, they actually end up with less accurate beliefs than before they began. This happens when experts put significant weight on their signals even though the signals are not particularly informative. When will they do this? The stronger the prior of the truth about question q^*, the more biased the researcher must be to choose q^*. Although the researcher gets to observe a signal of the truth in forming a posterior, her weight on that signal may be so unreasonable that it would be better not to have had the signal in the first place.

An extreme example might be helpful. Suppose that the prior is so strong and the question so unimportant that a researcher will only choose to study the question if she believes that her work will nearly answer the question. That is, $\bar{\sigma}^2_{q^*,1} \to 0$. Then, if she does study the question, she will put nearly full weight on the signal instead of the prior. If the true variance of the signal exceeds the variance of the prior, then her posterior will be inferior to her prior. This example is extreme, but it highlights the way that the signal can become so overweighted that it would be better not to have it.

What sorts of questions are associated with a precise prior of μ_{q^*}? These are questions about which much is known. They are likely to be older questions that have been heavily researched in the past, with high-quality data, opportunities for unconfounded trials, etc.—for example, questions such as "what is the speed of light?" or "what is the structure of a DNA molecule?"

Before moving on to the infinite-period case, I must discuss an alternative way that a policy maker or manager could elicit information from the expert that would be superior to simply asking about her posterior. The principal could simply ask the researcher for her signal $\hat{\mu}_{q^*}$, her prior $\tilde{\mu}_{q^*}$, and the quality of her prior, $\sigma^2_{q^*,0}$, and apply her own weighting to the new and old information. The expert would be happy to comply: after all, she really does want to honestly convey the truth. There are four reasons to believe that the model has relevance in the face of this possibility.

First, in practice, people really do ask experts their opinions and do not ask them about the individual pieces of information and their precisions. Therefore, the model is a positive statement about how most interactions with experts currently work.

Second, this provides a potential normative argument for how policy makers and managers should frame questions for experts. They should dig deeper into how the posterior was constructed and ask other experts and nonexperts to then reconstruct a posterior with their own weights.

Third, one reason that we may not ask for this deconstruction is that it is not really possible. In the model, I summarize a paper as being a draw of a random variable with a variance that becomes known. This is a nice modelling technique but does not really match reality (except for meta-analyses of randomized clinical trials, which can be aggregated in the way the model suggests). How would you frame a question to somebody about the fall of the Roman Empire and ask the person to describe the weights put on various original and nonoriginal sources, arguments, papers, books, etc.? It is not clear that it is possible.

Fourth, it remains unclear what weight to put upon the new and old information. If the policy maker or manager is aware of the selection problem, then she could put less weight on new information than the researcher would choose, but there is no way to know how much that weight should be reduced. The distribution from which the prior for $\tilde{\sigma}_{q^*}$ was drawn is unknown to everyone, and there is no reason to believe that the principal would be better at figuring out that distribution than the researcher herself.

4. The Infinite-Period Model

In this section, I make a number of changes that make the model richer and more easily applied to real-world settings. First, I move from a one-period setup to an infinite-period setup. This will allow me to model the evolution of knowledge and researcher opinion about a question over time. Second, I allow for multiple researchers in each time period. For simplicity, I do not allow them to see each other's research contemporaneously, though changing this assumption would not affect the qualitative results. Third, I allow the research technology to evolve over time, which provides interesting results about how recent shocks to that technology affects the precision of researcher opinion.

Throughout this section, a tilde over a variable will continue to represent a prior for the variable. Because there are multiple periods, it is important to note that the prior is for researchers as of time t. It will not be the same prior that researchers in a different period form. Similarly, a hat will continue to refer to a signal that a researcher receives, and a bar will refer to a threshold value. Posteriors will be subscripted with the term "posterior." The timeline within a given period will be identical to before. The only changes will be (i) multiple periods, (ii) multiple researchers, and (iii) the evolution of the research technology over time.

At time $t \geq 1$, the set of questions available for study is Q_t. Once a question has entered the field, it always remains, so $Q_{t-1} \subseteq Q_t$. The date at which a question enters the field is τ_q. There are N_t researchers in each period who each decide on a question to study. Each researcher lives for one period and then dies.

Let the set of researchers at time t who decide to study question $q \in Q_t$ be denoted as $R_{q,t}$, and let $|R_{q,t}| = n_{q,t} \geq 0$, where $\sum_{q \in Q_t} n_{q,t} = N_t$. Each question q continues to be defined by the number μ_q, which represents "the truth." When a question first enters the field, researchers form a common prior $\mu_{q,0}$, which is distributed normally with mean zero and variance $\sigma^2_{q,0}$.

A researcher's life is as follows: She is born and observes four variables for each available question, $\{\tilde{\sigma}^2_{r,q,t}, \tilde{\mu}_{q,t}, \sigma^2_{q,t-1}, V_q\}_{q \in Q_t}$. Each number corresponds to what the researcher observes in the one-period model; $\tilde{\sigma}^2_{r,q,t}$ is her prior for the quality of her own research on question q, $\tilde{\mu}_{q,t}$ is her prior for the truth about question q, $\sigma^2_{q,t-1}$ is the variance of $\tilde{\mu}_{q,t}$, and V_q is the importance of the question. She uses this information to choose a question $q^* \in Q_t$ and observes a signal $\hat{\mu}_{r,q^*,t}$ which is normally distributed with mean μ_{q^*} and variance $\sigma^2_{r,q^*,t}$.[9] She then combines her prior and signal into a posterior. I assume that all past signals for all questions are publicly observable. Furthermore, the variances of those signals are observable as well: both past work and its quality are common knowledge. However, in period t, the only party privy to research at time t is the person doing the research. Furthermore, the quality of that research is not yet known, even to her.

