http://social-epistemology.com
ISSN: 2471-9560

On Skipper's Humility Heuristic

Marco Meyer, University of York, marco.meyer@york.ac.uk

Meyer, Marco. 2020. "On Skipper's Humility Heuristic." *Social Epistemology Review and Reply Collective* 9 (11): 38-43. https://wp.me/p1Bfg0-5we.

SERRC
Social Epistemology
Review & Reply Collective

vol 9 (11): 38-43. 2020.
https://wp.me/p1Bfg0-5we

In '*The Humility Heuristic Or: People Worth Trusting Admit to What They Don't Know,"* Mattias Skipper defends a heuristic for identifying trustworthy people. In slogan form, the Humility Heuristic says that people worth trusting admit to what they don't know. As a general rule, you should trust my assertion that P more if I admit ignorance regarding some other matter Q.

What makes the Humility Heuristic intriguing is its minimalism: it does not require collaboration with others , and it does not require any prior education (as opposed to: Anderson, 2011; Brennan, 2020). Whether the advisee knows about the claim the advisor admits ignorance about makes no difference.

Skipper uses methods from formal epistemology. He derives the Humility Heuristic from three assumptions that appear highly plausible at first sight. Skipper manages both to make a strong case for each of the assumption, as well as pointing out scenarios in which they may not hold. The discussion is immensely helpful to clarify the conditions under which the Humility Heuristic holds.

I want to explore two questions with respect to the Humility Heuristic that Skipper leaves open: First, when should we use the humility heuristic, and how much weight should we give it? Second, what implications should advisors draw from the Humility Heuristic?

In discussing these questions, I advance two points. My first point is that the formal methods Skipper applies are insufficient to make progress on these questions. We need to complement the formal approach Skipper uses with an empirically informed approach. My second point is that Skipper's static analysis should be supplemented by a dynamic perspective. One way for the Humility heuristic to be effective is by making advisors more discerning in when to assert knowledge and when to admit ignorance. This is because the fact that advisees rely on the humility heuristic creates an incentive for well-intentioned advisors to resist overconfidence. This dynamic could greatly benefit from formalization using methods from game theory.

I hasten to add that Skipper does not claim otherwise. I submit the following not as a critique, but as an attempt to build on Skipper's paper.

The Limits of the Humility Heuristic

At its core, the heuristic relies on a two-step inference: we infer from the advisor's admission of ignorance that the advisor is more likely to assert knowledge if the advisor is confident they know; and from the advisor's confidence we infer that their assertion is more likely to be true.

The Humility Heuristic is a *heuristic* because each of these inferences are useful rules of thumb, but it can fail. The heuristic helps in guarding against placing trust in *overconfident* advisors. It fails, though, when dealing with *manipulative* advisors, undermining the first inference. It also fails when dealing with *incompetent* advisors, undermining the second inference.

Consider the case of manipulation first. The Humility Heuristic is open to exploitation. If an advisor knows that you rely on the humility heuristic, this makes it *easier* for the manipulative advisor to trick you into trusting them. All they need to do is to sprinkle some admissions of ignorance into the conversation. This is enough to increase your credence that they are trustworthy when asserting P, even though P may be a patent falsehood.

The Humility Heuristic is exploitable because it assumes that if people modulate the strength of their assertions, this is because they map the strength of their assertions to their credence in these assertions. But there are other reasons for modulating the strength of your assertion. The reason that directly undermines the Humility Heuristic is manipulative intent.

There are more reasons that can drive a wedge between how confident you are and the strength of your assertion. One example is culture. In certain countries, sects, or professional fields, people are socialized to habitually display humility. Yet this display of humility may be entirely divorced from the credence people have in their assertions.
Consider the case of incompetence. The Humility Heuristic backfires for advisors who are bad at acquiring knowledge. Suppose an advisor admits to ignorance in the absence of evidence. Yet if evidence is available, the advisor works as a 'negative Bayesian': evidence in favor of P reduces the advisor's credence that P, and *vice versa*. There is no reason to trust the assertion of this advisor that P more because the advisor admitted to ignorance about Q.

Admittedly, negative Bayesians are an extreme case. But the issue generalizes to 'bad Bayesians'. Similar to negative Bayesians, bad Bayesians claim ignorance if they have no evidence. Evidence in favor of P does not always reduce their credence that P, but nor does it reliably increase their credence that P. Sometimes evidence might have the desired effect. But equally often it does have the opposite effect. As in the case of negative Bayesians, there is no reason to trust the assertion of bad Bayesians that P more because they admitted to ignorance about Q.

