The Quality of Expertise

Edward D Van Wesep

May 7, 2013

1Contact: The University of North Carolina Kenan-Flagler Business School, Finance Area, Campus Box 3490 McColl Building, Chapel Hill, NC 27599-3490. Please address email correspondence to vanwesep@unc.edu. The author would like to thank David Dicks, Chris Parsons, Diego Garcia, Andrew Ang, Naveen Khanna, Nina Baranchuk, and seminar participants at Wharton, Michigan State, University of Texas at Dallas, Vanderbilt, and University of British Columbia for helpful advice. The many errors are mine.
Abstract

Policy-makers and managers often turn to experts when in need of information: because they are more informed than others of the content and quality of current and past research, they should provide the best advice. I show, however, that we should expect experts to be systematically biased, potentially to the point that they are less reliable sources of information than non-experts. This is because the decision to research a question implies a belief that research will be fruitful. If priors about the impact of current work are correct, on average, then those who select into researching a question are optimistic about the quality of current work. In areas that are new, or feature new research technologies (e.g., data sources, technical methods, or paradigms), the selection problem is less important than the benefit of greater knowledge: experts will indeed be experts. In areas that are old and lack new research technologies, there will be significant bias. Furthermore, consistent with a large body of empirical research, this selection problem implies that experts who express greater confidence in their beliefs will be, on average, less accurate. This paper provides many empirical implications for expert accuracy, as well as mechanism design implications for hiring, task assignment, and referee assignment.

JEL Classifications: D80, J44

Keywords: Expertise, Learning, Prediction
Policy-makers and managers often need information in order to make a decision, and turn to “experts” for that information. Experts have a deep knowledge of the theory and evidence concerning their subject of expertise, and 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 to the standard: do unbiased experts even exist 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 policy-makers trust that the expert’s information is correct? I show that experts are, on average, biased and may 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 recent empirical work (e.g., Tetlock, 2006). I provide parameters that are, in principle at least, observable to the 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 manuscript, I focus on what is, for practical purposes, academic 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, like medicine, economics, or finance. While the model will be precise in defining what constitutes a question or a field, in practice it may be more difficult to draw a distinction. Is “the downfall of the Roman Empire” a field, in which “what caused it?” is a question? Or is the question “What caused the downfall of the Roman Empire?” while the field is Roman History? I will leave
the interpretation to the reader.

The model works as follows: in each period, researchers enter fields (e.g., start PhDs) and decide which questions to research (e.g., decide upon 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 private beliefs of the likely usefulness of future work on each question. These beliefs are, in expectation, correct. Each researcher ultimately chooses a question that she considers important, and for which she believes her work will be productive. After choosing a question, the researcher studies the question and updates her beliefs, taking into account past and current work. She does not observe current research on other questions, and therefore does not update her beliefs about the answers to those questions. 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 and non-experts alike, where a non-expert is defined to be a researcher in the field who has not seen new work (e.g., since graduate school) on the specific question of interest. Note that a non-expert is not a layman – she is familiar with past research on a question, but her knowledge is not up-to-date.

The fundamental trade-off for the policy-maker is that, while experts are more knowledgeable about current research, the very fact that they have chosen to study a particular question suggests that they may fundamentally disagree with others – who have chosen different questions – about the likely productivity of research on that question. At each step in the analysis, I make assumptions that are as favorable as possible for the conclusion that experts provide better advice than non-experts. I assume that: (i) only experts can observe current research on a topic, (ii) their utilities depend only upon their ultimate contributions to knowledge, and (iii) they want to convey the most accurate information possible to a querying policy-maker. The key assumption driving the results is that ex-
perts do not understand the selection problem. That is, when forming an opinion about the true answer to the question they study, they assume that their beliefs of the quality of current research are unbiased. In truth, while their beliefs are unbiased \textit{ex ante}, they are not unbiased \textit{conditional on having chosen a particular question}. It would perhaps be reasonable to allow for more significant bias among researchers; after all, the psychological evidence appears to confirm that many experts appear to suffer from decidedly first-order biases, such as overconfidence and narcissism. I show, however, that first-order biases are unnecessary to cast doubt on expertise.\footnote{The assumption that researchers use their priors, and do not adjust their beliefs to account for selection into an area of expertise is, for two reasons, weaker than it may appear. First, there is no evidence that one could offer a researcher that her prior is incorrect. That is, within the model, there are no past or present data that could refute the prior. Second, even if the theoretical argument in this paper were sufficient to convince her to adjust her beliefs, there is no alternative belief that is superior to the one I assume. She would know that she should put less weight on current work, but no specific alternative weight would be superior.}

The model yields several appealing results. 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 studied historically. 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 such 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.
The second set of results concerns how properties of a question relate to the likely accuracy of expert and non-expert beliefs. Experts put too much weight on current work, on average, but some questions have a more severe problem than others. Specifically, questions that are less important, more heavily studied historically, less heavily studied currently, 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, non-expert knowledge is always increasing. However, as knowledge about a question 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 non-expert opinion when a question is new: there is little published work for non-experts to observe, and little selection into becoming an expert. Expert opinion is less better 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 non-expert opinion.

The third set of results concerns how the properties of the field relate to the likely accuracy of expert and non-expert 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 relates to the quality of their opinions. Because the selection into expertise is more severe over time, older researchers are generally less biased than younger. This means that, for
the most part, it is better to solicit the views of older researchers about current work. The exception is when significant new research technologies become available concerning a question. In this case, work on the 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 studying 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).

I also show 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 upon 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 also prefer questions where past work was more precise, because 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 current work. Researchers that are pessimistic, in the sense that they believe current work is worse than it
actually is, will report lower confidence in their beliefs than researchers with correct priors. Because they underweight current work, 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 current 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, but because pessimistic researchers will choose alternative areas of expertise, the empirical correlation among experts will be negative. This relationship has been observed in dozens of studies, and this model provides an endogenous basis for it.

The sixth set of results concern mechanism design applications: (i) who should write tenure letters? (ii) when hiring a worker (e.g., as a term-structure modeler at an investment bank), should one hire a more intrinsically motivated expert (e.g., a person with a Ph.D. in finance) or a more financially motivated expert (e.g., a physicist)? (iii) when assigning workers to tasks, what discretion should they have over the choice of task?

This model relates to literatures in economics and psychology. Much of the early work in optimal learning assumes that people are Bayesian, and centers upon one of two questions. First, how should one trade off experimentation, which improves one’s knowledge, and using that knowledge productively? For example, the “multi-armed bandit problem” (Robbins, 1952) assumes several gambles with unknown payoffs, and a gambler that must choose which gamble to take in each period. There is value to experimenting, in that the gambler learns about the payoff distribution from each gamble, but there is also value to choosing gambles that have paid off well in the past. Second, how should one incorporate information

---

2 The relationship with no selection would be non-linear, but the linear trend would be close to zero.
provided by others into one’s own beliefs, and how does this rule affect the opportunity for information transmission in different settings? This so-called “cheap talk” literature began with Crawford and Sobel (1982) and has become quite extensive (see, e.g., Krishna and Morgan, 2001, 2004, and Aumann and Hart, 2003). It is typically assumed that the one providing the information has different preferences from the one receiving the information, providing an incentive to muddle the information and making the inference problem difficult. In this paper, I assume that the one providing the information genuinely wishes to be accurate, but ask when we should expect her information to be accurate. This is one step primitive to the standard question of how to elicit and interpret her information.

In the last decade, this literature has flourished in a variety of ways. Aragones et al (2005) separate two mechanisms of learning – collecting and analyzing information – and show that sorting through information is considerably more difficult than one might assume. This provides a theoretical basis for the widely observed phenomenon that theories that are “obvious” once pointed out can nonetheless be insightful when first presented. Epstein and Schneider (2007) study the problem of learning under ambiguity, offering an alternative to the Bayesian world in which the “truth” does not become perfectly understood over time.