4.1. The Formation of Priors and Posteriors

Because knowledge accumulates over time, a researcher's priors are formed by observing past research. In this section, I derive the values of a researcher's priors for each question. A question may never have been researched before period t, either because it is new or because no researcher has ever chosen it. In this case, the common prior for the question has variance $\sigma^2_{q,0}$. If there has been research on a question prior to time t, then that research is publicly observable and there is

[9] Allowing researchers to study more than one question would not qualitatively affect the results, so long as they cannot study *every* question.

a common Bayesian prior for the question. The precision for that prior is simply the sum of the precisions of the signals that comprise it:

$$1/\sigma^2_{q,\,t-1} = 1/\sigma^2_{q,\,0} + \sum_{s=\tau_q}^{t-1} \sum_{r \in R_{q,\,s}} 1/\sigma^2_{r,\,q,\,s}.$$

The prior for the truth about question q at time t, common to all researchers, is just the weighted average of the initial prior and the signals about q, where the weight of each signal is proportional to the precision of the signal:

$$\tilde{\mu}_{q,\,t} = \frac{1/\sigma^2_{q,\,0}}{1/\sigma^2_{q,\,t-1}} \mu_{q,0} + \sum_{s=\tau_q}^{t-1} \sum_{r \in R_{q,\,s}} \frac{1/\sigma^2_{r,\,q,\,s}}{1/\sigma^2_{q,\,t-1}} \hat{\mu}_{r,\,q,\,s}.$$

Researcher r's prior of the variance of her signal of μ_q at time t is $\tilde{\sigma}^2_{r,\,q,\,t} \sim G_{r,\,q,\,t}(\cdot)$, where $G_{r,\,q,\,t}$ is increasing and differentiable over the entire support $[0, \bar{G}]$, with $G_{r,\,q,\,t}(0) = 0$ and $G_{r,\,q,\,t}(\bar{G}) = 1$. Note that $G_{r,\,q,\,t}$ is unknown to the researcher, but researchers are, on average, unbiased in their beliefs: $E_{G_{r,\,q,\,t}}(\tilde{\sigma}^2_{r,\,q,\,t}) = \sigma^2_{r,\,q,\,t}$. Researchers differ and choose different questions, because $\tilde{\sigma}^2_{r,\,q,\,t} \neq \tilde{\sigma}^2_{\rho,\,q,\,t}$ for $r \neq \rho$.

Researcher r's posterior for her chosen question q^* is simply the weighted average of the common prior and her signal:

$$\mu_{r,\,q^*,\,t,\,\text{posterior}} = \frac{1/\sigma^2_{q^*,\,t-1}}{1/\sigma^2_{q^*,\,t-1} + 1/\tilde{\sigma}^2_{r,\,q^*,\,t}} \tilde{\mu}_{q^*,\,t} + \frac{1/\tilde{\sigma}^2_{r,\,q^*,\,t}}{1/\sigma^2_{q^*,\,t-1} + 1/\tilde{\sigma}^2_{r,\,q^*,\,t}} \hat{\mu}_{r,\,q^*,\,t}.$$

The analysis will center on the quality of that posterior. Before solving for equilibrium decisions and outcomes in the model, I discuss three assumptions. First, questions are exogenously generated over time. Although this is unrealistic, it is a useful simplification.

Second, researchers agree on the facts about past research. This assumption may not be correct in practice, but allowing disagreement about the quality of past work would only serve to reinforce the primary conclusions of the paper while imposing a more significant departure from the rational model.

Third, as in the one-period model, I assume that, in forming an opinion of the quality of her work, a researcher assumes that the variance of her signal equals the prior for that variance. Also, as before, the model could be adjusted so that she gets a signal of $\sigma^2_{r,\,q,\,t}$ after she observes $\hat{\mu}_{r,\,q^*,\,t}$. So long as the eventual weight of $\hat{\mu}_{r,\,q^*,\,t}$ in the posterior depends at least slightly on her prior for $\sigma^2_{r,\,q,\,t}$, the qualitative results would stand. This assumption is weaker than it may first appear: there is no evidence that a researcher could observe that her prior is incorrect; there is only a theoretical argument. Moreover, even if she could be persuaded by the theoretical argument, there is no way for her to know how much to adjust her beliefs about $\sigma^2_{r,\,q,\,t}$ upward. Note that $G_{r,\,q,\,t}$ is unknown, unique to each researcher, and not drawn from a larger distribution. Any adjustment would be arbitrary and incorrect with probability 1. A perfectly rational and informed researcher could say, at best, "I know that I am likely to be biased in favor of my own work but cannot provide any estimate of how much."

4.2. The Decision to Research a Question q

I assume, as before, that the researcher cares about how much she expects to advance knowledge: $u_{r,\,q,\,t} = u(V_q, \sigma^2_{q,\,t-1}, \tilde{\sigma}^2_{r,\,q,\,t})$, where u is increasing in the importance of the question, V_q; increasing in the variance of her prior, $\sigma^2_{q,\,t-1}$; and decreasing in her prior of the variance of the signal for her chosen question. Note that the researcher's utility depends only on how her research improves knowledge about a question relative to where it was when she began her work. It does not depend on what other researchers are learning contemporaneously, what research she expects to be done in the future, or what would have been done if she had chosen a different question. Results would be similar if any of these three alternatives were incorporated, but at the expense of significant added complexity. By using a utility function that ignores others' choices, I can use partial equilibrium analysis rather than searching for Nash equilibria.

Lemma 4. *A researcher is weakly more likely, ceterus paribus, to study question q if*

(1) *she believes her signal will be more precise: $\tilde{\sigma}^2_{r,\,q,\,t}$ is lower;*

(2) *the question is newer: $t - \tau_q$ is lower;*

(3) *fewer researchers have worked on the question historically: $n_{q,\,s}$ is lower for any $s \in \{\tau_q, \tau_q + 1, \ldots, t-1\}$;*

(4) *historic signals were less precise: $\sigma^2_{r,\,q,\,s}$ is higher, $s < t$, for any r; and*

(5) *the question is more important: V_q is higher.*

Corollary 3. *Researchers are more likely to study questions that experience an improvement in the research technology.*

An improvement in the research technology may be, in practice, an increase in the (i) quality of available data, (ii) quantity of available data, or (iii) available methods of analysis. An improvement in the research technology implies more productive research on the question and therefore more people studying it.

These results are straightforward and, rather than being seen as novel aspects of the model, should be seen as confirming the sensibility of the assumptions.