Skipper is well aware that the Humility Heuristic holds only conditionally. Importantly, he is clear that the Humility Heuristic makes an ordinal claim. It says that you should trust my assertion that P more if I admitted ignorance about Q, but it does not say *how much more* you should trust my assertion that P. Still, to warrant the name *heuristic*, the Humility Heuristic should be applicable at least in some context, enabling its users to place their trust more intelligently than they otherwise would. In the following sections, I discuss how empirical research can help to determine whether and when the Humility Heuristic is useful.

When Should We Use the Humility Heuristic?

The humility heuristic is most useful when advisors are prone to overconfidence, and manipulation and incompetence are comparatively smaller problems. We should rely on the humility heuristic when we know we are dealing with well-intentioned, competent individuals, who might however suffer from overconfidence. This is clearly an important use case. But it takes an independent assessment of the likely motivations and skills of your advisor to determine whether to lean on the Humility Heuristic.

Empirical research can bolster the case for the importance of a heuristic screening for overconfidence (Moore, 2020). Werner De Bondt and Richard Thaler have stated "Perhaps the most robust finding in the psychology of judgment is that people are overconfident" (De Bondt & Thaler, 1995). Other things equal, the assertions of people who are overconfident are less trustworthy, and they are less likely to admit to ignorance. This is precisely the scenario in which the Humility Heuristic shines. The prevalence of overconfidence thus provides a strong reason for using the Humility Heuristic.

What about the undermining conditions of the Humility Heuristic? The likelihood of manipulation is context specific. Whether there is an incentive for manipulation depends, among other things, on how closely the interests of advisor and the advisee are aligned. To manage the risk of manipulation, other heuristics can come to the rescue. For instance, advisees might only rely on the Humility Heuristic if their advisor's interests are aligned with theirs.

One worry about the Humility Heuristic is that it becomes *more* vulnerable to exploitation the more prevalent it is. If the Humility Heuristic is not typically used (and is not known to be typically used), malicious actors are less likely to try to exploit the Humility Heuristic, because the attempt is less likely to be successful. By contrast, if the Humility Heuristic became a household technique, malicious actors would have a strong incentive to exploit it.

What about incompetence? How often do we encounter bad or even negative Bayesians who also admit to ignorance some of the time? At first glance, one might assume that absent manipulative intent, readiness to admitting ignorance is an indicator of a certain level of competence in converting information into credences that track the truth. However, I can think of counter-examples.

One case to consider is the professional economist acting as a pundit. Consider a (perhaps not entirely fictional) world in which the work of the best economists is split in two: As professional economists, they produce impressive academic publications, which adhere to the highest standards of their profession. These publications investigate narrow hypotheses and are careful not to overstate their case. These articles feature discussions of the limitations of their research and admissions of ignorance. By contrast, in their role as pundits, they take bold positions on policy in accordance with their political leanings. These op-eds appeal to economic theory to bolster their case, but they lack the nuance characteristic of the professional work. Importantly, economists-turned-pundits find themselves on opposite ends of the political argument.

For instance: Will the Greek GDP rebound if it adopts austerity measures to overcome their debt crisis? As pundits, these researchers invoke their economic expertise to defend opposite ends of the argument. Assuming that they have access to the same information, at least some of these economists acting as pundits would at least be bad Bayesians.

More generally, research in psychology suggests that experts may sometimes be *worse* than laypeople in predicting outcomes (Tetlock, 2005). One explanation may be that experts get invested in the hypothesis they investigate. This need not undermine the prospects for scientific progress. In fact, the scientific method can be seen as an aid for experts to guard against overstating their case. However, in contexts where the standards of the scientific method are not enforced, the very fact that one is an expert might make it more likely that one is a bad Bayesian.

Let's consider the inverse case of highly competent advisors. To simplify, let's focus on the edge case of the well-intentioned, omniscient advisor. If asked for advice, this highly competent advisor provides it freely, and is never wrong. It is curious that according to the Humility Heuristic, we should discount testimony from this highly competent advisor. The reason is that an omniscient advisor never has an opportunity to admit ignorance – they just always happen to know. This result is curious because the Humility Heuristic downgrades testimony from the most reliable source imaginable.

To move from the edge case to an example more relevant in the real world, consider a child with respect to two potential advisors: a friend their age and an adult. Suppose the child and the adult are equally humble, in that they are equally likely to admit when they don't know. It is plausible that the child knows less than the adult. In fact, it is possible that the adult will have known the answer to all of the child's questions, whereas their friend regularly draws a blank. As a result, the Humility Heuristic would favor the child's young friend.

A way of addressing this issue would be to amend the Humility Heuristic. For instance, rather than relying on admissions of ignorance, one might discount advisors whose advise turns out to be false. A sophisticated way of assessing the performance of experts is to track their Brier score, measuring the accuracy of predictions relative to the credence advisors assign to it (Brier, 1950).