Many equilibrium concepts in economics rely upon the game being common knowledge, but that knowledge must often be learned. It is not sufficient for me to learn the game – other players must learn as well. Cripps et al (2008) provide a simple example in which the size of the parameter space defining the game determines whether common learning occurs.

In somewhat more light-hearted work, Fudenberg and Levine (2006) ask why superstitions persist even in a world that many believe to be rational. They show that superstitions

---

3 They offer the example of the roll-aboard suitcase. We knew suitcases were heavy. We knew that it was smart to put wheels on heavy things to move them. But it took centuries for people to put wheels on a suitcase.
may survive so long as they rely upon off-equilibrium beliefs that are two steps off the equilibrium path. This work provides explanations for currently popular superstitions as well as a framework for the design of religion.

Second, there is a growing empirical literature in psychology concerning the specific question of this paper: are experts even expert and, if so, when? Repeated studies have found examples in which expert opinion can be less accurate than non-expert 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).\textsuperscript{4} There is significant work showing that information often reduces the precision of predictions.\textsuperscript{5} 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 world-view 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, to some extent, 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 due to 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

\textsuperscript{4}For example, Dawes, Faust and Meehl (1989) show that doctors are less accurate in predicting patient outcomes than a simple statistical model. Dawes and Corrigan (1974) write: “the statistical 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.”

\textsuperscript{5}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.
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 more is more
theoretical. Further, I seek endogenous, rather than exogenous, explanations for the failure
of expert predictions.

The rest of the paper is as follows. Section 1 provides the basic model and results
concerning the decision to study a question and the implications of that decision on expert
bias. Section 2 presents several examples that allow us to evaluation the evolution of
knowledge and accuracy over time. These examples also provide illustrations of the general
results of the preceding section. Section 3 offers some mechanism design applications, and
Section 4 concludes. Proofs can be found in the appendix.

1 The model

This model concerns a single field, with an exogenously given set of researchers choosing
among an exogenously set number of questions is each period. At time $t \geq 1$, the set of
questions available for study is $Q_t$, where $|Q_t| = k_t$. Once a question has entered the field,
it always remains, so $Q_{t-1} \subseteq Q_t$. 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
$R_{q,t}$, and let $|R_{q,t}| = n_{q,t} \geq 0$, where $\sum_{q=1}^{k_t} n_{q,t} = N_t$. Each question $q$ is defined by the
number $q$, which represents “the truth”. Researchers’ common prior about $\mu_q$ is that it is
distributed normally with mean zero and variance $\sigma_{q,0}^2$, and the goal of research is to learn
$\mu_q$.

Researcher $r$ researching question $q$ at time $t$ receives a signal $\hat{\mu}_{r,q,t} = \mu_q + z_{r,q,t}$, where
the standard deviation of $z_{r,q,t}$ is $\sigma_{r,q,t}$. Researchers report their signals publicly.\footnote{A public report of the signal $\hat{\mu}_{r,q,t}$ may appear in practice as a journal article, book, report, or speech.} Past
reports are common knowledge, as are the precisions of those reports. Current reports about question \( q \) are only observable to researchers that also study question \( q \), and the precisions of those reports are unknown.

Let researcher \( r \)'s prior of the variance of her signal of \( \mu_q \) at time \( t \) be \( \hat{\sigma}_{r,q,t}^2 \sim G_{r,q,t}(\bullet) \) where \( G \) is increasing and differentiable over the entire support \([0, \bar{G}]\), with \( G_{r,q,t}(0) = 0 \) and \( G_{r,q,t}(\bar{G}) = 1 \). \( G \) is unknown to the researcher, but researchers are, on average, unbiased in their beliefs: \( E_{G_{r,q,t}}(\hat{\sigma}_{r,q,t}^2) = \sigma_{r,q,t}^2 \).

Let the precision of research be denoted \( \gamma_{r,q,t} = 1/\sigma_{r,q,t}^2 \), and let researcher \( r \)'s belief of that precision be denoted \( \gamma_{r,q,t} = 1/\hat{\sigma}_{r,q,t}^2 \). Given that past reports and their precisions are common knowledge, at each time \( t \) there is a common prior for each question \( q \). The precision for that prior is simply the sum of the precisions of the signals that comprise it.

\[
\tilde{\gamma}_{q,t-1} = \frac{1}{\sigma_{q,0}^2} + \sum_{s=t+1}^{t-1} \sum_{r \in R_{q,s}} \gamma_{r,q,s},
\]

where \( \tau_q \) is the first period that question \( q \) enters the field.\(^7\)

I assume that the researcher cares about how much she has advanced knowledge: \( u_{r,q,t} = u(V_q, \tilde{\gamma}_{q,t-1}, \hat{\gamma}_{r,q,t}) \), where \( u \) is increasing in the importance of the question, \( V_q \), decreasing in the precision of her prior, \( \tilde{\gamma}_{q,t-1} \), and increasing in the precision of the signal that she expects to get, \( \hat{\gamma}_{r,q,t} \). There are a variety of functional forms that satisfy these requirements, all of which yield qualitatively identical results to those I present below. For simplicity, I assume that her utility equals the difference between the variance of the prior and the variance of the posterior if she researches the question, scaled by the value of the question, under the assumption that no other researchers study the question:

\[
u_{r,q,t} = V_q \times \left[ (\tilde{\gamma}_{q,t-1}^{-1} - (\hat{\gamma}_{q,t-1} + \hat{\gamma}_{r,q,t})^{-1}) \right].
\]

\(^7\)That is, \( q \in Q_{\tau_q} \) but \( q \notin Q_{\tau_{q-1}} \).
Note that this utility function ignores the research that other experts do on a question. The researcher’s utility depends only upon how her research improves knowledge about a question relative to where it was when she began her work. It does not depend upon 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.

Note also that I could have also defined her utility to be the ratio of her perceived signal precision to the precision of the current estimate of $\mu_q$, which would yield $u_{r,q,t} = V_q \times \frac{\gamma_{r,q,t}}{\gamma_{q,t-1}}$. All comparative statics in the paper would be identical with this alternative choice for utility.

A researcher’s life is as follows. She is born, and observes a vector $\{\tilde{\sigma}_{r,1,t}, \tilde{\sigma}_{r,2,t}, ..., \tilde{\sigma}_{r,k,t}\}$, where $\tilde{\sigma}_{r,q,t} \sim G_{r,q,t}(\bullet)$. She then chooses a question $q \in Q_t$, and observes a signal $\tilde{\mu}_{r,q,t} = \mu_q + z_{r,q,t}$, where the standard deviation of $z_{r,q,t}$ is $\sigma_{r,q,t}$. She then reports $\tilde{\mu}_{r,q,t}$ publicly and honestly. Researchers differ, and choose different questions, because $\tilde{\sigma}_{r,q,t} \neq \tilde{\sigma}_{r',q,t}$ for $r \neq r'$.

After observing all reports associated with her chosen question, $\tilde{\mu}_{r',q,t}, r' \in R_{q,t}$, she forms a posterior about $\mu_q$. In forming a posterior, she must choose a weight to place upon current work, whose quality is not contemporaneously observable. I assume that she assigns $\tilde{\sigma}_{r,q,t}^2$ as the variance of current work on question $q$, an assumption I discuss below.