4.3. The Precision of Expert and Nonexpert Opinion

As in the one-period model, for any set of priors and parameter values, there will be a threshold variance such that a researcher chooses question q^* if her prior for the variance of her signal of q^* is less than the threshold, and she will not study it otherwise. Formally stated, for any set of parameters regarding alternative questions $\{V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t}\}_{q\in Q_t\setminus q^*}$, and for any values for q^* $\{V_{q^*}, \sigma^2_{q^*,t-1}\}$, there exists a threshold value $\bar{\sigma}^2_{r,q^*,t} \geq 0$ such that researcher r studies question q^* if $\tilde{\sigma}^2_{r,q^*,t} < \bar{\sigma}^2_{r,q^*,t}$ and does not study it if $\tilde{\sigma}^2_{r,q^*,t} > \bar{\sigma}^2_{r,q^*,t}$. This means that if a change in some parameter increases (decreases) the threshold $\bar{\sigma}^2_{r,q^*,t}$, then $E(\tilde{\sigma}^2_{r,q^*,t} \mid q^* \text{ chosen})$ must increase (decrease) as well. Because it is possible that $\bar{\sigma}^2_{r,q^*,t} = 0$ (the question is never studied) or $\bar{\sigma}^2_{r,q^*,t}$ is very large (the question is always studied), the comparative statics will only be strict for interior values of $\bar{\sigma}^2_{r,q^*,t}$, i.e., those for which the question may or may not be studied, depending on $\bar{\sigma}^2_{r,q,t}$.

Proposition 2. *If a question q^* is sometimes studied, then as the level of knowledge about q^* increases, expected expert bias increases. That is, if $\bar{\sigma}^2_{r,q^*,t} \in (0, \bar{G})$ for researcher r, then $(d/d\sigma^2_{q^*,t-1})E(\tilde{\sigma}^2_{r,q^*,t} \mid q^* \text{ chosen}) < 0$.*

Proposition 2 establishes that as researchers learn more about a question, the bias of the researchers who continue to study it increases as well. Therefore, anything that increases the precision of knowledge about a question will also increase the level of bias among researchers studying the question. The following results are immediate.

Corollary 4. *Expected expert bias is higher, ceterus paribus, if*

(1) *the question is older: t is higher for any given τ;*

(2) *more researchers have worked on the question historically: $n_{q,s}$ is higher for any $s \in \{\tau_q, \tau_q+1, \ldots, t-1\}$;*

(3) *historic signals were more precise: $\sigma^2_{r,q,s}$ is lower for any $s \in \{\tau_q, \tau_q+1, \ldots, t-1\}$; and*

(4) *the question is less important: V_q is lower.*

Just as in the one-period case, if q^* is less important, then fewer researchers will choose to study it. Those that do will be more biased. Also just as in the one-period case, a stronger prior for the truth about q^* causes fewer researchers to choose it. Unlike the one-period case, having multiple periods allows us to be more precise about what causes a prior to be strong. Questions that are old have had more time to be researched and thus allow for stronger priors. Similarly, for questions of any given age, if more people worked on a question in the past, the current prior will be more precise. Finally, if a given number of people have studied a question, then the more precise their signals were, the more precise the current prior for the question.

Just as in the one-period case, expert opinion can be less accurate than nonexpert opinion for a wide variety of types of nonexpert.

Corollary 5. *An expert's posterior for her chosen question q^* can be less precise, in expectation, than her prior.*

Corollary 6. *An expert's posterior for q^* can be less precise, in expectation, than the opinion of a dilettante, a person with access to a random draw of at least one prior signal for q^*.*

The intuition for these results is simple and is the same as in the one-period case. The threshold such that question q^* is chosen can be arbitrarily close to zero. This means that the expected value of a researcher's prior for the variance of her signal can be arbitrarily close to zero. This means that she may place up to full weight on her own signal when forming a posterior for μ_{q^*}, yielding a precision of her posterior of $1/\tilde{\sigma}^2_{r,q^*,t}$. The precision of her prior, before doing research, is $1/\sigma^2_{q^*,t-1} = 1/\sigma^2_{q^*,0} + \sum_{s=\tau_{q^*}}^{t-1}\sum_{r\in R_{q^*,s}} 1/\sigma^2_{r,q^*,s}$, which is greater than $1/\tilde{\sigma}^2_{r,q^*,t}$ so long as her signal is not, in fact, more precise than the combined signals of all past researchers.

Not only are experts biased but they may also become so biased that they are less reliable than (i) their past selves, who had a good understanding of past research but were not privy to new work, and (ii) dilettantes, who are aware of only some past work. Indeed, this may be quite common in practice, as many questions are "settled" in the minds of many researchers who have moved elsewhere for research questions. Tetlock (2006) finds extensive evidence that dilettantes have superior predictive power to subject-matter experts, across a range of fields, time periods, and types of question. I would suggest that this model speaks to these data, and it also provides a suggestion for what further tests to run on the data.

Questions for which expert opinion is less accurate than nonexpert opinion feature few researchers, whereas those for which experts are more accurate feature many researchers. Therefore, although a large majority of *researchers* may work on questions for which they truly are the best source of information, it could be the case that the majority of *questions* feature experts who are less reliable than nonexperts.

Thus far I have focused on the bias generated by properties of the question at the start of period t, but the likely productivity of current research is important as well. Holding constant all parameters except for $\tilde{\sigma}^2_{r,q^*,t}$, we can evaluate how the level of bias changes as the productivity of current research, $\sigma^2_{r,q^*,t}$, changes. To do this, let $\tilde{\sigma}^2_{r,q^*,t} = K_{r,q^*,t}\sigma^2_{r,q^*,t}$

and let $E(K_{r,q^*,t}) = 1$. The distribution of $\sigma^2_{r,q^*,t}$ is $G_{r,q^*,t}(\cdot)$, so the distribution of $K_{r,q^*,t}$ can be denoted as $H_{r,q^*,t}(x/\sigma^2_{r,q^*,t}) \equiv G_{r,q^*,t}(x)$, which has support $[0, \bar{G}/\sigma^2_{r,q^*,t}]$ and density $h_{r,q^*,t}(\cdot)$. The following result is immediate.

