

When Hindsight Bias is not a Bias

1 Introduction

My favourite fallacy is the fallacy fallacy. It's the fallacy of thinking that something is a fallacy when it isn't. This paper concerns a high-profile instance, namely the phenomenon of hindsight bias. Roughly, it is the phenomenon of being more confident that some body of evidence supports a hypothesis when one knows that the hypothesis is true, than when one doesn't.

Here are a couple illustrations. A juror hears evidence concerning a railroad with a dangerous stretch of track and must judge how probable a derailment was, given the evidence available at the time. Given hindsight bias, her estimate of the probability of derailment is higher if she knows that a train in fact derailed, and she is more likely to deem the railroad company negligent.¹ Second illustration: Subjects are given a case in which a therapist meets with a psychiatric patient who tells her he has been having violent thoughts about harming a third party, but she does not report the threat. Subjects who are also told that the patient in fact injured the third party rate the therapist's *ex ante* evidence as more strongly suggesting the patient would become violent than those who are not informed about the outcome.²

Hindsight bias is almost universally regarded as irrational. After all, that's why it's called a bias. In his seminal 1975 paper, Fischhoff says that those with knowledge of the outcome 'overestimated' the degree to which it would have been reasonable to predict the outcome

¹The train derailment case is adapted from Hastie et al (1999).

²This summarizes results from LaBine and LaBine (1996).

ex ante. And in a recent literature review, Roese and Vohs (2012, 411) state, ‘When there is a need to understand past events as they were experienced in situ, hindsight bias thwarts sound appraisal.’

Why regard hindsight bias as irrational? First, evidence is sometimes misleading. So the fact that some event occurs does not mean that it wasn’t appropriate to assign it a low probability beforehand. Second, the truth or falsity of a hypothesis does not affect how strongly it is supported by the evidence, and so it seems that to determine the degree to which the evidence supports the hypothesis, we should just look at the evidence itself and consider how much reason it alone gives us for believing the hypothesis.

It is true that evidence can be misleading and that the truth value of a hypothesis does not affect the degree to which it is supported by the evidence. But I will argue that, notwithstanding these points, hindsight bias is often perfectly rational.

The biases and heuristics research program has yielded results that threaten the view of humans as by and large rational. It has also generated significant pushback. Some pushback takes the form of conceding that a given phenomenon constitutes a deviation from ideal rationality, but arguing that it is ‘ecologically rational,’ meaning roughly that it constitutes a favourable adaptation to typical environments (Gigerenzer 2008). But other forms of pushback dispute the claim that the phenomenon really does constitute a violation of ideal rationality (cf. Kelly 2004 on sunk costs). My defence of hindsight bias is of the latter sort.

2 What is Hindsight Bias?

Unfortunately, there is no universally accepted characterisation of hindsight bias.

Definitions vary, and many are vague or otherwise problematic.³ We need to more

³Here is a small sample of characterisations of hindsight bias: Fischhoff (1975, 288) describes it as the phenomenon whereby ‘Reporting an outcome’s occurrence increases its perceived probability of occurrence.’ Bodenhausen (1990, 1113) describes it as follows: ‘Subjects who had been given outcome information judged the described outcome to be more strongly implied by the facts at hand than did subjects who were given no outcome information.’ Hastie et al (1999, 610) say it is the tendency ‘to judge that another person *ex ante* would have made judgments consistent with the *ex post* judgments.’ Harley (2007, 48) defines it as ‘the tendency

precisely characterise our target phenomenon.

Suppose you must judge the degree to which H is supported by the evidence possessed by some agent A (who does not know whether H). Let P be your credence function and $ES_A(H)$ be the degree to which A 's evidence supports H (between 0 and 1, inclusive).