Because researcher $r'$s prior for the standard deviation of $z_{r,q,t}$ is $\tilde{\sigma}_{r,q,t}$, her updated estimate of $\mu_q$ after observing $\tilde{\mu}_{r,q,t}$ will put the wrong weight on her current signal relative to past signals. If $\tilde{\sigma}_{r,q,t} > \sigma_{r,q,t}$, then she will put too little weight on current work, and if $\tilde{\sigma}_{r,q,t} < \sigma_{r,q,t}$, she will put too much weight on current work. Ex ante, by assump-
tion, \( E(\hat{\sigma}_{r,q,t}) = \sigma_{r,q,t} \). She will also assign a different weight to current work than other researchers studying the same question, so experts will disagree.

Before solving for equilibrium decisions and outcomes in the model, I discuss two assumptions. First, questions are exogenously generated over time. While 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, the quality of current research is not contemporaneously observable. It takes one period for the quality of work to become known. This assumption is critical for the results: if experts observe the quality of current work in forming posteriors, there is no scope for disagreement.

Fourth, researchers assume that their priors for the quality of current work are correct. \( \text{Ex ante} \), they are correct on average, but after selecting into a question, they may be incorrect on average. Note that this assumption is fairly weak: there is no evidence that they could observe that their priors are incorrect, only a theoretical argument. Moreover, even if they could be persuaded by the theoretical argument, there is no way for them to know how much to adjust their beliefs about \( \sigma_{r,q,t} \) upward, as \( G_{r,q,t} \) is unknown. Any adjustment would be arbitrary and incorrect, with probability one. A perfectly rational and informed expert could say, at best, “I know that I am likely to be biased in favor of current work, but cannot provide any estimate of how much.”

1.1 The decision to research a question \( q \)

The first period in which question \( q \) comes into existence is \( \tau_q \). The number of people working on question \( q \) in periods \( \tau_q \) to \( t - 1 \) is \( \{n_{q,\tau_q}, n_{q,\tau_q+1}, ..., n_{q,t-1}\} \). The prior for all
researchers at the start of time $t$ for question $q$ can thus be written

$$
\bar{\mu}_{q,t-1} = \sum_{s=\tau_q}^{t-1} \sum_{m \in R_{q,s}} \left( \frac{\gamma_{m,q,s}}{\gamma_{q,t-1}} \right) \hat{\mu}_{m,q,s}.
$$

In period $s$, there are $n_{q,s}$ researchers, each of whom draws signal $\hat{\mu}_{r,q,s} = \mu_q + z_{r,q,s} \sigma_{r,q,s}$. These signals are reported and observed by experts. At the end of period $t$, expert $r$ therefore has posterior

$$
\bar{\mu}_{r,q,t} = \sum_{s=\tau_q}^{t-1} \sum_{m \in R_{r,s}} \left( \frac{\gamma_{m,q,s}}{\gamma_{q,t-1} + n_{q,t} \tilde{\gamma}_{r,q,t}} \right) \hat{\mu}_{m,q,s} + \sum_{m \in R_{r,t}} \left( \frac{\tilde{\gamma}_{r,q,t}}{\gamma_{q,t-1} + n_{q,t} \tilde{\gamma}_{r,q,t}} \right) \hat{\mu}_{m,q,t},
$$

where I have noted that the posterior is unique to expert $r$. The posterior precision is believed by expert $r$ to be $\tilde{\gamma}_{q,t-1} + n_{q,t} \tilde{\gamma}_{r,q,t}$. Non-experts do not observe contemporaneous research, and therefore have a posterior identical to the prior, and do not adjust the precision.

**Proposition 1** A researcher is weakly more likely, ceterus paribus, to choose to study question $q$ if:

1. She believes her signal will be more precise: $\tilde{\sigma}_{r,q,t}$ 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: $\sigma_{r,q,s}$ is higher, $s < t$, for any $r,$

5. The question is more important: $V_q$ is higher.
Corollary 1 Researchers are more likely to study questions that experience a rapid increase in the (i) quality of available data, (ii) quantity of available data, (iii) available methods of analysis.

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.

1.2 The precision of expert and non-expert opinion

The probability that researcher \( r \) chooses question \( q \) is

\[
pr \left( \tilde{\sigma}_{r,q,t}^2 < (V_q/V_{q'})\tilde{\gamma}_{t-1}^{-2} \left[ \tilde{\gamma}_{q,t-1}^{-2}\tilde{\sigma}_{r,q',t}^2 + \tilde{\gamma}_{q',t-1} \right] - \tilde{\gamma}_{q,t-1}^{-1} \right) \text{ for all } q' \in Q_t \setminus q.
\]

This is given by the first-order statistic:

\[
\int_0^{\infty} \left( g_{r,q,t}(x) \times \prod_{q' \in Q_t \setminus q} \left[ 1 - G_{r,q',t} \left( (V_q/V_{q'}) \tilde{\gamma}_{q,t-1}^{-2} \left( \tilde{\gamma}_{q,t-1}^{-1}x + \tilde{\gamma}_{q,t-1} - \tilde{\gamma}_{q',t-1}^{-1} \right) \right) \right] \right) dx,
\]

where \( g_{r,q,t} \) is the (strictly positive) derivative of \( G_{r,q,t} \). The probability that question \( q \) is chosen equals, for a given \( \tilde{\sigma}_{r,q,t} \), the probability that all other questions have values of \( \tilde{\sigma}_{r,q',t} \) that are high enough to make question \( q \) the most attractive. I then integrate over all possible values of \( \tilde{\sigma}_{r,q,t} \), multiplying by the density at \( \tilde{\sigma}_{r,q,t} \).

I cannot calculate the expected value \( E(\tilde{\sigma}_{r,q,t}) \) analytically, but I can perform comparative statics with it relatively easily. For any set of values \( V_q \) and \( \{V_{q'}, \tilde{\gamma}_{q',t-1}\}_{q' \in Q_t \setminus q} \), which are known at time \( t \), and for any draws of \( \tilde{\sigma}_{r,q',t}, q' \in Q_t \setminus q \), there exists a threshold value \( \tilde{\sigma}_{r,q,t} \geq 0 \) such that researcher \( r \) studies question \( q \) if and only if \( \tilde{\sigma}_{r,q,t} \in [0, \tilde{\sigma}_{r,q,t}^r] \). Therefore, if I find that a change in some parameter increases (decreases) the threshold \( \tilde{\sigma}_{r,q,t} \), then the expectation of \( \tilde{\sigma}_{r,q,t} \) must increase (decrease) as well.
Proposition 2  If a question $q$ is worth studying for some vector of beliefs $(\hat{\tau}_{r,q,t})_{q' \in Q_t}$, then as the level of knowledge about $q$ increases, expected expert bias weakly increases. That is, if $\hat{\tau}_{r,q,t} > 0$, then $\frac{d}{d\hat{\tau}_{r,q,t-1}} E(\hat{\tau}_{r,q,t} \mid q \text{ chosen}) < 0$.

Proposition 2 establishes that as researchers learn more about a question, the bias of the researchers that 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 2  Expected expert bias is higher, ceterus paribus, if:

1. The question is older: $t$ is higher for any given $\tau$,

2. More researchers have worked on the question historically: $n_{q,s}$ is higher for any $s \in \{\tau_q, \tau_q + 1, ..., t - 1\}$,

3. Historic signals were more precise: $\sigma_{r,q,s}$ is lower, so $\gamma_{r,q,s}$ is higher, for any $s \in \{\tau_q, \tau_q + 1, ..., t - 1\}$,

4. The question is less important: $V_q$ is lower.

Not only are experts biased, but they may become so biased that they are less reliable than non-experts. Indeed, this may be quite common in practice, as many questions are “settled” in the minds of non-experts, who have moved elsewhere for research questions. Questions for which expert opinion is less accurate than non-expert opinion feature few researchers, while those for which experts are more accurate feature many researchers. Therefore, while 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 non-experts.
Corollary 3  *Expert opinion can be less precise than non-expert opinion.*

Thus far I have focused upon the bias generated by properties of the question at the start of period $t$, but the likely productivity of future research is important as well. Holding constant all values $V_q$ and $\{\hat{V}_q', t_{-1}\}_{q' \in Q_t \setminus q}$, and draws of $\hat{\sigma}_{r,q', t}, q' \in Q_t \setminus q$, I can evaluate how the level of bias changes as the productivity of current research, $\sigma_{r,q,t}$, changes. To do this, I write $\hat{\sigma}_{r,q,t} = K\sigma_{r,q,t}$ so that $E(K) = 1$. The distribution of $\hat{\sigma}_{r,q,t}$ is $G_{r,q,t}(\cdot)$, so the distribution of $K$ can be denoted $H_{r,q,t}(x/\sigma_{r,q,t}) \equiv G_{r,q,t}(x)$, which has support $[0, \bar{G}/\sigma_{r,q,t}]$ 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 Section 3 as it relates to the age of researchers.