Proposition 3. *As the precision of current research on a question increases, expected expert bias decreases.*

Proposition 3 should not be surprising: as the productivity of current research rises, more researchers will choose question q. The marginal researcher is less optimistic than the average. Expected bias therefore falls. This fact has important empirical implications. When research on a question becomes more productive, whether via newly available data, new techniques, new paradigms, etc., researchers studying the question will be more accurate in their opinions. I explore this result further in §5 as it relates to the age of researchers.

4.4. The Relationship Between the Field and the Question

I have focused thus far on how the properties of the question affect the choice of potential researchers to study the question and the quality of their posteriors. An equally interesting analysis focuses on the properties of the field more generally and how that affects bias on a question. Fast-growing fields will feature particularly biased experts when it comes to older questions. The reason is that quickly growing fields will siphon off many researchers into new questions where the fruit is low hanging. Researchers who choose to study old questions in quickly expanding fields must believe that the precision of their signals about those old questions is very high to be willing to forgo the option to research new questions. On average, this implies that they are significantly biased. In slowly growing fields, there are few new questions to research, so researchers are forced to study old questions, regardless of their bias. The average researcher on one of these questions, then, is not particularly biased.

For example, medicine is a field that is currently growing quickly, with new questions, methods, and opportunities arising frequently. One must wonder, therefore, who continues to study the effect of cigarettes on lung cancer. Although the question is not completely understood (nor should it ever be, according to the model), it is highly studied, and future work has a low likelihood of impact. Researchers who choose to study it now are likely very biased, more so because other opportunities abound.

For another example, history is a field that is growing slowly, with only one year of new history being created each year. A young researcher must choose some question to study, but there are few opportunities for seminal work in a field with such slow question generation. Therefore, a researcher choosing to study the fall of the Roman Empire is likely to be less biased *because* there are few other opportunities for work that are clearly superior.

Proposition 4. *For old questions, experts are more biased if the field is growing more quickly: $E(\tilde{\sigma}^2_{r,q,t} \mid q$ chosen$)$ is weakly decreasing in the number of available questions, $|Q_t|$, ceterus paribus.*

4.5. Expert Confidence and Accuracy

I have focused thus far on questions of expert accuracy and ways to predict expert accuracy from properties of the question and the field. Another way one might predict such accuracy is to simply ask the expert, "How accurate do you think your beliefs are?" One might expect that experts who are more accurate would have more confidence in their posteriors, so this confidence would be an additional useful signal. As discussed in the introduction, however, empirical work in psychology suggests that experts that are more confident are actually less accurate, on average. I show in this section that this relationship arises endogenously in the model.

Because past reports and their precisions are common knowledge, experts will report greater precision of their posteriors when their draws of $\tilde{\sigma}^2_{r,q,t}$ are lower. As $\tilde{\sigma}^2_{r,q,t}$ approaches $\sigma^2_{r,q,t}$ from above, experts move from putting too little weight on current work to putting the correct weight on current work. They will *report* greater posterior precision and will *actually have* greater posterior precision. As $\tilde{\sigma}^2_{r,q,t}$ passes $\sigma^2_{r,q,t}$ and continues to fall, experts will report greater precision, but their posteriors will have *less* precision: they put too much weight on current work. In this range, then, there is a negative relationship between reported accuracy and actual accuracy.

When the field is large and there are many interesting questions, most researchers will choose a question for which $\tilde{\sigma}^2_{r,q,t} < \sigma^2_{r,q,t}$. Therefore, the empirical domain of $\tilde{\sigma}^2_{r,q,t}$ will tend to be that for which there is a negative relationship between expert confidence and accuracy. This widely observed phenomenon arises endogenously, even though I assume that people are not biased on average.

5. Applications

In this section, I discuss a variety of implications that these insights have for real-world practice. For the sake of brevity, and because the implications should be obvious, with a bit of reflection, I do not prove these claims formally.

5.1. Application: Golden Cohorts

There will sometimes arise a golden cohort of researchers working on a question, a group whose opinions are more accurate than contemporaries at all points of their careers. These cohorts arise when there is a significant drop in σ_q^2 at the time they are choosing questions.

This drop makes current work impactful and draws in researchers, even if they are not particularly biased, making young researchers for q less biased than old. This cohort stays with the question over time, and these researchers remain less biased than younger cohorts who start their careers when more is known about the question, and incremental research is therefore less valuable.

Finding golden cohorts requires finding situations where there are significant improvements in data or techniques in a short span of time. The marginal revolution of the 1890s, the game theory revolution in the 1970s, and the new trade theory of the 1980s provide potential examples in the field of economics. These tools for thinking about prices, information, and trade policy, respectively, were associated with cohorts of researchers that have been seen since as leading experts on their respective questions.

A recent, important example may be business-cycle macroeconomics. Macroeconomic data are not generated quickly—recessions are blissfully infrequent. It can be argued that, at least in the West, there was little new business-cycle data generated in the decades leading to the great recession. The recession, however, provides a wealth of data, as different states and countries have experienced different types and degrees of shocks and responded with highly varied monetary and fiscal policies. It remains to be seen, though I believe it likely, whether this recession yields a golden cohort in business-cycle macroeconomics.

I do not argue that this is the only explanation for the rise of golden cohorts. An alternative explanation is that young researchers have more of an incentive to invest in new technologies, allowing them to eclipse older researchers. As the technology ages, new young researchers find themselves in the diminishing returns phase of a technology, leading them elsewhere. These explanations for the existence of golden cohorts are not mutually exclusive.

5.2. Application: Academic vs. Professional Experts

Expert opinion can be less precise than nonexpert opinion because selection into studying a question yields researchers who are biased in favor of their own research. This naturally implies that the problem can be solved by changing researchers' utility functions. I asserted at the outset that researchers are motivated by the search for knowledge, but there are clearly other motivations, such as money, that I could have assumed. Professional experts may choose questions for which there is significant prior research, because this helps them form accurate—and profitable—beliefs. Academics shy away from these questions because the marginal impact of their work is likely to be low. Therefore, when choosing an expert to query, the manager or policy maker must first determine how heavily researched is the question.

