The Bayesian Prisoners' Dilemma

Suppose someone sends you a new article claiming X. Intuitively, we think, "This will either make you more likely to believe X, or have no effect." Once you understand Bayesian reasoning, however, this makes no sense. When someone sends you an article claiming X, you should ask yourself, *"Is this evidence stronger or weaker than I would have expected?"* If the answer is "stronger, " then you should become more likely to believe X. However, if the answer is "weaker," then you should become *less* likely to believe X.

Thus, suppose you initially consider X absurd. When someone sends you some evidence in favor of X, you should update in favor of X if the evidence is less awful than expected. You should update against X, in contrast, only if the evidence is *even more awful* than expected.

Similarly, suppose you initially consider X absurd, but your brilliant friend nevertheless defends it. The fact that a brilliant person believes X is evidence in its favor. *Given* his brilliance, however, his arguments should only persuade you if they are *even better than you would have expected from one so brilliant*. When a great mind offers mediocre arguments, you shouldn't merely be unmoved; you should be actively repelled: "That's the best you can do?!"

Example: One of the smartest people I know routinely sends me pro-"social justice" links on Twitter. As a result, I think even less of the movement than I previously did. If even he fails to defend his view effectively, the view is probably truly devoid of merit.

What, however, should I conclude if this mighty intellect simply stopped sending me links? One possibility, of course, is that he's given up on me. Another possibility, though, is that he's exhausted his supply of evidence. At this point, he's got nothing better than… nothing.

The strange upshot: While Bayesian reasoning seems to imply that persuasive efforts are, on average, ineffective, there is a reason to keep arguing. Namely: Failure to argue is, on average, an admission of intellectual defeat. And by basic Bayesian principles, this in turn implies that the continuation of argument is at least weak evidence in favor of whatever you're arguing.

Stepping back, you can see a somewhat depressing conclusion. When people are perfect Bayesians, argument is a kind of Prisoners' Dilemma.

If your opponent keeps arguing, you want to keep arguing so it doesn't look like you've run out of arguments.

If your opponent stops arguing, you want to keep arguing to emphasize that your opponent has run out of arguments.

As a result, both sides have an incentive to argue interminably. Which, as you may have noticed, they usually do.

Is there any ejector seat out of this intellectual trap? Yes. You could build a credible reputation for talking only when you have something novel to add to the conversation. Then instead of interpreting your silence as, “I’ve got nothing,” Bayesian listeners will interpret it as, “I’ve rested my case.”

[silence]