



July 21, 2018

Dear TOK Society List,

As many of you are aware, there has been an unfortunate development involving intellectual property and Hugo Mercier and Dan Sperber's new book *The Enigma of Reason: A New Theory of Human Understanding* (Harvard, paperback 2018). From the book cover summary, their new theory of human reason is that it functions to "help us justify our beliefs and actions to others, convince them through argumentation, and evaluate the justifications and arguments that others address to us. In other words, reason helps humans better exploit their uniquely rich social environment." As readers on this list know, this formulation is remarkably close to the Justification Hypothesis (JH) account of human reasoning, but the book does not mention the JH at all.

If this were simply an issue of the JH being overlooked, that would be a frustrating example of a poor literature review, but I would let it slide. The problem is that I explicitly made Mercier and Sperber's aware of the JH and they seem to have drawn from it. Their work in this area was originally published in 2011, and at that time it was framed as the "argumentation theory" of human reason. When I read it in 2013, it was close enough to the JH that I reached out to the first author (Dr. Mercier) and, in a friendly and collegial way, informed him of the JH and how it was similar. He replied, read what I sent, found it "interesting" and acknowledged that the JH was really a broader theory, and that he and Sperber were focused more on the "cognitive mechanisms."

In the four years since that time, their theory has gone from being close to the JH account to being almost identical to it. Namely, it went from just focusing on argumentation to now being a joint theory of argumentation and justification. And they are still claiming this is a new and original formulation of human reasoning. What this means is that we are looking at a case of either (a) grossly irresponsible and shoddy scholarship (every undergraduate knows that when proposing an idea, one reviews and cites relevant literature); or (b) something more sinister. For purposes of integrity, I believe that the situation must be addressed, and the record needs to be publicly set straight.

As is ethically appropriate, we have been in contact with the authors. They have engaged in strong denial and pushback. The attached packet documents all of what has occurred. I believe all the bases have been covered. For example, in addition to contacting them, I have also consulted some lawyers, including the JMU attorney who, speaking as the JMU attorney agreed this was an "appropriate and professional" response. We have also opened a line of communication with Harvard Publishing.

The next phase is the "release" of this information, which is currently planned for occurring on Monday, August 6th. At that time, the blog in the attached document will be posted, and links to the supporting material will be set up. In addition, a copy of the packet and a copy of the chapter on the Justification Hypothesis will be officially delivered to Harvard University Press, Dan Sperber, Hugo Mercier and their

places of employment. In addition, the packet will be sent to the following media sties and individuals who have covered or reviewed this work. These include: Henry Farrell at the Washington Post, Elizabeth Kolbert at the New Yorker, Olivia Goodhill at Quartz, Ian Sample and Nicola Davis at the Guardian, Tom Stafford of the The MindHacksblog, Cecilia Heyes, Julian Baggini (review in Financial Times), Erik Hare at Barataria, Jaroslav Peregrin, Brian Boyd, and John Brockman at The Edge. If folks become aware of others who might be interested, I would welcome information about that.

We will then see what happens. The goal of this action is to:

1. Set the record straight and make as public as possible (and appropriate) that theirs is not (i) a new theory of human reasoning and that they (ii) were made directly aware of the JH and subsequent moved their work toward it and thus they have not acted with the integrity or responsibility expected of academic scholars.
2. Raise questions/create a teachable moment/foster discussion about what is the line between shoddy scholarship and something more sinister and what should we do in situations like these (i.e., should this be ignored, should legal action be taken, should it just be pointed out in a minor way and give benefit of doubt; how do we decide what our values are and so forth).
3. Use this as an example to understand processes of justification, investment, social influence and the like. We should be able to see these events through the lenses provided by the ToK/UTUA framework and learn from them that way.

These reflections are offered for this list to consider. It would be great if this was something that our community achieved relative consensus about. It is unclear about what will happen after all this information is released. However, to the extent that it is deemed an interesting story and picks up traction, it very well will likely be something that will result in some significant blowback. Folks' reputations are at stake, and that can result in some very strong defensive attacks. So, I would like folks to know that I might be in for some counterattacks here.

