Karen Ruth Adams: “Reflections of a (Female Subject-Matter Expert) Superforecaster”

Posted on October 1, 2014 by admin

In August, I attended the Good Judgment Project’s 2014 Superforecaster Conference, where the top 2% of last year’s 7,200 forecasters and top forecasters from previous seasons met with the principal investigators. We discussed the project’s findings to date, changes for Season 4, and plans for the future.

At the conference, I was struck by both the diversity and lack of diversity among the forecasters. There was a lot of occupational diversity. I met people who work in finance, IT, materials science, law, and other commercial sectors. Yet, considering that the aim of the study is to improve national security forecasting by US intelligence agencies, there was a notable lack of subject-matter experts. I met just a handful of security scholars from academia and think tanks and policy makers and practitioners from government and non-profits.

I was also surprised to see few women. It wasn’t that I expected a 50/50 ratio. I thought women would be 25-33% of forecasters, mirroring the percentage of women among American faculty in political science, security scholars at the International Studies Association, policy analysts and leadership staff at Washington think tanks, and senior US national security and foreign policy officials. Instead I learned from GJP researcher Pavel Atanasov that at the beginning of Season 3 (Fall 2013), women were just 17% of GJP forecasters. By the end of the season (Spring 2014), women had dropped out at higher rates (35%) than men (29%). Among this year’s superforecasters (last year’s top 2%), just 7% are women.

As a woman who has spent decades developing expertise on international relations and human, national, and international security, and as a citizen who would like US security forecasting and policy to improve, this concerns me. It also concerns GJP’s principal investigators, who have asked forecasters to offer suggestions for improving the mix. This post is a contribution to that conversation. I explain why I joined the project and what I’ve done and learned so far. I also offer some thoughts about what remains to be discovered and improved about gender and expertise among forecasters.

Why I Joined the Project

In March 2011, I received an intriguing email via a listserv of strategic studies scholars. Bob Jervis, a noted expert on national and international security, was looking for

> knowledgeable people to participate in a quite unprecedented study of forecasting sponsored by Intelligence Advanced Research Projects Activity (“IARPA”) and focused on a wide range of political, economic and military trends around the globe. The goal of this unclassified project is to explore the effectiveness of techniques such as prediction markets, probability elicitation, training, incentives and aggregation that the research literature suggests offer some hope of helping forecasters see at least a little further and more reliably into the future.

Bob was recruiting for the GJP team on behalf of principal investigators Barbara Mellers, Don Moore, and Phil Tetlock. According to Bob, the “minimum time commitment would be several hours in passing training exercises, grappling with forecasting problems, and updating your forecasting response to new evidence throughout the
year.” The rewards would be “Socratic self-knowledge,” the opportunity to learn and be assessed on “state-of-the-art techniques (training and incentive systems) designed to augment accuracy,” a $150 honorarium, and the opportunity to compete anonymously with the freedom to go public later. In addition, Bob said he thought it would be fun.

I immediately said yes, for two reasons. First, I remembered Phil Tetlock from my time as a political science graduate student at UC Berkeley, and I trusted him to run an interesting and high-quality study in which my anonymity would be protected. That was important to me because I wanted to take the risk of forecasting without worrying about the effects on my scholarly reputation. After all, in *Expert Political Judgment*, Phil had shown that experts (highly educated professionals in academia, think tanks, government, and international organizations) weren’t much better at forecasting than “dilettantes” (undergraduates). Moreover, they had trouble outperforming “dart-throwing chimps” (chance) and generally underperformed computers (extrapolation algorithms). As an expert on security studies, this was my chance to try to prove him wrong.

Second, as a recently-tenured political science professor, I had begun to expand my research program, focusing less on individual publications and more on my career contribution. For several years, I had been developing a new framework for studying, teaching, and improving human, national, and international security. One element of the project is helping students evaluate and explain the historical and current security levels of various actors (individuals, social groups, and states) and predict future security levels. Thus the opportunity to find out more about forecasting and try my hand as a participant-observer was too good to ignore.

What I’ve Done So Far

Since Fall 2011, I’ve participated on the GJP team in all three years of IARPA’s ACE tournament. In Season 1, I was in an individual prediction polling group, making individual forecasts on a survey platform with no interaction with other participants. In Season 2, I was on an interactive individual survey platform, where participants could explain their forecasts and see their own and others’ accuracy ratings (Brier scores). In Season 3, I was on the Inkling platform, one of two large prediction markets, where participants were given 50,000 “inkles” to buy and sell “stocks” in answers to questions, with probabilities expressed as prices. We could also make comments and see one another’s scores (earnings).