But note that tracking the accuracy of advice chips away at the minimalism of the Humility Heuristic. All the Humility Heuristic requires is information about whether an advisor admits to ignorance. While the Humility Heuristic disadvantages highly competent advisors, the alternative of making accuracy assessments has more demanding informational requirements. For we need to track not only the content of the advisor's testimony, but also assess its accuracy.

How much weight, then, should we give the Humility Heuristic in our reasoning? What the discussion shows is that working out the weights is not just a matter of extending the formal analysis that Skipper develops. At least, we need to know how important the threat of overconfidence is relative to the threats of manipulation and incompetence. This question can be approached on two levels: empirical research can help to determine how important these three threats are in general. Since the answer is dependent on circumstance, we also need to determine in individual cases whether overconfidence is likely the most important threat. That means that the Humility Heuristic is likely most useful when used alongside other heuristics. There are heuristics such as the tracking of accuracy in terms of Brier scores that the Humility Heuristic may well be inferior to. These heuristics, however, have higher informational requirements.

What Can Advisors Learn from the Humility Heuristic?

Skipper's analysis is static, in the sense that he takes it as a given whether an advisor is trustworthy or otherwise. The function of the Humility Heuristic is to equip advisees with a tool to weed out the untrustworthy, and to double down on trustworthy advisors.
Yet, shifting to a dynamic perspective reveals an additional way in which the Humility Heuristic may be effective. Advisors have an incentive to *become more trustworthy* if advisees rely on the Humility Heuristic.

To see why, let's put ourselves in an advisor's shoes. Given our advisees rely on the Humility Heuristic, how should we behave? I argued above that with respect to the possibility of manipulation, the Humility Heuristic has a self-undermining tendency. The more widely the Heuristic is used, the more attractive it becomes to exploit it for malicious actors.

But assume we are a bunch of well-intentioned advisors. If asked for advice, we want to help. Occasionally, we fall prey to overconfidence. Shifting perspective to occasionally overconfident but well-intentioned advisors reveals that the Humility Heuristic also has a self-reinforcing tendency. By rewarding advisors who admit when they don't know, the Humility Heuristic can help to check advisor overconfidence.

When asked for advice on a given issue, advisors face a choice: to submit an opinion or to admit ignorance. If advisees do not rely on the Humility Heuristic, advisors who seek to influence have an incentive to authoritatively state their point, regardless of their credence. By contrast, if advisors know advisees rely on the Humility Heuristic, they also have a countervailing incentive to enhance their trustworthiness by admitting ignorance. On any given occasion, advisors need to decide whether to attempt to influence, or to invest in their trustworthiness.

Hence the Humility Heuristic can start a virtuous circle: advisees scanning advisors for admissions of ignorance can make advisors more trustworthy. This will only happen given a number of requirements. Here are two I can think of: First, interactions between advisors and advisees need to be repeated. Second, advisors need to care more about influencing advisees if they are more confident.

Formal methods from game theory would be hugely helpful in clarifying the requirements for getting this virtuous circle started. Adding a dynamic lens could also help to analyze which of the two effects is more important, the static or the dynamic. Moreover, the dynamic lens could be studies using agent-based models.

I have argued for supplementing Skipper's static formal analysis with empirical methods and with a dynamic analysis. Another obvious extension, which I have not discussed here, is to analyze the Humility Heuristic from the perspective of virtue ethics. All of these proposed extensions are testament to the fruitfulness of the Humility Heuristic beyond Skipper's current paper. I hope they go also some way to illustrate how fruitful it is to use a range of formal and empirical methods as well as theoretical approaches for addressing questions in applied epistemology.

References

Anderson, Elizabeth. 2011. "Democracy, Public Policy, and Lay Assessments of Scientific Testimony." *Episteme*, *8* (2): 144–164.

Brennan, Johnny. 2020. "Can Novices Trust Themselves to Choose Trustworthy Experts? Reasons for (Reserved) Optimism." *Social Epistemology*, *34* (3): 227–240.

Brier, Glenn. W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review*, *78*(1): 1–3.

De Bondt, Werner. F. M., and Richard H. Thaler. 1995. "Financial Decision-Making in Markets and Firms: A Behavioral Perspective." In *Handbooks in Operations Research and Management Science* Vol. 9 edited by Robert Jarrow, Vojislav Maksimovic, William T. Ziembapp, 385–410. Science Direct: Elsevier.

Moore, Don. A. 2020. *Perfectly Confident*. Harper Collins.

Skipper, Mattias. 2020. The Humility Heuristic Or: People Worth Trusting Admit to What They Don't Know. *Social Epistemology* doi: 10.1080/02691728.2020.1809744.

Tetlock, Philip. 2005. *Expert Political Judgment: How Good is It? How Can We Know?* Princeton University Press.