### 1.3 The relationship between the field and the question

I have focused thus far upon 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 upon 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 a 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. While the question is not completely understood (nor should it ever be, according to the model), it is highly studied and future work likely has a low likelihood of impact. Researchers who choose to study it now are likely very biased in favor of current work, 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: $E(\hat{\sigma}_{r,q,t})$ is weakly decreasing in $k_t$, ceterus paribus.

1.4 Expert confidence and accuracy

I have focused thus far upon 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 that 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 $\tilde{\sigma}_{r,q,t}^2$ are lower. As $\tilde{\sigma}_{r,q,t}^2$ approaches $\sigma_{r,q,t}^2$ 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 $\tilde{\sigma}_{r,q,t}^2$ passes $\sigma_{r,q,t}^2$ 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 $\tilde{\sigma}_{r,q,t}^2 < \sigma_{r,q,t}^2$. Therefore, the empirical domain of $\tilde{\sigma}_{r,q,t}^2$ will tend to be that for which there is a negative relationship between expert confidence and accuracy. I provide a representative example in Section 2.4.

I now turn to some numerical examples to elucidate the preceding results.

## 2 Numerical examples

The selection of researchers into questions drives bias in a relatively straightforward way. In order to clarify the preceding results, and provide some additional intuition for the evolution of expertise over time, I now present some numerical examples. The examples involve deterministic priors in order to allow for analytical results.
First, I assume a non-random distribution of potential experts arriving each period and show, for a variety of parameters, how the quality of expert and non-expert opinion evolves over time. For simplicity, I assume one question and a fixed outside option $\bar{u}$, which is appropriate for a large, growing field.\footnote{A constant outside option requires that there are always new questions arising. In a static field, the utility from studying questions declines over time, as all questions age together.}

Second, I shock the quality of data and map impulse-response functions to see how changes in the research technology affect the quality and quantity of research on the question.

Third, I assume two questions in the field, but remove the outside option. This analysis will allow me to study how the quality of one question in a field affects the bias of researchers in another question in the same field. It will also allow us to see how the preceding results change when the field is static, rather than growing. I show that, over time, the field reaches a steady-state in which all experts are less reliable than non-experts. This is because the importance of incremental information must tend to zero over time, while the average bias does not. In a static field, experts simply cannot be relied upon.

I also show that the quality of expert opinion within a field can oscillate between two questions, as researchers flock from one to the other, appearing to chase trends. In the long run, more important questions will be more heavily researched and better understood, but in the short run, when the impact of new research is relatively large, researchers may oscillate between choosing questions that are important, and those that are relatively less studied.

### 2.1 The quality of opinion over time in a large, growing field

Let there be one question $q$, with $\sigma_{r,q,t} = 1$ for all $t > 0$ and $\sigma_{q,0}^2 = 0.1$, and let $\tau_q = 1$.

Let there be 39 potential researchers every period, $r \in \{1, 2, \ldots, 39\}$, and let $\sigma_{r,q,t}^2 = r / 20$,
so $\hat{\sigma}_{r,q,t}^2$ ranges from 0.05 to 1.95, in increments of 0.05. Clearly $E(\hat{\sigma}_{r,q,t}^2) = \sigma_{r,q,t}^2 = 1$. I represent the existence of other questions as a fixed outside option, $\bar{\pi}$. The fact that the quality of the outside option is unchanging is consistent with a large field that is growing over time, so that good opportunities for research consistently arise.

The upper graphs in Figure 1 plot the precision of expert and non-expert estimates of $\mu_q$ over time, for values of the question relative to the outside option of $V_q/\bar{\pi} = 10,000$ and $V_q/\bar{\pi} = 100,000$. The lower graphs plot the number of researchers, out of 39, that choose to study $q$ over the outside option.

When $V_q/\bar{\pi} = 100,000$, shown in the left-hand graphs, the question is very important relative to the outside option, driving all 39 potential experts to study $q$ in early periods. Non-experts, if they were to exist, would have access only to the prior, which has a precision of 10, while experts have access to all 39 expert reports as well, and are on average unbiased. The median expert opinion will be unbiased and have precision of 49, nearly five times the precision of a non-expert opinion. As the number of periods increases, non-experts gain access to this information and gain precision. Once some potential experts begin to choose the outside option, in period 7, the increasing bias and information over time offset, leading the precision of expert opinion to level off. The same is not true for non-experts, who become better informed and more reliable.

When $V_q/\bar{\pi} = 10,000$, shown in the right-hand graphs, all potential experts choose to study the question in period 1. Beginning in period 2, many or most potential experts choose the outside option, reducing the rate of knowledge growth and reducing the quality of expert opinion. It is interesting to note that expert opinion does not, under these parameters, become more precise over time. The increase in knowledge is met with an approximately corresponding increase in bias in favor of current work, yielding approximately constant precision over time.
These examples highlight that experts will generally be less expert when questions are old and well studied.

2.2 The quality of opinion and shocks to the research technology

It should be intuitively clear that when research becomes more effective, potential researchers are more willing to study a question. An increase in efficacy could be through improved technology (e.g., the development of computers), improved theory (e.g., the marginal revolution or capital asset pricing model), improved data availability (e.g., new proprietary sources of data), or increased funding (e.g., additional grants from the NSF). Each of these changes can be modeled in this framework with a reduction in $\sigma_{r,q,t}$ at time $t$: the impact of a fixed effort toward research yields an improved signal of the truth. In Figure 2, I plot our four graphs from Figure 1, with a small change. At time $t = 7$, $\sigma_{r,q,t}^2$ drops from 1 to 0.1: research is ten times as effective.

In the upper-left graph, representing the case where the question is very important, the drop in $\sigma$ does not affect the level of bias among experts because all potential experts choose to study $q$ both before and after the drop in $\sigma$. Despite this, the quality of research produced increases significantly because of the drop in $\sigma$, meaning that experts become much better informed than non-experts, who lack access to current research. In later periods, the drop in the number of researchers studying $q$ more than offsets the improved productivity of research, and expert opinion stagnates, while non-expert opinion continues to improve.

In the upper-right graph, representing the case where the question is not very important, expert and non-expert opinion sharply diverge when $\sigma$ falls, as many researchers decide to study $q$, both increasing the knowledge available only to experts, and reducing their average bias. Over the subsequent periods, non-experts gain access to this new information, sharply
increasing the precision of their opinions. Meanwhile, the opportunity for novel research falls, leading to even greater selection against unbiased experts than prior to the drop in \( \sigma \).

These examples highlight that shocks to the research technology are accompanied by shocks to the quality of expert opinion. The quality of the technology itself is less relevant than changes to that technology. When looking for external indicators of whether experts are to be believed, look for changes in the processing and creation of knowledge.

### 2.3 The evolution of expert opinion in a small, static field

Figure 3 plots the precision of expert and non-expert opinion for two questions when there is no outside option. These two questions comprise the entire field, which is therefore static. These figures help clarify how changes to one question in a field affect properties of opinion regarding the other question.