Over time, my accuracy has improved. I moved from the top 18% in Season 1 (Brier score of .42) to the top 8% in Season 2 (Brier score of .34), and top 1% in Season 3 (no Brier score because I was in the prediction market, where I more than tripled my “money,” finishing 6th of 693 forecasters). My best categories have consistently been those in which I have the most expertise: international relations, military conflict, and diplomacy.

How GJP Has Enhanced My Skills and Confidence

Participating in the study has done what I hoped it would. It has improved my forecasting skill. By compelling me to express forecasts in stark probabilistic terms and by using clear and generally fair rules to score them, GJP has given me a laboratory in which to learn how to balance the forecasting skills of “calibration” (understanding base rates) and “discrimination” (identifying exceptions).

Now, when a colleague or reporter asks what I think will happen in an international conflict or international organization, I think more clearly about the theories and facts I’m using to arrive at my answer, the probabilities I assign to various outcomes, and the confidence I wish to express. In my security class, I model this process for my
students and ask them to make their own forecasts.

Participating in the GJP has also increased my confidence. Like most experts, I used to be reluctant to make point predictions that could run afoul of complexity or chance and be taken out of context. Moreover, like many professional women, I suffered from both “imposter syndrome” — the feeling that I don’t know enough and will be found out — and the knowledge that women’s qualifications and contributions are systematically discounted. Thus it never helped to be told that I should bluff, like men.

Thanks to the GJP, I’ve learned I don’t have to pretend to be something I’m not. I have a good sense of where my expertise lies, where it makes the most difference in improving group accuracy, and when and on what terms I wish to go public with predictions. I also know it’s not a weakness but a strength to approach forecasting with humility.

As Bob Jervis predicted, participating in the GJP has also been fun. I don’t worry about being right all of the time. In Season 3, I was one of the most frequent commenters, revealing my forecasts and logic, and asking for feedback. When I’m wrong, I have other forecasters to laugh with and learn from.

In August, before the superforecaster conference, I revealed my identity. That was a surprise to some of my fellow forecasters, who had assumed Fork Aster was a man.

What Remains to be Discovered about Gender and Expertise

Before the superforecaster conference, I wondered if being out as a female subject-matter expert would change the dynamics of my participation. Would I speak up less often for fear of being dubbed a pointy-headed intellectual? Would GJP turn into a forum in which women’s comments were discounted or ignored, or in which successful women were deemed unlikable?

My concerns about gender were allayed at the conference. Although women are just 7% of superforecasters, they were not segregated by choice or default. Women sat and stood and worked in groups with men. To me, this shows the value of initial anonymity. Superforecasters were known to be good forecasters, whether or not they were known to be women. The women who were there had made a cut based on merit, so they were accepted and confident.

But they weren’t overconfident. After all, the GJP’s major findings are that forecasters perform best when they understand probability, are open-minded, and are scored for accuracy. Together, this means superforecasters of all stripes know that perfection is unattainable, there’s a good chance they’re wrong and should listen to other views, and the best way to improve is to put themselves out there to be scored.

My concerns about expertise were allayed in the first month of Season 4. In the superforecaster market, I’ve been speaking up about as much as I did last year. Moreover, I’ve found that instead of saying less for fear of being wrong, I’ve been tempted to say more than I can confidently support simply to burnish my credentials. Thus this year is shaping up to be a test of whether I can remain open-minded despite having my reputation at stake. With the recent brouhaha about faux experts and experts-for-hire, this will be very interesting indeed.

Like all forecasters, I care a great deal about how “right” and “wrong” are scored. As someone who is trying to build her confidence, I also care about ratings and rankings. But I’m not motivated by fake money, and I doubt
most subject-matter experts are either. So this year, I’m looking forward to a new metric, “market contribution,” which will summarize each market forecaster’s contribution to the market’s accuracy (Brier score). To understand what motivates forecasters of various types, I hope GJP will track via surveys and team and market behavior the extent to which individuals seem to be motivated by problem-solving, competition, social interaction, accuracy, and other goals.

