

# SSC Journal Club: Expert Prediction Of Experiments

Posted on November 27, 2016 by Scott Alexander



## I

It's been a good month for fretting over failures of expert opinion, so let's look at DellaVigna & Pope, [Predicting Experimental Results: Who Knows What?](#) The authors ran a pretty standard behavioral economics experiment where they asked people on Mechanical Turk to do a boring task while being graded on speed and accuracy. Then they offered one of fifteen different incentive schemes, like "we'll pay you extra if you do well" or "your score will be publicly visible".

But the point of the study wasn't to determine which incentive scheme worked the best, it would determine *who could best predict* which incentive scheme worked the best. The researchers surveyed a bunch of people – economics professors, psychology professors, PhD students, undergrads, business students, and random Internet users on Mechanical Turk – and asked them to predict the experimental results. Since this was a pretty standard sort of behavioral economics experiment, they were wondering whether

people with expertise and knowledge in the field might be better than randos at figuring out which schemes would work.

They found that knowledgeable academics had some advantage over randos, but with enough caveats that it's worth going over in more detail.

First, they found that prestigious academics did no better (and possibly slightly worse) than less prestigious academics. Full professors did no better than associate professors, assistant professors, or PhD students. People with many publications and citations did no better than people with fewer publications and citations.

Second, they found that field didn't matter. Behavioral economists did as well as microeconomists did as well as experimental psychologists did as well as theoretical psychologists. To be fair, this experiment was kind of in the intersection of economics and psychology, so all of these fields had equal claim to it. I would have liked to see some geologists or political scientists involved, but they weren't.

Third, the expert advantage was present in one measure of accuracy (absolute forecast error), but not in another (rank-order correlation). On this second measure, experts and randos did about equally well. In other words, experts were better at guessing the exact number for each condition, but not any better at guessing which conditions would do better or worse relative to one another.

Fourth, the expert advantage was pretty small. Professors got an average error of 169, PhD students of 171, undergrads of 187, MBA students of 198, and MTurk users of 271 (random guessing gave an error of about 416). So the difference between undergrads and experts, although statistically significant, was hardly overwhelming.

Fifth, even the slightest use of “wisdom of crowds” was enough to overwhelm the expert advantage. A group of five undergrads averaged together had average error 115, again compared to individual experts’ error of 169! Five undergrads averaged together (115) did about as well as five experts averaged together (114). Twenty undergrads averaged together (95) did about as well as twenty experts averaged together (99).

Sixth, having even a little knowledge of individuals’ forecasting ability screened off expert status. The researchers gave forecasters some experimental data about the effects of a one-cent incentive and a ten-cent incentive, and asked them to predict the scores after a four-cent incentive – a simple, mechanical problem that just requires common sense. Randos who can do well on this problem do just as well as experts on the experiment as a whole. Likewise, randos who are noticed to do well on the first half of the experiment will do just as well as experts on the second half too. In other words, we’re back to finding “superforecasters”, people who are just consistently good at this kind of thing.

None of this seems to be too confounded by effort. The researchers are able to measure how much time people take on the

task, whether they read the instructions carefully, etc. There is some advantage to not rushing through the task, but after that it doesn't seem to matter much. They also try offering some of the Mechanical Turkers lots of money for getting the answers right. That doesn't seem to help much either.

The researchers ask the experts to predict the results of this experiment. They (incorrectly) predict that prestigious academics with full professorships and lots of citations will do better than mere PhD students. They (incorrectly) predict that psychologists will do better than non-psychologists. They (correctly) predict that professors and PhD students will do better than undergrads and randos.



What do we make of this?

I would *tentatively* suggest it doesn't look like experts' expertise is helping them very much here. Part of this is that experts in three different fields did about equally well in predicting the experimental results. But this is only weak evidence; it could be that the necessary expertise is shared among those three fields, or that each field contains one helpful insight and someone who knew all three fields would do better than any of the single-field experts.

But more important, randos who are able to answer a very simple question, or who do well on other similar problems, do just as well

as the experts. This suggests it's possible to get expert-level performance just by being clever, without any particular expertise.

So is it just IQ? This is a tempting explanation. The US average IQ is 100. The undergrads in this experiment came from Berkeley, and Berkeley undergrads have an average SAT of 1375 = average IQ of 133 (this seems really high, but apparently matches estimates from [The Bell Curve](#) and the [Brain Size blog](#) ; however, see Vaniver's point [here](#)). That same Brain Size post proposes that the average professor has an IQ of 133, but I would expect psychology/economics professors to be higher, plus most of the people in this experiment were from really good schools. If we assume professors are 135-140, then this would neatly predict the differences seen from MTurkers to undergrads to professors.

But the MBA students *really* don't fit into this model. The experiment gets them from the University of Chicago Booth School of Business, which is the top business school in the country and has an average GMAT score of 740. That corresponds to an IQ of almost 150, meaning this should be the highest-IQ sample in the study, yet the MBAs do worse than the undergrads. Unless I'm missing something, this is fatal to an IQ-based explanation.

I think that, as in *Superforecasting*, the best explanation is a separate "rationality" skill which is somewhat predicted by high IQ and scientific training, but not identical to either of them. Although some scientific fields can help you learn the basics of thinking clearly, it doesn't matter what field you're in or whether you're in any field at all as long as you get there somehow.

I'm still confused by the MBA students, and expect to remain so. All MBA students were undergraduates once upon a time. Most of them probably took at least one economics class, which was where the researchers found and recruited their own undergraduates from. And most of them were probably top students from top institutions, given that they made it into the best business school in the US. So how come Berkeley undergraduates taking an econ class outperform people who *used to be* Berkeley undergraduates taking an econ class, but are now older and wiser and probably a little more selected? It might be that business school selects against the rationality skill, or it might be that business students learn some kind of anti-insight that systematically misleads them in these kinds of problems.

(note that the MBAs don't put in less effort than the other groups; if anything, the reverse pattern is found)



Does this relate to interesting real-world issues like people's trouble predicting this election?

One important caveat: this is all atheoretical. As far as I know, there's no theory of psychology or economics that should let people predict how the incentive experiment would go. So it's asking experts to use their intuition, supposedly primed by their expertise, to predict something they have no direct knowledge about. If the experiment were, say, physicists being asked to predict the speed

of a falling object, or biologists being asked to predict how quickly a gene with a selective advantage would reach fixation, then we'd be in a very different position.

Another important caveat: predictive tasks are different than interpretative tasks. Ability to predict how an experiment will go without having any data differs from ability to crunch data in a complicated field and conclude that eg saturated fat causes/doesn't cause heart attacks. I worry that a study like this might be used to discredit eg nutritional experts, and to argue that they might not be any better at nutrition than smart laymen. Whether or not this is true, the study doesn't support it.

So one way of looking at it might be that this is a critique not of expertise, but of "punditry". Engineers are still great at building bridges, doctors are still great at curing cancer, physicists are still great at knowing physics – but if you ask someone to predict something vaguely related to their field that they haven't specifically developed and tested a theory to cope with, they won't perform too far above bright undergrads. I think this is an important distinction.

But let's also not get too complacent. The experts in this study clearly *thought* they would do better than PhD students. They *thought* that their professorships and studies and citations would help them. They were wrong. The distinction between punditry and expertise is pretty fuzzy. Had this study come out differently, I could have argued for placing nice clear lab experiments about in-

centive schemes in the “theory-based and amenable to expertise” category. You can spin a lot of things either direction.

I guess really the only conclusion you can draw from all of this is not to put any important decisions in the hands of [people from top business schools](#).