There are five potential values of \( \hat{\sigma}_{r,q,t} \) for each question \( q \in \{ Q1, Q2 \} \), and the spread between them is defined by \( Spread q \in [0, 1/2] \), a parameter I allow to vary:

\[
\hat{\sigma}_{r,q,t} \in \{ 1 - 2 \cdot Spread q, 1 - Spread q, 1 + Spread q, 1 + 2 \cdot Spread q \}.
\]

For simplicity, there is one researcher for each pair \( \{ \hat{\sigma}_{r,Q1,t}, \hat{\sigma}_{r,Q2,t} \} \), so there are 25 researchers in the field. The upper figure in Figure 3 assumes that \( Spread Q1 = Spread Q2 = 0.4 \), so \( \hat{\sigma}_{r,q,t} \) takes values of 0.2, 0.6, 1, 1.4, and 1.8 for each question.\(^9\)

---

\(^9\)The set of researcher priors each period can be represented in a table. Let the row variable be values of \( \hat{\sigma}_{r,Q1,t} \) and the column variable be values of \( \hat{\sigma}_{r,Q2,t} \). Then the set of researchers born each period are

\[
\begin{array}{cccc|cccc}
0.2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0.6 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1.4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1.8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

and

\[
\begin{array}{cccc|cccc}
0.8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0.9 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1.1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1.2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

in the upper and lower panels of Figure 3, respectively.
Initially, the experts for both questions have more precise opinions than non-experts, but as the marginal importance of current work decreases, the bias overcomes the benefit of knowledge of current work.

In the lower figure in Figure 3, the prior precisions are raised to 5 for both questions, and the spread for question Q1 is reduced to 0.1. Beliefs about $\sigma_{r,q,t}$ are less diffuse for question Q1, so the types who choose question Q1 are less biased. This does not affect the level of bias for researchers in question Q2, but clearly shows that experts can be superior sources of information, even if biased, for an extended period. In deciding whether to query an expert or non-expert, a policy-maker or manager must consider the dispersion of beliefs about $\sigma_{q,t}$.

Figure 4 plots the precision of expert and non-expert opinion for the two questions over time, where question Q1 is twice as important as question Q2. Prior precisions and spreads are the same for both questions, at 10 and 0.3, respectively. Because question Q1 is more important, most researchers choose question Q1 in period 1, with only the most biased in favor of Q2 choosing to study Q2. In period 2, however, question Q1 has been relatively heavily studied, causing researchers to flock to question Q2, even though it is less important. In period 3, the questions are again equally well studied, and researchers return to question Q1. This means that the level of bias among experts oscillates as well, with high bias in periods when the question is less studied. Over time, as the value of marginal research decreases, a steady-state is reached in which the types choosing each question are constant and question Q1 receives more research than Q2.

2.4 Expert confidence and accuracy in a large, growing field

To evaluate the relationship between actual and self-assessed expert accuracy, I return to the assumptions in Section 2.1, representing a question in a large, growing field. Figure
5 plots the precision of expert reports on the vertical axis and the expert’s belief of her precision on the horizontal axis, assuming $V_q/i = 10,000$. In period 1, all potential researchers study the question. Increasing self-assessed precision is associated with more optimistic priors, i.e., lower draws of $\tilde{\sigma}_{r,q,t}^2$. For those researchers whose priors are pessimistic, $\tilde{\sigma}_{r,q,t}^2 > \sigma_{r,q,t}^2$, increasing perceived precision is associated with increased actual precision. As $\tilde{\sigma}_{r,q,t}^2$ falls, the weight on current work increases toward the correct level, and the researcher’s belief of her posterior precision is higher. For those researchers whose priors are optimistic, $\tilde{\sigma}_{r,q,t}^2 < \sigma_{r,q,t}^2$, increasing perceived precision is associated with decreased actual precision. These researchers already put too much weight on current work, and increasing that weight further is not helpful.

The linear regression line is plotted as well, showing that the observed relationship between self-assessed and actual accuracy is negative, even without a selection bias. This is due to a modeling assumption. I assumed that $E(\tilde{\sigma}_{r,q,t}^2) = \sigma_{r,q,t}^2$. In forming posteriors, a low draw of $\tilde{\sigma}_{r,q,t}^2$ has a disproportionate impact relative to a high draw of $\tilde{\sigma}_{r,q,t}^2$. Had I assumed that $E(\tilde{\gamma}_{r,q,t}) = \gamma_{r,q,t}$, this negative relationship would not be present without any selection bias.

In the right panel, the same scatter is presented for experts in period 5. Many potential experts choose the outside option, so only optimistic experts remain. Among this group, there is a clear negative relationship between self-assessed and actual precisions. Because there is, in practice, always some selection into any question, we should expect that this negative relationship will be the empirical norm.

10 An easy way to see this is to imagine a simple distribution of $G_{r,q,t}$, which assigns a 50% probability to a draw of $\tilde{\sigma}_{r,q,t}^2 = 0$ and a 50% probability to a draw of $\tilde{\sigma}_{r,q,t}^2 = 2\sigma_{r,q,t}^2$. The posterior of the former researcher will be formed entirely of current work, and the reported precision will be infinite. The posterior of the latter will put too little weight on current work, and the posterior precision will be finite. Even without selection bias, the negative relationship will be present.
3 Which experts should be consulted?

Thus far, there has been little discussion of which experts should be consulted and how their opinions should be weighed. This is because experts are functionally identical in the current model. In practice, however, there may be many experts for a given question, and policy-makers and managers may have limited ability to query them. It is therefore important to address precisely which experts should be consulted in a given case.

In this section, I split experts along two dimensions. First, experts may be older or younger. The prior results naturally extend to the bias of experts at different points in their careers. I will argue that older experts are typically more reliable than younger, with the exception that shocks to the research technology can make younger researchers temporarily more reliable. Older researchers that study, for example, severe recessions, were likely better sources of information than younger researchers before the recession of 2008 but, post-recession, younger researchers may be more reliable for some time. Indeed, this cohort of researchers is likely to be the most reliable source of expertise on business cycle macroeconomics for their entire careers.

Second, experts may be academic or professional, in the sense that they may choose research questions for intrinsic or extrinsic reasons. Professional experts may be less able to discern the truth, but also lack the bias that comes from selecting into an area of expertise. For older questions, like macroeconomic forecasting, professionals are likely to be more reliable than academics while for new questions, like the relationship between specific genes and cancers, academics may be superior.

This analysis also has implications for task assignment within a firm: allowing workers to choose areas of expertise may make them happier, induce effort, and reduce wage bills, but it also likely leads to bias in their beliefs. There are similar implications for hiring: should a hedge fund hire workers who have advanced degrees in finance or in physics?
The former have more knowledge of financial markets, but are likely biased about their functioning.

For the most part, the claims in this section follow immediately from the intuition in the model, so I do not prove them formally.

### 3.1 The effect of expert age on precision

To understand the effect of expert age, I must introduce some variation along this dimension. I allow researchers to live for $s$ periods, rather than one, but require that they maintain their initial question for their full careers. A more important assumption concerns how I assign expert beliefs when they are older. One option would be to allow a new, independent draw of $\tilde{\sigma}_{r,q,t}$. This would clearly imply that older experts would be unbiased, on average, and therefore clearly the best source of information. I do not believe this is a particularly reasonable assumption, given that biases are likely to persist, even if in weakened form, throughout a career.

I assume instead that the level of bias is fixed over time. Define $K_{r,q,t} = \tilde{\sigma}_{r,q,t}/\sigma_{r,q,t}$ to be the proportionate bias for researcher $r$ on question $q$ in the first period of the her career. I assume that $K_{r,q,t} = K_{r,q,t'}$ for $t \neq t'$, so her bias on all questions is constant. Allowing partial or full mean-reversion would strengthen the results.