There are then two natural ways of characterising hindsight bias. Importantly, it will not matter for my purposes which one we adopt, for my argument for the rationality of hindsight bias goes through either way. First, we might say that you have hindsight bias just in case your credence that A 's evidence strongly supports H (above some threshold t) is higher conditional on H 's truth than not:

$$P(ES_A(H) > t \mid H) > P(ES_A(H) > t)$$

Second, we might say that you have hindsight bias just in case your expectation of the degree to which A 's evidence supports H is higher conditional on H 's truth than not:

$$\sum_n P(ES_A(H) = n \mid H) \times n > \sum_n P(ES_A(H) = n) \times n$$

There are two independent reasons why hindsight bias, characterised in either of these ways, is often rational. First, the truth of H provides some evidence about what A 's evidence is. Given the assumption that evidence is less likely to be misleading than not, H therefore provides evidence that A 's total evidence (whatever it is) supports H . Second, even if you know exactly what evidence A has, learning the truth of H provides some evidence about the degree to which that evidence supports H . Even if you evaluated the *ex ante* evidence correctly, learning H provides some evidence that if you made a mistake, you are more likely to have erred low than to have erred high in estimating the degree to which that evidence supports H .⁴

to exaggerate the likelihood of a given outcome compared to its foresight predictability.' Finally, Roese and Vohs (2012, 411) describe it as 'the belief that an event is more predictable after it becomes known than it was before it became known.'

⁴Instead of characterising hindsight bias in terms of your credences about how strongly A 's evidence in fact supports H , we might characterise it in terms of your credences about how strongly A (or perhaps an average human) would in fact believe H in light of that evidence. But my arguments will carry over to support the rationality of hindsight bias characterised in this way. After all, it is plausible that people's credences are usually at least roughly proportional

3 Hindsight as Evidence of Evidence

Let's take these points in turn. Start with a case where you know nothing about A, about A's evidence, or about the proposition H . You will no doubt find it difficult to assign credences to the various hypotheses about the degree to which A's evidence supports H . Nevertheless, if you are then told that H is in fact true, you should increase your credence that A's evidence supports H (as well as your expectation of the degree to which it does so).

Why is this? Well, to begin with, if you know H , you can rule out the hypothesis that A's evidence decisively tells against H . After all, truths cannot entail a falsehood, so if evidence can include only truths (Williamson 2000, Littlejohn 2013), then you can rule out A's having evidence which logically entails $\neg H$. **Moreover, it is rational to believe that in the actual world, evidence is not generally misleading, and hence that a randomly selected body of evidence is more likely to support truths than falsehoods.** Note that this is a much more plausible assumption than that the world is highly orderly or predictable; indeed, skepticism looms if one denies it.

So, when you know nothing about what A's evidence actually is, it is clear that learning H should, in general, increase your credence that A's evidence strongly supports H , and also your expectation of the degree to which it does so. (This is not to say that it should always do so; if you were informed that an evil demon was systematically planting misleading evidence, then learning H should decrease your credence that A's evidence supports H .)

Of course, study participants are told a great deal about the evidence available *ex ante*. For instance, in the study which used the example of railroad derailment (Hastie et al 1999), participants were given extensive background materials including expert testimony on potential causes of derailment, a declaration by the National Transportation Safety Board that

to the degree to which their evidence supports a given proposition. People are not typically anti-reliable in judging what their evidence supports. Then, if learning H rationally makes you more confident that A's evidence supports H to some very high degree, it should also make you more confident that A had high credence in H . And if learning H rationally raises your expectation of the degree to which A's evidence supports H , it should also raise your expectation of the credence A had in H .

the track was hazardous, and an appeal of that declaration by the railroad company.

Still, it's important to recognize that third parties rarely have *all* of the evidence possessed by those in the *ex ante* situation. A long-time railroad employee or therapist will have lots of relevant evidence, including first-person observations, that cannot be conveyed through relatively short briefing materials. In light of this, the fact that the event in question occurred still provides some evidence about what further evidence was possessed *ex ante*. Given the further assumptions (i) that evidence consists of truths and hence cannot logically entail falsehoods, and/or (ii) that evidence is less likely to be misleading than not, it follows that learning a hypothesis should increase your credence that the *ex ante* evidence supported that hypothesis (as well as your expectation of the degree to which it did so).

4 Hindsight as Evidence about Evidential Support

Turn now to the second reason why hindsight bias can be rational. Suppose that you possess all of the evidence had by A *ex ante*. It is tempting to think that in this case, learning H wouldn't provide any evidence about the import of that *ex ante* evidence. After all, you could just look at that evidence and tell the degree to which it supports H . But facts about evidential support are not transparent in this way. You are not always in a position to know for certain the degree to which a body of evidence supports a given hypothesis. And if you are uncertain to what degree the *ex ante* evidence supports H , then upon learning H , you should increase your expectation of the degree to which it supports H .