Examples may include macroeconomic or financial forecasting. These areas are well established, and there is significant financial reward for the successful forecaster. When a policy maker wants to know the effect of a policy on GDP growth or interest rates, for example, she may want to ask a professional rather than an academic.

5.3. Application: Task Assignment and Hiring

Managers often face a problem of task assignment, one important element of which is the delegation of decision making to a worker. For example, consider a lender assigning employees to build a risk model. There are many potential risk factors for a borrower, such as current purchasing behavior (what stores does the borrower currently visit?), pay-down behavior (to what creditors is the borrower directing payments?), traveling behavior (is the borrower moving around more than usual, or does the borrower appear to have moved without notifying the lender?), etc. One task assignment option would be to allow the employees complete discretion in choosing research areas and proposing modeling solutions. Another option would be assigning employees to specific areas and allowing discretion only within a narrow band.

It is reasonable to assume that employees would be weakly harder working if allowed to choose their focuses, but there could be significant bias implicit in their choices. Suppose that pay-down behavior has been heavily studied at the lender, and many elements of pay-down behavior are already used in the risk model. Then an employee given discretion who still chooses to focus on pay-down behavior may be severely biased. The model suggests, therefore, that young firms with many novel jobs to do should allow more employee discretion, whereas older firms, whose jobs are more "maintenance" than "construction," should allow less.

These insights apply to hiring as well. Consider a hedge fund hiring a Ph.D. for a quantitative modelling role. The fund could hire a Ph.D. in finance or a Ph.D. in physics. The physics Ph.D. likely has an intrinsic interest in physics but is joining the hedge fund for money. The Ph.D. in finance may join the firm better prepared, needing less training in the details of financial markets, but has clearly chosen the questions of finance, in part, for intrinsic reasons. Depending on the degree of relative intrinsic motivation, the fund may prefer either employee.

6. Conclusion

This paper is dedicated to the development and application of a simple insight: the selection into an area of expertise causes a bias on the part of experts. When the selection is strong, experts may not, in fact, be expert, in the sense that they are not the best sources of available information. This insight immediately raises two questions. First, how can a policy maker or manager distinguish between experts who are more and less reliable, given that the experts themselves cannot? Properties of the question are useful in this regard: questions that are less important, more highly researched, and for which there are little new data, few new tools, etc., are subject to more severe selection bias, and they imply more questionable expertise. These implications are fairly general and should apply in essentially any model that captures the basic insight of this paper. Attributes of the expert are also useful: experts who began their research when new data and tools were becoming available and those who are extrinsically motivated will often be more reliable. Finally, an expert's self-reported confidence in her posterior is also a useful guide to her accuracy. Experts who express greater confidence are less precise. This implication of the model aligns well with empirical work and arises endogenously.

Second, how should we design mechanisms to solve this problem? There are a surprising number of settings where this insight has implications for mechanism design, within and outside academia. There are clear implications for hiring, task-assignment, and even the refereeing process. Journals should generally require both expert (referee) and nonexpert (editor) opinions. The expert judges the quality of work within the question (e.g., by observing a signal of the value of $\sigma^2_{r,q,t}$ associated with the report $\hat{\mu}_{r,q,t}$), and the nonexpert judges the likely bias among researchers studying the question more generally. Because experts disagree, it may be useful to ask multiple experts (referees). Because nonexperts agree, one nonexpert (editor) is sufficient.

It is important to note that the fact that research experts tend to be biased does not suggest that the method by which people choose to become researchers should be changed. Knowledge about questions increases over time as those questions are researched, regardless of whether the contemporaneous researchers themselves have a biased view of the quality of their work. It is only in soliciting their views that outsiders must be careful to account for this bias.

Acknowledgments

The author thanks Andrew Ang, Nina Baranchuk, David Dicks, Diego Garcia, Naveen Khanna, Chris Parsons, Luigi Zingales, and seminar participants at Wharton, Michigan State, the University of Texas at Dallas, Vanderbilt, the University of British Columbia, Frankfurt School of Finance and Management, Vienna Graduate School of Finance, and the University of Colorado Boulder for helpful advice. All errors are the author's own.

Appendix. Proofs

Proof of Lemma 2. Note that

$$q^* \in \arg\max_q u(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1}).$$

Therefore, anything that increases her utility from choosing a research question weakly increases the likelihood of her choosing that question. And anything that decreases her utility from choosing an alternative weakly increases the likelihood of her choosing that question. By assumption, u_q is increasing in V_q, increasing in $\sigma^2_{q,0}$, and decreasing in $\tilde{\sigma}^2_{q,1}$. Also by assumption, $u_{q'}$ is increasing in $V_{q'}$, increasing in $\sigma^2_{q',0}$, and decreasing in $\tilde{\sigma}^2_{q',1}$. □

Proof of Lemma 1. Taking a derivative of the variance of the posterior with respect to the prior for the precision of her signal yields

$$\frac{d}{d(1/\tilde{\sigma}^2_{q^*,1})}\frac{1/\sigma^2_{q^*,0}+\sigma^2_{q^*,1}/(\tilde{\sigma}^2_{q^*,1})^2}{(1/\sigma^2_{q^*,0}+1/\tilde{\sigma}^2_{q^*,1})^2}$$
$$=2\frac{\sigma^2_{q^*,1}/\tilde{\sigma}^2_{q^*,1}-1}{(1/\sigma^2_{q^*,0}+1/\tilde{\sigma}^2_{q^*,1})^3\sigma^2_{q^*,0}}.$$

This derivative equals zero when her prior is correct: $\sigma^2_{q^*,1}=\tilde{\sigma}^2_{q^*,1}$. It is negative when $\sigma^2_{q^*,1}/\tilde{\sigma}^2_{q^*,1}<1$, which is when she believes her work is less accurate than it is, and positive when $\sigma^2_{q^*,1}/\tilde{\sigma}^2_{q^*,1}>1$, which is when she believes that her work is more accurate than it really is. Therefore $\sigma^2_{q^*,1}=\tilde{\sigma}^2_{q^*,1}$ is a minimum. □