By assumption, the only difference, on average, between researchers of different ages is the level of selection into a question when the researcher was young. I therefore abstract from other differences that may occur in practice, such as differences in awareness of the literature, research effort, intellectual flexibility, etc. Results below should therefore be seen as *ceterus paribus.*
3.1.1 Application: Tenure letters and the preference for “grey hair”

It is typical for universities to solicit views of faculty at other schools in reviewing a professor’s tenure case. The “letter-writers” are typically very senior, but it is not clear why this should be so. Less senior faculty often publish papers at equally high rates, suggesting a similar level of knowledge about a candidate’s contribution. Indeed, one could argue that older faculty are more likely to be academically disengaged, delegating work to research assistants and spending more time with their families.

This paper provides some support for the focus on “grey hair”:

Claim 1 In expanding fields, at a given time, older researchers have more precise opinions. In static fields, at any given time, the precision of opinions is un-correlated with age.

The intuition is straightforward, and follows immediately from Corollary 2: as a question ages, more is known about it. So long as there are new questions to draw researcher interest, those researchers who choose to study the older question are increasingly biased over time. Older researchers chose to study the question when it was less well understood, so are less biased on average.

If the field is static, however, no new questions arise: all questions age together so there is no worsening selection into any given question. When younger researchers start their careers, all questions are better understood than when their elders began as researchers, so the decision to choose one question over another does not involve increasing selection.

Because most fields are expanding, it is typical that older faculty may be more able to judge the quality of a candidate’s contribution because they are less biased about the importance of current work in the candidate’s area. Younger faculty suffer, potentially, from an inverse problem, in which the less impactful is current work, the more impactful they believe it to be! Older faculty, having chosen research questions long ago, are more
able to note the lack of interesting work done on those questions and can report this to a candidate’s school. However, provosts and deans should note that the “grey hair” preference in tenure letters should be reconsidered in slowly expanding or static fields like History, in which there are few new questions and little new data.

3.1.2 Application: New information and the preference for youth

There is an important counter-example, however, to the claim that older experts tend to be less biased.

Claim 2  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 $\sigma_q$ at the time they are choosing questions.

This drop makes current work impactful, and draws in researchers, even if they are not particularly biased. This is a rare case in which young researchers are more reliable than old. This cohort stays with the question over time, and remain less biased than younger cohorts who start their careers when more is known about the question, and incremental research is therefore less valuable. An especially stark example is shown in the right-hand panel of Figure 2.

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 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.

3.2 The decision to become an expert and precision

Expert opinion can be less precise than non-expert opinion because selection into studying a question yields researchers who are biased in favor of current 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 that I could have assumed.

Consider the following change to the model. After a researcher chooses a question, she forms a posterior $P_r^{P_r}$. A noisy signal of the truth is publicly observed, $M_{q,t} = \mu_q + m_{q,t}$, where $m_{q,t}$ is an independent mean zero random variable with precision $\gamma_{q,t}$. The researcher receives utility based upon making accurate predictions about important topics:

$$u = V_q - \left(\mu_{q,t}^{P_r} - M_{q,t}\right)^2.$$  

The researcher’s perceived expected utility would then be

$$u_{r,q,t} = E\left[ V_q - \left(\mu_{q,t}^{P_r} - M_{q,t}\right)^2 \mid \tilde{\sigma}_{r,q,t} \right].$$  

With these preferences, it should be clear which forces drive entry to a question.

**Proposition 5** An expert with preferences $u_{r,q,t} = E\left[ V_q - \left(\mu_{q,t}^{P_r} - M_{q,t}\right)^2 \mid \tilde{\sigma}_{r,q,t} \right]$ is weakly more likely, ceterus paribus, to study question $q$ if:

1. She believes her signal will be more precise: $\tilde{\sigma}_{r,q,t}$ is lower,

---

11The expectation is not the true expectation, but her belief of the expectation, given her belief of $\tilde{\sigma}_{r,q,t}$.
2. The prior signal precision is higher: $\tilde{\eta}_{q,t-1}$ is higher,

3. The question is more important: $V_q$ is higher.

While a researcher with these preferences will still choose questions for which she is more biased, she also will choose questions for which there is more prior knowledge, counteracting this bias. Indeed, depending upon parameters, experts may be more reliable than non-experts for all questions. This alternative utility function may be implementable in practice, and two examples are provided below.

3.2.1 Application: Academic vs. professional experts

For many questions, there are both academic and professional experts making claims about the answer to the question. Professionals study a question (e.g., form growth forecasts, profit forecasts, etc.) for money, and are rewarded for accuracy. Academics may have some extrinsic reward for accuracy, but are generally motivated by intrinsic rewards. Even if professionals are less aware of, or able to process, current and past research, this handicap may be less severe than the bias associated with research expertise.

Claim 3 Professionals may have more precise opinions than academics. The quality of professional expert opinion, relative to academic, is higher when a question is older and more researched.

Proposition 5 establishes that professional experts 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.
3.2.2 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. In the setting of this paper, I can consider whether a worker should be allowed to choose a question to research or have that question assigned.

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, while 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 upon the degree of relative intrinsic motivation, the fund may prefer either employee.

4 Conclusion

This paper is dedicated to the development and application of a simple insight: the selection into an area of expertise implies 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 is little new data, few new tools, etc., are subject to more severe selection bias, and 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: older experts, those who entered when new data and tools were becoming available, and those who are extrinsically motivated will often be more reliable. The result concerning golden cohorts should generalize to most any model, while the results that older and more extrinsically motivated experts are more precise should be seen as conditional on equal skill and ability. Finally, the 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.
1. Tenure letter writers: The results concerning age suggest a rationale for asking older researchers to write tenure letters. Assuming that letter writers should have the same area of expertise as the candidate, it is important to maximize the expertise of the person evaluating the candidate.

2. The refereeing process: Journals should generally require both expert (referee) and non-expert (editor) opinions. The expert judges the quality of work within the question (e.g., by observing a signal of the value of $\sigma_{r,q,t}$ associated with the report), and the non-expert judges the likely bias in the question more generally. Because experts disagree, it may be useful to ask multiple experts (referees). Because non-experts agree, one non-expert (editor) is sufficient.

3. Hiring: One might imagine that it is superior to hire intrinsically motivated agents, as the individual rationality constraint may be more easily met, the greater the utility the agent gets from working. Intrinsic motivation, however, is the source of selection bias that yields biased experts. The principal should determine whether the motivation existed prior to learning of past work on the question. If so, then the motivation reflects inherent interest and is beneficial for the employer. If the interest arose or was strengthened after the agent began to study the question, however, this may imply likely bias. An example would be a hedge fun hiring a Ph.D. in finance or physics. The former may be more knowledgeable and easier to train, but also chose a course of study that may suggest bias.

4. Task assignment: Similar to the hiring problem, a principal could choose how to assign employees to tasks. While employees likely prefer discretion in their area of expertise, random assignment may be superior to discretion, as it eliminates the source of bias once they become expert. For example, it is common for PhD students in the sciences
to become expert in whatever questions senior researchers in the department happen to ask. While there is some discretion regarding which lab or program to choose, it is limited. In the social sciences, the opportunity for self-directed study is greater. This suggests the potential for greater bias among researchers in the social sciences than the hard sciences.

The fact that research experts tend to be biased, importantly, 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.

5 References


6 Appendix – Proofs

**Proof of Proposition 1.** A researcher chooses question $q$ if and only if $u_{r,q,t} \geq u_{r,q',t}, \forall q' \in Q_t \setminus q$.