**Why and How Female and Expert Participation Should Be Improved**

In one field after another, studies have found that groups with more diverse participants and in particular more women make more accurate decisions. That is reason enough for GJP to redouble its efforts to recruit and retain women. It also speaks to the importance of preserving occupational diversity. Yet GJP should also make an effort to recruit more subject-matter experts. Otherwise, it will be hard to evaluate whether Tetlock’s earlier findings about the overall unreliability of expert political judgment are valid.

Although I’m not an expert on the effects of gender and expertise on participation in scientific studies, I have some thoughts about how GJP could recruit and retain more women and subject-matter experts.

First, it’s important to think about how the recruiting pitch sounds. The one I got was perfect for me. It appealed to my expertise and love of learning, my desire to improve US security policy, and my sense of fun. It also seemed reasonable. A few hours, a few updates… no big commitment. In fact, the requirement has been higher. In the first three years, it took me about 5 hours per week on top of my regular current events reading to research, answer, and discuss the required 25 questions. Since women spend more time than men on the “second shift” of family and household work, the time requirement probably depresses female participation and retention rates. If security scholars and practitioners have heavier work obligations than individuals in the private sector, high time commitments could depress their participation as well. Since the whole point of expertise is to be good at something in particular instead of everything in general, perhaps GJP should set different participation expectations for subject-matter experts. It would be more fair to all forecasters, though, either to reduce the time requirements overall or to provide more financial or reputational compensation.

Second, where does the recruiting pitch go? According to Project Director Terry Murray, the GJP has not made a systematic effort to recruit a diverse pool of forecasters. Instead, the project has relied on word of mouth by the principal investigators and advisors (most of whom are men) and serendipitous media coverage. To include more women across the professions, GJP should reach out to interest groups such as the World Affairs Council and skill-building networks such as Lean In. (For a big bang, GJP could collaborate with Lean In to produce a video about how to improve forecasting skills). To reach more subject-matter experts, GJP should recruit through professional organizations such as Women in International Security (which has both male and female members), and the international security studies divisions of the American and international political science associations.

Third, what are GJP training materials, questions, and discussions like? Since women dropped out of Season 3 at higher rates than men, there may be something about the experience itself that’s a turn-off. As a middle-aged, female security expert, I was used to a lot of the bravado I saw on the GJP boards, and I was willing to live with it because I wanted to learn something. I also thought my anonymous participation might improve things. Other women may drop out because they find the exchanges unpleasant or irrelevant, or because they lack the confidence to weigh in.
Still other women may be turned off by some of the recommended reading. I devoured Kathryn Schulz’s *Being Wrong*. But it was a shock to open Daniel Kahneman’s *Thinking Fast and Slow* and discover that the first chapter features the cognitive challenge posed by a photograph of an angry woman. Later, it emerges that that one of the most debated problems in cognitive science relates to experiments in which people erroneously assume that a woman who lives in Berkeley is more likely to be a feminist and a bank teller than simply a bank teller. To increase female participation, researchers and forecasters need to think carefully about their language and examples so they don’t evoke what Kahneman refers to as “System 1 errors” in a whole subset of participants. Although researchers don’t intend for their work to have these effects, if they’re not attentive, it can.

**A Continuing Conversation**

At the superforecaster conference, many of the male participants asked me how I thought female participation could be improved. When they found out I’m a political science Ph.D. and professor, they asked me the same thing about including more security scholars. They were not just being polite. Over the past three years, as we’ve contributed to GJP [reportedly outperforming intelligence analysts](#), we’ve all learned the value of open-minded thinking, and we know it’s more likely in groups with diverse participants whose individual contributions are heard and valued.

For now, my recommendations for GJP are to review the participation requirements, reach out to organizations and networks populated by women and subject-matter experts, and survey current and past participants about their impressions of the work load and content and their reasons for staying in or leaving the project.

My recommendation to women and subject-matter experts is to give forecasting a shot. Decide what you want to get out of the project and what kind of participant you want to be. Then do your best. See what you learn and what others learn from you. Forecast anonymously at first, then come out if you like. I predict a lot of wonky fun – intellectual puzzles, interesting exchanges of ideas, head-to-head competition, some memorable “Aha!” moments, and the pride of knowing that you have contributed in some small way to improving security forecasting.

And you? What do you suggest? Let’s continue the conversation.

>>>>

Karen Ruth Adams (aka Fork Aster) is an associate professor of political science at the University of Montana.

---

This entry was posted in [IARPA ACE Tournament](#), [Recruitment](#), [superforecasters](#) and tagged [Karen Ruth Adams](#). Bookmark the [permalink](#).