Proof of Lemma 3. Note that

$$q^*=\arg\max_q u(V_q, \sigma^2_{q,0}, \tilde{\sigma}^2_{q,1}).$$

Define an alternative vector of priors $\{s^2_{q,1}\}_{q\in Q}$, which equals $\{\tilde{\sigma}^2_{q,1}\}_{q\in Q}$ except for question q^*. Let $s^2_{q^*,1}<\tilde{\sigma}^2_{q^*,1}$. Because $u(V_q,\sigma^2_{q,0},\tilde{\sigma}^2_{q,1})$ is decreasing in $\tilde{\sigma}^2_{q,1}$, if $q^*=\arg\max_q u(V_q,\sigma^2_{q,0},\tilde{\sigma}^2_{q,1})$, then $q^*=\arg\max_q u(V_q,\sigma^2_{q,0},s^2_{q,1})$. Similarly, if we instead let $s^2_{q^*,1}>\tilde{\sigma}^2_{q^*,1}$, then if $q^*\neq\arg\max_q u(V_q,\sigma^2_{q,0},\tilde{\sigma}^2_{q,1})$, then $q^*\neq\arg\max_q u(V_q,\sigma^2_{q,0},s^2_{q,1})$. Taken together, these two facts mean that there exists a threshold value $\bar{\sigma}^2_{q^*,1}(\{V_q,\sigma^2_{q,0},\tilde{\sigma}^2_{q,1}\}_{q\in Q\backslash q^*},V_{q^*},\sigma^2_{q^*,0})$ such that question q^* is chosen if $\tilde{\sigma}^2_{q^*,1}<\bar{\sigma}^2_{q^*,1}$ and is not chosen if $\tilde{\sigma}^2_{q^*,1}>\bar{\sigma}^2_{q^*,1}$. Because the support of $\tilde{\sigma}^2_{q^*,1}$ is $[0,\infty)$ and $E(\tilde{\sigma}^2_{q^*,1})=\sigma^2_{q^*,1}$, for any finite $\bar{\sigma}^2_{q^*,1}$, $E(\tilde{\sigma}^2_{q^*,1}\mid\tilde{\sigma}^2_{q^*,1}<\bar{\sigma}^2_{q^*,1})<\sigma^2_{q^*,1}$. □

Proof of Proposition 1. Because q^* is sometimes chosen, $\bar{\sigma}^2_{q^*,1}$ is nonzero and finite. From Corollary 1, the precision of the posterior is decreasing as $\tilde{\sigma}^2_{q^*,1}$ decreases for $\tilde{\sigma}^2_{q^*,1}<\sigma^2_{q^*,1}$. From Lemma 3, $E(\tilde{\sigma}^2_{q^*,1}\mid\tilde{\sigma}^2_{q^*,1}<\bar{\sigma}^2_{q^*,1})<\sigma^2_{q^*,1}$ for a finite $\bar{\sigma}^2_{q^*,1}$. Further, $E(\tilde{\sigma}^2_{q^*,1}\mid\tilde{\sigma}^2_{q^*,1}<\bar{\sigma}^2_{q^*,1})$ is increasing in $\bar{\sigma}^2_{q^*,1}$. Therefore, as $\bar{\sigma}^2_{q^*,1}$ increases, the expected precision of expert opinion for question q^* increases; $\bar{\sigma}^2_{q^*,1}$ is increasing in V_{q^*} and decreasing in $\sigma^2_{q^*,0}$, which establishes the proposition. □

Proof of Corollary 2. Let parameters be such that $\bar{\sigma}^2_{q^*,1} \to 0$. Then, conditional on choosing question q^*, the researcher will put full weight on her signal in forming a posterior, and that posterior will have variance approaching $\sigma^2_{q^*,1}$. The prior has variance of $\sigma^2_{q^*,0}$. If $\sigma^2_{q^*,0} < \sigma^2_{q^*,1}$, then the posterior is inferior to the prior. □

Proof of Lemma 4. Note that

$$q^* = \arg\max_q u(V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t}).$$

Therefore, anything that increases her utility from choosing a research question weakly increases the likelihood of her choosing that question. And anything that decreases her utility from choosing an alternative weakly increases the likelihood of her choosing that question. Note that $u(V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t})$ is increasing in V_{q^*}, increasing in $\sigma^2_{q^*,t-1}$, and decreasing in $\tilde{\sigma}^2_{r,q^*,t}$. Also by assumption, $u_{q'}$ is increasing in $V_{q'}$, increasing in $\sigma^2_{q',t-1}$, and decreasing in $\tilde{\sigma}^2_{r,q',t}$. Because researchers are Bayesian, the prior for any question q has a variance satisfying $1/\sigma^2_{q,t-1} = 1/\sigma^2_{q,0} + \sum_{s=\tau_q}^{t-1}\sum_{r\in R_{q,s}} 1/\sigma^2_{r,q,s}$. Therefore, $\sigma^2_{q,t-1}$ is increasing as elements are removed from $R_{q,s}$ for $s \in \{\tau_q, \tau_q+1, \ldots, t-1\}$. Holding $R_{q,s}$ and $\sigma^2_{r,q,s}$ constant for each period s, reducing the difference $t-\tau_q$ increases $\sigma^2_{q,t-1}$. In each case, a higher value of $\sigma^2_{q,t-1}$ implies a higher value of $u(V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t})$ and a higher likelihood of q being chosen. □

Proof of Proposition 2. It is sufficient to show that, for every set of values $\{\sigma^2_{q,t-1}, V_q\}_{q\in Q_t\backslash q^*}$ and V_{q^*}, if $\bar{\sigma}^2_{r,q^*,t} > 0$, then it is increasing in $\sigma^2_{q^*,t-1}$. Let question q' be the best available alternative to question q^* for researcher r:

$$q'(r,q,t) = \arg\max_q u(V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t})_{q\in Q_t\backslash q^*}.$$