Therefore, anything that increases her utility from choosing a re-

\footnote{Note that the probability of $u_{r,q,t} = u_{r,q',t}$ for some $q \neq q'$ is zero.}
search question weakly increases the likelihood of her choosing that question. Utility can be written 
\[ u_{r,q,t} = V_q \times \left( \sigma_{q,t}^2 + \sum_{s=\tau_q}^{t-1} \sum_{r \in R_{q,s}} \gamma_{r,q,s} \right)^2, \]
which is decreasing in \( \sigma_{r,q,t}, n_{q,s} \) and \( \gamma_{r,q,s} \), for \( s \in \{ \tau_q, \tau_q + 1, \ldots, t - 1 \} \), and increasing in \( V_q \). Holding \( n_{q,s} \) and \( \gamma_{r,q,s} \) constant for each period \( s \), reducing the difference \( t - \tau_q \) increases \( u_{r,q,t} \).

\[ \square \]

**Proof of Proposition 2.** It is sufficient to show that, for every set of values \( \{ \tilde{q}_{q',t-1}, V_{q'} \}_{q' \in Q_q \setminus q} \) and \( V_q \), if \( \sigma_{r,q,t} > 0 \), then it is decreasing in \( \tilde{q}_{q,t-1} \). Let question \( q' \) be the best available alternative to question \( q \) for researcher \( r \):

\[ q'(r,q,t) = \arg \max_{m} \{ V_m / [\tilde{q}_{m,t-1}^2 \sigma_{r,m,t} + \tilde{q}_{m,t-1}] \}_{m \in Q_q \setminus q}. \]

Then \( \sigma_{r,q,t} = (V_q/V_{q'})\tilde{q}_{q,t-1} \left[ \tilde{q}_{q',t-1}^2 \sigma_{r,q',t} + \tilde{q}_{q',t-1} \right] - \tilde{q}_{q,t-1}^{-1} \). I take a derivative of \( \sigma_{r,q,t} \) with respect to \( \tilde{q}_{q,t-1} \), and use the fact that \( \sigma_{r,q,t} > 0 \) to sign the derivative.

\[ \frac{\partial}{\partial \tilde{q}_{q,t-1}} \left[ (V_q/V_{q'})\tilde{q}_{q,t-1} \left[ \tilde{q}_{q',t-1}^2 \sigma_{r,q',t} + \tilde{q}_{q',t-1} \right] - \tilde{q}_{q,t-1}^{-1} \right] \]
\[ = \tilde{q}_{q,t-1}^{-2} - (V_q/V_{q'})\tilde{q}_{q,t-1} \left[ \tilde{q}_{q',t-1}^2 \sigma_{r,q',t} + \tilde{q}_{q',t-1} \right] \times 2\tilde{q}_{q,t-1}^{-1} \]
\[ = \left( \tilde{q}_{q,t-1}^{-1} - (V_q/V_{q'})\tilde{q}_{q,t-1} \left[ \tilde{q}_{q',t-1}^2 \sigma_{r,q',t} + \tilde{q}_{q',t-1} \right] \right) \tilde{q}_{q,t-1}^{-1} \]
\[ < -(V_q/V_{q'})\tilde{q}_{q,t-1} \left[ \tilde{q}_{q',t-1}^2 \sigma_{r,q',t} + \tilde{q}_{q',t-1} \right] \tilde{q}_{q,t-1}^{-1} \]
\[ < 0. \]

\[ \square \]

**Proof of Corollary 2.** \( \tilde{q}_{q,t-1} = \frac{1}{\sigma_{q,t}} + \sum_{s=\tau_q}^{t-1} \sum_{r \in R_{q,s}} \gamma_{r,q,s} \). The derivatives with respect to \( n_{q,s} \) and \( \gamma_{r,q,s} \) are clearly positive. Further, fixing \( \{ R_{q,s}, \{ \gamma_{r,q,s} \}_{r \in R_{q,s}} \}_{s=\tau_q \ldots \infty} \), increasing \( t \)
to \( t' \) strictly increases \( \sum_{s=r}^{t-1} \sum_{r \in R_{q,s}} \gamma_{r,q,s} \) so long as \( n_{q,s} > 0 \) for some \( s \in \{ t, t+1, \ldots, t'-1 \} \). To see how the importance of the question enters, I follow the lines of the proof of Proposition 2 and take a derivative of the threshold value \( \bar{\sigma}_{r,q,t} \), assuming that it is positive. If it is zero, then nobody is studying the question so researchers are weakly more biased as any parameters are changed. The derivative is

\[
\frac{\partial}{\partial V_q} (V_q/V_{q'}) \tilde{\gamma}_{q,t-1}^{-2} \left[ \tilde{\gamma}_{q,t-1}^2 \tilde{\sigma}_{r,q',t}^2 + \tilde{\gamma}_{q',t-1} \right] - \tilde{\gamma}_{q,t-1}^{-1} \\
= (1/V_{q'}) \tilde{\gamma}_{q,t-1}^{-2} \left[ \tilde{\gamma}_{q,t-1}^2 \tilde{\sigma}_{r,q',t}^2 + \tilde{\gamma}_{q',t-1} \right] \\
> 0.
\]

\[\Box\]

**Proof of Corollary 3.** The variance of the non-expert opinion in period \( t \) is \( 1/\tilde{\gamma}_{q,t-1} \).

The variance of the opinion of expert \( r \) is

\[
\text{var} \left[ \sum_{s=r}^{t-1} \sum_{m \in R_{q,s}} \left( \frac{\gamma_{m,q,s}}{\tilde{\gamma}_{q,t-1} + n_{q,t} \tilde{\gamma}_{r,q,t}} \right) \mu_{m,q,s} + \sum_{m \in R_{q,t}} \left( \frac{\tilde{\gamma}_{r,q,t}}{\tilde{\gamma}_{q,t} + n_{q,t} \tilde{\gamma}_{r,q,t}} \right) \mu_{m,q,t} \right] \\
= \sum_{s=r}^{t-1} \sum_{m \in R_{q,s}} \left( \tilde{\gamma}_{q,t-1} + n_{q,t} \tilde{\gamma}_{r,q,t} \right)^{-2} \sigma_{m,q,s}^2 + \sum_{m \in R_{q,t}} \left( \tilde{\gamma}_{r,q,t} \right)^{-2} \sigma_{m,q,t}^2 \\
= \sum_{s=r}^{t-1} \sum_{m \in R_{q,s}} \left( \tilde{\gamma}_{q,t-1} + n_{q,t} \tilde{\gamma}_{r,q,t} \right)^{-2} + \sum_{m \in R_{q,t}} \left( \tilde{\gamma}_{r,q,t} \right)^{-2} \left( \sigma_{m,q,t}^2 / \sigma_{r,q,t}^2 \right).
\]

Because \( \tilde{\sigma}_{r,q,t}^2 \) has support \([0, \bar{\sigma}]\), \( \sigma_{m,q,t}^2 / \sigma_{r,q,t}^2 \) can be arbitrarily high, so an expert may put nearly full weight on current research, regardless of the volume and quality of past research. \[\Box\]

**Proof of Proposition 3.** As \( \sigma_{r,q,t} \) decreases, \( G_{r,q,t}(x) \) increases for all \( x \in [0, \bar{\sigma}/\sigma_{r,q,t}] \).