I have no interest in getting into a nasty, personal exchange. My plan is to simply focus on the key facts at hand. The idea that human reason evolved in a way that it functions to "justify our beliefs and actions to others" has been clearly presented in the form of the JH since 2003. The authors were made explicitly aware of it five years ago, and since that time their theory moved from being very similar to the JH to being almost identical to it. And yet they continue to claim they have a novel idea and fail to mention the JH in their work. Thus, the facts point clearly to either (a) irresponsible and shoddy scholarship or (b) something more sinister. Full stop.

Thoughts and reflections on this are welcome.

Sincerely,

A handwritten signature in black ink, appearing to read "Greg Hurst". The signature is fluid and cursive, with a large, sweeping flourish at the end.

PS Appendix A to this letter provides folks a snapshot of the three possible scenarios of what happened, with the data pointing strongly to the third scenario.

Appendix A:

Analyses of What Might Have Happened

Whenever we are faced with a situation of conflict involving claims and counterclaims, an outsider generally wants to know what is the most likely thing that has happened. It seems there are basically three possible scenarios, with the third being by far the most likely in my opinion.

Scenario 1. The authors were completely ignorant of the JH (and, to take their account even further, it is debatable if the work should have been cited). This is the scenario that Dan Sperber argues for in his email. Namely, that Hugo Mercier glanced at the work I sent, kindly and quickly replied and never really processed any of it. Thus, the JH did not influence their work at all. In this scenario they are guilty of shoddy scholarship and a poor literature review. That is, unless you take another step in their narrative, which is the claim by Sperber that they were justified in not reviewing work on the JH. His justification for this defense is unclear, because he makes this claim while also stating that he was not at all familiar with the work on the JH. One possible justification is to emphasize that the JH is not well-known in the human reasoning literature. Of course, that is a very weak reason (pun intended) to not cite it, but that is a possible narrative. Bottom line, their case is that this is a minor issue of overlooking some questionably relevant research on partially overlapping ideas.

Scenario 2. Full-fledged and intentional plagiarism by both. In this scenario, the authors were both aware of the JH and deliberately and intentionally stole from it. In this scenario, Mercier notifies Sperber about the JH upon receiving it and both are influenced by the ideas. However, it does not fit exactly with their frame or worldview, but they take the key insight about reason and forge ahead. They are conscious of being influenced by it, read up on it carefully, and consciously engaged in deception.

For a host of reasons which are found in the packet, I find both scenarios not very likely.

Scenario 3. Between shoddy scholarship and something more sinister. This is the in-between scenario that much data point support. Here, Mercier reads the JH and finds it interesting and informative. It plants the seed clearly in him that justification is the key to understanding the function of human reasoning. It is a broader idea that really subsumes the argumentation theory. He begins to raise this idea with Sperber. Sperber, who had been kicking around vaguely similar ideas, sees the light when Mercier prompts him and the two of them run with it. In this scenario Mercier might have alluded to the JH and Sperber blew him off (he clearly is a dominant, alpha male type), or did not say anything. As the ideas took shape, Mercier was aware of the overlap at one level, but wanted to follow Sperber's lead. So, he rationalized to himself that he did not need to say anything strongly about the need to credit the JH. There are several reasons. First, one can point out what Mercier did in his note to me, which is that there is a difference between a cognitive reasoning mechanism (their theory) and a self-conscious interpreter system (a primary focus of the JH). Second, the JH is not salient in the human reasoning literature and they can claim to be experts in that literature so they can rationalize that it is not relevant based on that. And third, if they did cite the JH, then their claims to a new theory of human reasoning would be shot. Together, these elements serve as motives that justify their not focusing on it. In short, between Sperber's irresponsible arrogance and Mercier's deference and deception via omission, they proceeded to convince themselves they were building a new theory, even though the facts are clearly otherwise.