Then $\bar{\sigma}^2_{r,q^*,t}$ satisfies $u(V_{q'}, \sigma^2_{q',t-1}, \tilde{\sigma}^2_{r,q',t}) = u(V_{q^*}, \sigma^2_{q^*,t-1}, \bar{\sigma}^2_{r,q^*,t})$. Write the threshold value $\bar{\sigma}^2_{r,q^*,t}$ as a function of $\sigma^2_{q^*,t-1}$ and take a total derivative of both sides of the equation to get

$$\begin{aligned}
&\frac{d}{d\sigma^2_{q',t-1}} u(V_{q'}, \sigma^2_{q',t-1}, \tilde{\sigma}^2_{r,q',t})\\
&\quad = \frac{d}{d\sigma^2_{q^*,t-1}} u(V_{q^*}, \sigma^2_{q^*,t-1}, \bar{\sigma}^2_{r,q^*,t}(\sigma^2_{q^*,t-1}))\\
0 &= u_2(V_{q^*}, \sigma^2_{q^*,t-1}, \bar{\sigma}^2_{r,q^*,t}(\sigma^2_{q^*,t-1}))\frac{d\bar{\sigma}^2_{r,q^*,t}(\sigma^2_{q^*,t-1})}{d\sigma^2_{q^*,t-1}}\\
&\quad + u_3(V_{q^*}, \sigma^2_{q^*,t-1}, \bar{\sigma}^2_{r,q^*,t}(\sigma^2_{q^*,t-1})),
\end{aligned}$$

which yields

$$\frac{d\bar{\sigma}^2_{r,q^*,t}}{d\sigma^2_{q^*,t-1}} = -\frac{u_3}{u_2}.$$

Because $u_3 > 0$ and $u_2 < 0$, the right-hand side is positive, implying that $d\bar{\sigma}^2_{r,q^*,t}/d\sigma^2_{q^*,t-1} > 0$. □

Proof of Corollary 4. Note that $1/\sigma^2_{q,t-1} = 1/\sigma^2_{q,0} + \sum_{s=\tau_q}^{t-1}\sum_{r\in R_{q,s}} 1/\sigma^2_{r,q,s}$ because researchers are Bayesian; $\sigma^2_{q,t-1}$ is increasing as elements are removed from $R_{q,s}$ for $s \in \{\tau_q, \tau_q+1, \ldots, t-1\}$. Holding $R_{q,s}$ and $\sigma^2_{r,q,s}$ constant for each period s, reducing the difference $t-\tau_q$ increases $\sigma^2_{q,t-1}$. And holding $R_{q,s}$ constant, higher values of $\sigma^2_{r,q,s}$ imply higher values of $\sigma^2_{q,t-1}$. Because $d\bar{\sigma}^2_{r,q,t}/d\sigma^2_{q,t-1} > 0$ (from the proof of Proposition 2), these three facts imply a higher threshold for choosing the question and less biased researchers. The remaining element is V_q, for which the result follows immediately from Proposition 4. □

Proof of Corollary 5. Because $\bar{\sigma}^2_{r,q^*,t}$ can be arbitrarily close to zero, $E(\tilde{\sigma}^2_{r,q^*,t} \mid \tilde{\sigma}^2_{r,q^*,t} \le \bar{\sigma}^2_{r,q^*,t})$ can be arbitrarily close to zero. The variance of expert opinion on a selected question q^* is therefore

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1/\sigma^2_{q^*,t-1}}{1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t}}\tilde{\mu}_{q^*,t} + \frac{1/\tilde{\sigma}^2_{r,q^*,t}}{1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t}}\hat{\mu}_{r,q^*,t}\right)\\
&= \mathrm{Var}\left(\frac{1/\sigma^2_{q^*,t-1}}{1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t}}(\mu_{q^*}+z_{q^*,0})\right.\\
&\qquad\left. + \frac{1/\tilde{\sigma}^2_{r,q^*,t}}{1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t}}(\mu_{q^*}+z_{q^*,1})\right)\\
&= \left(\frac{1/\sigma^2_{q^*,t-1}}{1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t}}\right)^2\sigma^2_{q^*,0} + \left(\frac{1/\tilde{\sigma}^2_{r,q^*,t}}{1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t}}\right)^2\sigma^2_{q^*,1}\\
&= \frac{1/\sigma^2_{q^*,t-1}+\sigma^2_{r,q^*,1}/(\tilde{\sigma}^2_{r,q^*,t})^2}{(1/\sigma^2_{q^*,t-1}+1/\tilde{\sigma}^2_{r,q^*,t})^2}.
\end{aligned}$$

Letting $\tilde{\sigma}^2_{r,q^*,t} \to 0$, and applying L'Hôpital's rule, the variance of expert opinion is $\sigma^2_{r,q^*,t}$: she puts full weight on her signal and no weight on past work. The variance of the prior is $\sigma^2_{q^*,t-1}$, which can be arbitrarily close to zero. Therefore, it is possible that $\sigma^2_{r,q,t} > \sigma^2_{q^*,t-1}$, and the expert's posterior is less precise than her prior. □

Proof of Corollary 6. From the proof of Corollary 5, as parameters drive $\bar{\sigma}^2_{q^*,t-1} \to 0$, the expert's posterior will have variance approaching $\sigma^2_{r,q^*,t}$. A dilettante observes a random sample of at least one past signal. If the result is true for a dilettante that observes one prior signal, it will be true for a dilettante that observes more than one, so we restrict attention to the case of a single past signal. The total number of past signals on q^* is S. The expected precision of the dilettante's belief is $1/\sigma^2_{q^*,0} + (1/S)\sum_{s=\tau_{q^*}}^{t-1}\sum_{r\in R_{q^*,s}} 1/\sigma^2_{r,q^*,s}$. Suppose that $\sigma^2_{r,q^*,s} = \sigma^2_{q^*,0}$ for all r, s. Then the dilettante's belief has precision $2/\sigma^2_{q^*,0}$, and the expert's belief has precision $1/\sigma^2_{q^*,0}$. This establishes the possibility that a dilettante's belief is more accurate, in expectation, than an expert's. □