For every set of values \( \{ \tilde{\gamma}_{q',t-1}^{-2}, V_{q'} \} \) \( q' \in Q_t \setminus q \) and \( V_q \), the threshold for a researcher to choose
question $q$ is $\bar{\sigma}_{r,q,t}$. I show that expected bias is increasing in $\sigma_{r,q,t}$ for any given values $\{\gamma_{q',t-1}^{-2}, V_{q'}\}_{q'\in Q_t\setminus q}$ and $V_q$, so it is increasing in $\sigma_{r,q,t}$. The expected bias given $\{\gamma_{q',t-1}^{-2}, V_{q'}\}_{q'\in Q_t\setminus q}$ and $V_q$ is $E(K \mid r \text{ chosen}) = \frac{1}{H_{r,q,t}(\bar{\sigma}_{r,q,t}/\sigma_{r,q,t})} \int_0^{\sigma_{r,q,t}} x h_{r,q,t}(x) dx$. Taking a derivative with respect to $\sigma_{r,q,t}$ yields

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

$$- \frac{h_{r,q,t}(\bar{\sigma}_{r,q,t}/\sigma_{r,q,t})}{H_{r,q,t}(\bar{\sigma}_{r,q,t}/\sigma_{r,q,t})} (\bar{\sigma}_{r,q,t}^2/\sigma_{r,q,t}) \left[ \int_0^{\sigma_{r,q,t}/\sigma_{r,q,t}} x h_{r,q,t}(x) dx - \bar{\sigma}_{r,q,t} \right]$$

$$= \frac{h_{r,q,t}(\bar{\sigma}_{r,q,t}/\sigma_{r,q,t})}{H_{r,q,t}(\bar{\sigma}_{r,q,t}/\sigma_{r,q,t})} \left[ E(K \mid K \leq \bar{\sigma}_{r,q,t}/\sigma_{r,q,t}) - \bar{\sigma}_{r,q,t} \right].$$

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 $\bar{\sigma}_{r,q,t}$ such that question $q$ is studied is

$$\bar{\sigma}_{r,q,t} = \left( V_q/V_{q'} \right) \gamma_{q',t-1}^{-2} \left[ \gamma_{q',t-1}^2 \bar{\sigma}_{r,q',t}^2 + \gamma_{q',t-1} \right] - \gamma_{q,t-1}^{-1}$$

for the best alternative $q'$, as perceived by researcher $r$. This alternative is given by

$$q'(r, q, t) = \arg \min_m \left\{ \left[ \gamma_{m,t-1}^2 \bar{\sigma}_{r,m,t}^2 + \gamma_{m,t-1} \right] / V_m \right\} \quad \text{for } q' \in Q_t \setminus q.$$

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 min operator is weakly decreasing as arguments are added, so letting $q''(r, q, t) = \arg \min_m \left\{ \left[ \gamma_{m,t-1}^2 \bar{\sigma}_{r,m,t}^2 + \gamma_{m,t-1} \right] / V_m \right\} \quad \text{for } q'' \in Q'_t \setminus q$, we have

$$\left[ \gamma_{q'',t-1}^2 \bar{\sigma}_{r,q'',t}^2 + \gamma_{q'',t-1} \right] / V_{q''} \leq \left[ \gamma_{q',t-1}^2 \bar{\sigma}_{r,q',t}^2 + \gamma_{q',t-1} \right] / V_{q'}.$$
Therefore, $\bar{\sigma}_{r,q,t}(Q_t') \leq \bar{\sigma}_{r,q,t}(Q_t)$, where other arguments of $\bar{\sigma}_{r,q,t}$ are suppressed for clarity.

**Proof of Proposition 5.** A researcher chooses question $q$ if and only if $u_{r,q,t} \geq u_{r,q',t}, \forall q' \in Q_t \setminus q$. Therefore, anything that increases her utility from choosing a research question weakly increases the likelihood of her choosing that question.

\[
u_{r,q,t} = E \left[ V_q - \left( \mu_{q,t}^P - M_{q,t} \right)^2 \mid \widetilde{\sigma}_{r,q,t} \right]
= V_q - E \left( \mu_{q,t}^P \mid \widetilde{\sigma}_{r,q,t} \right)^2 - E \left( M_{q,t} \mid \widetilde{\sigma}_{r,q,t} \right)^2 + 2 E \widetilde{\sigma}_{q,t} \left( \mu_{q,t}^P M_{q,t} \mid \widetilde{\sigma}_{r,q,t} \right).
\]

Because $\mu_{q,t}^P$ and $M_{q,t}$ have mean $\mu_{q,t}$, $E \left( \mu_{q,t}^P \mid \widetilde{\sigma}_{r,q,t} \right)^2 = \mu_{q,t}^2 + (\sigma_{q,t}^P)^2$, where $\sigma_{q,t}^P$ is researcher $r$’s belief of the standard deviation of her posterior. $E \left( M_{q,t} \mid \widetilde{\sigma}_{r,q,t} \right)^2 = \mu_{q,t}^2 + (\sigma_{q,t}^M)^2$, and because $\mu_{q,t}^P$ and $M_{q,t}$ are independent, $E \widetilde{\sigma}_{q,t} \left( \mu_{q,t}^P M_{q,t} \mid \widetilde{\sigma}_{r,q,t} \right) = \mu_{q,t}^2$. I can therefore re-write the utility function as

\[
u_{r,q,t} = V_q - (\widetilde{\sigma}_{q,t})^2 - (\sigma_{q,t}^M)^2.\]

Part 1 of proposition 5 follows because $(\sigma_{q,t}^P)^2$ is increasing in $\widetilde{\sigma}_{r,q,t}$, so $\nu_{r,q,t}$ is decreasing in $\widetilde{\sigma}_{r,q,t}$. Part 2 follows because $\sigma_{q,t}^P$ is decreasing in $\widetilde{\gamma}_{q,t-1}$, so $\nu_{r,q,t}$ is increasing in $\sigma_{q,t}^P$. Part 3 follows because $\nu_{r,q,t}$ increases one for one with $V_q$. ■

7 Figures
Figure 1: The upper two figures display expert and non-expert precision for a more (left) and a less (right) important question, over time. The bottom figures display the number of researchers, out of a potential 39, choosing those questions over the outside option. As the marginal contribution of research decreases, researchers begin choosing the outside option, and the bias among experts outweighs their superior knowledge. These figures are appropriate for a large, growing field.
Figure 2: The upper two figures display expert and non-expert precision for a more (left) and a less (right) important question, over time. The bottom figures display the number of researchers, out of a potential 39, choosing those questions over the outside option. As the marginal contribution of research decreases, researchers begin choosing the outside option, and the bias among experts outweighs their superior knowledge. At $t = 7$, the research technology becomes 10 times more productive, causing a surge of work and improving the quality of expert opinion. The surge is temporary. These figures are appropriate for a large, growing field.
Figure 3: These figures display average expert and non-expert precision over time when the field has two questions and no outside option. In the top figure, the questions are identical. Even though the types of expert choosing the question do not change over time, the marginal impact of their work decreases, so the problem of bias eventually overtakes the benefit of that marginal work. In the bottom figure, the distribution of expert beliefs is four times wider for question Q2 than for question Q1. This does not affect the types choosing to study each question, but means that the bias in question Q1 is not sufficient, in only 10 periods, to make experts in that question less reliable than non-experts.
Figure 4: These figures plot the precision of expert and non-expert opinion for a field with two questions and no outside option. Question Q1 is twice as important, but otherwise the questions are identical. In the first few periods, researchers oscillate between choosing the question that is more important and the one that is less researched. This yields oscillation in the quality of expert opinion. At any given time, the less studied question features experts that are little better than non-experts, while the heavily studied question features superior experts. In the long run, a steady-state is reached in which more researchers choose the more important question.
Figure 5: These figures plot experts’ reported and actual posterior precisions for a question in a large, growing field. In period 1, by construction, all potential experts choose to study the question. For experts with pessimistic priors regarding current work, actual and perceived accuracy increase together. For experts with optimistic priors, the relationship is negative. By period 5, pessimistic researchers choose the outside option and only optimists are left. The relationship between actual and self-assessed accuracy is always negative, but moreso for older questions. Linear regression lines are plotted.