Proof of Proposition 3. As $\sigma^2_{r,q,t}$ decreases, $G_{r,q,t}(x)$ increases for all $x \in [0, \bar{G}/\sigma^2_{r,q,t}]$. I show that expected bias is increasing in $\sigma^2_{r,q,t}$ for any given values $\{\sigma^2_{q',t-1}, V_{q'}\}_{q'\in Q_t\backslash q}$ and V_q, so it is increasing in $\sigma^2_{r,q,t}$. The expected bias given $\{\sigma^2_{q',t-1}, V_{q'}\}_{q'\in Q_t\backslash q}$ and V_q is

$$E(K \mid r \text{ chosen}) = \frac{1}{H_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})}\int_0^{\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t}} x h_{r,q,t}(x)\,dx.$$

Taking a derivative with respect to $\sigma^2_{r,q,t}$ yields

$$\begin{aligned}
&\frac{d}{d\sigma^2_{r,q,t}} E(K \mid r \text{ chosen}) \\
&\quad = \frac{h_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})\bar{\sigma}^2_{r,q,t}}{[H_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})\sigma^2_{r,q,t}]^2} \int_0^{\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t}} x h_{r,q,t}(x)\,dx \\
&\qquad - \frac{h_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})}{H_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})\sigma^2_{r,q,t}} \left(\frac{\bar{\sigma}^2_{r,q,t}}{\sigma^2_{r,q,t}}\right)^2 \\
&\quad = \frac{h_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})}{H_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})} \left[\int_0^{\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t}} x h_{r,q,t}(x)\,dx - \frac{\bar{\sigma}^2_{r,q,t}}{\sigma^2_{r,q,t}}\right] \\
&\quad = \frac{h_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})}{H_{r,q,t}(\bar{\sigma}^2_{r,q,t}/\sigma^2_{r,q,t})} \left[E\left(K \,\middle|\, K \le \frac{\bar{\sigma}^2_{r,q,t}}{\sigma^2_{r,q,t}}\right) - \frac{\bar{\sigma}^2_{r,q,t}}{\sigma^2_{r,q,t}}\right].
\end{aligned}$$

The first term is positive and the second is negative, so the product is negative. □

Proof of Proposition 4. Recall that the threshold value of $\tilde{\sigma}^2_{r,q,t}$ such that question q is studied is $\bar{\sigma}^2_{r,q,t}$. $\bar{\sigma}^2_{r,q^*,t}$ satisfies $u(V_{q'}, \sigma^2_{q',t-1}, \tilde{\sigma}^2_{r,q',t}) = u(V_{q^*}, \sigma^2_{q^*,t-1}, \bar{\sigma}^2_{r,q^*,t})$ for the best alternative question q', as perceived by researcher r, given by $q'(r,q,t) = \arg\max_q u(V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t})_{q \in Q_t \setminus q^*}$. Increasing the number of questions at time t from k_t to k'_t is equivalent to the set of questions increasing from Q_t to Q'_t, where $Q_t \subset Q'_t$. The max operator is weakly increasing as arguments are added, so letting $q''(r,q,t) = \arg\max_q u(V_q, \sigma^2_{q,t-1}, \tilde{\sigma}^2_{r,q,t})_{q \in Q'_t \setminus q^*}$, we have

$$u(V_{q'}, \sigma^2_{q',t-1}, \tilde{\sigma}^2_{r,q',t}) \le u(V_{q''}, \sigma^2_{q'',t-1}, \tilde{\sigma}^2_{r,q'',t}).$$

Therefore, $\bar{\sigma}^2_{r,q,t}(Q'_t) \le \bar{\sigma}^2_{r,q,t}(Q_t)$, where other arguments of $\bar{\sigma}^2_{r,q,t}$ are suppressed for clarity. □

References

Aumann R, Hart S (2003) Long cheap talk. *Econometrica* 71(6):1619–1660.

Benoit J-P, Dubra J (2011) Apparent overconfidence. *Econometrica* 79(5):1591–1625.

Camerer CF, Johnson EJ (1991) The process-performance paradox in expert judgement: How can experts know so much and predict so badly? Ericsson KA, Smith J, eds. *Toward and General Theory of Expertise*: *Prospects and Limits* (Cambridge University Press, Cambridge, UK), 195–217.

Che Y-K, Kartik N (2009) Opinions as incentives. *J. Political Econom.* 117(5):815–860.

Compte O, Postlewaite A (2004) Confidence enhanced performance. *Amer. Econom. Rev.* 94(5):1536–1557.

Crawford V, Sobel J (1982) Strategic information transmission. *Econometrica* 50(6):1431–1451.

Dawes RM, Corrigan B (1974) Linear models in decision making. *Psych. Bull.* 81(2):95–106.

Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial judgment. *Science* 243(4899:1668–1674.

Goldstein DG, Gigerenzer G (2002) Models of ecological rationality: The recognition heuristic. *Psych. Rev.* 109(1):75–90.

Krishna V, Morgan J (2001) A model of expertise. *Quart. J. Econom.* 116(2):747–775.

Krishna V, Morgan J (2004) The art of conversation: Eliciting information from experts through multi-stage communication. *J. Econom. Theory* 117(2):147–179.

Morris S (1995) The common prior assumption in economic theory. *Econom. Philos.* 11(2):227–253.

Santos-Pinto L, Sobel J (2005) A model of positive self-image in subjective assessments. *Amer. Econom. Rev.* 95(5):1386–1402.

Sethi R, Yildiz M (2012) Public disagreement. *Amer. Econom. J.: Microeconom.* 4(3):57–95.

Tetlock P (2006) *Expert Political Judgement*: *How Good Is It? How Can We Know?* (Princeton University Press, Princeton, NJ).

Van den Steen E (2004) Rational overoptimism (and other biases). *Amer. Econom. Rev.* 94(4):1141–1151.

Van den Steen E (2010a) Interpersonal authority in a theory of the firm. *Amer. Econom. Rev.* 100(1):466–490.

Van den Steen E (2010b) On the origin of shared beliefs (and corporate culture). *RAND J. Econom.* 41(4):617–648.

Van Wesep ED (2015) The shape of cooperative communication. Working paper, University of Colorado Boulder, Boulder.

Zabojnik Y (2004) A model of rational bias in self-assessments. *Econom. J.* 23(2):259–282.