This is most clear in cases involving complex logical relations. Suppose you're given some true premises, and you're told that they either entail H or entail $\neg H$. You try to figure out which it is, but with no success (**suppose the problem is Fields Medal-worthy**), so you wind up uncertain about which one is entailed by the premises. Then, you're told that in fact H is true. Clearly you should conclude that the premises entail H !

We can make the same point (albeit slightly less obviously) in cases where the

evidence and the hypothesis are logically independent. Suppose you are given some *ex ante* evidence that consists of a bunch of raw statistical data concerning the relationship between exposure to some chemical and being diagnosed with a rare cancer. You try to run a multiple regression analysis, but you're not sure about how to write down the formula used for one step (your university statistics course was several years ago!). On one way of proceeding, you get the result that the probability of cancer given exposure to the chemical is 0.999, while on the other way of proceeding, you get the result that the probability of cancer given exposure is only 0.001. Being uncertain about which way of writing down the formula was the right one, you are 0.5 confident that a given person who was exposed to the chemical will get the cancer. If you are then told that the person did indeed get the cancer, it seems you should increase your confidence that the right answer was 0.999.

Consider also garden variety cases of abductive inference. On one standard way of thinking about evidential support, the degree to which a body of evidence supports a hypothesis depends on a variety of theoretical virtues, including the simplicity and naturalness of the hypothesis, the probability it assigns to the evidence, and the degree to which it is potentially explanatory. But these theoretical virtues can conflict. The degree to which a body of evidence supports a hypothesis depends, then, on the correct way of trading off these different theoretical virtues against each other.⁵ Learning which hypotheses are true can then provide some evidence about which way(s) of trading off these theoretical virtues is rational. Suppose, for instance, that you are uncertain whether simplicity should be given more, less, or the same weight as fit with the evidence (given some scale for measuring simplicity and fit). Given some evidence, hypothesis H_1 fares best

⁵My argument does not depend on there being one privileged way of assigning weights to the different theoretical virtues, so long as some weightings are permissible and others impermissible.

when simplicity is given more weight than fit, H_2 fares best when they are given equal weight, and H_3 fares best when simplicity is given less weight than fit. It is plausible that learning H_1 would provide some evidence that simplicity should be given more weight than fit, and hence that the *ex ante* evidence more strongly supported H_1 than H_2 or H_3 .

I have suggested that learning the truth of H typically provides evidence about how strongly a body of evidence supports H . We might refer to this as the epistemic significance of ‘lower-order evidence.’ The obverse case of higher-order evidence has been much discussed. Most theorists think that higher-order evidence (i.e. evidence about what your evidence supports) should often affect your credences in first-order propositions (Feldman 2005, Elga 2007, Christensen 2010). For instance, upon learning that your evidence supports H to a degree n higher than your actual credence in H , you should raise your credence in H . Importantly, it follows that upon learning H , you should raise your credence that that evidence supports H to degree n . After all, positive relevance is a symmetric relation: $P(A | B) > P(A)$ iff $P(B | A) > P(B)$.⁶ So, your credence in H should be higher conditional on the claim that your *ex ante* evidence E supports H to degree n than not iff your credence that E supports H to degree n is higher conditional on H ’s truth than not.⁷ Thus, my claim about the epistemic significance of ‘lower-order evidence’ is supported not only by the cases considered above, but also by appeal to the orthodox position on higher-order evidence.

⁶**Proof:** Suppose $P(A | B) > P(A)$. By the ratio analysis, $\frac{P(A \wedge B)}{P(B)} > P(A)$. Rearranging, we have $\frac{P(A \wedge B)}{P(A)} > P(B)$. By the ratio analysis again, we have $P(B | A) > P(B)$.

⁷**In symbols:** $P(H | ES_E(H) = n) > P(H)$ iff $P(ES_E(H) = n | H) > P(ES_E(H) = n)$, where $ES_E(H) = n$ is the proposition that evidence E supports H to degree n .

5 Hindsight Bias and Ideal Rationality

I claim that hindsight bias is often consistent with ideal rationality. But one might object that my argument in the previous section does not support this strong conclusion. I argued that if you are uncertain about facts about evidential support, then you often ought to display hindsight bias. However, this argument provides no support for the claim that hindsight bias is ideally rational unless uncertainty about the evidential support facts is itself consistent with ideal rationality.⁸ (No such objection can be levelled against my first argument, for it is obviously consistent with ideal rationality that one be uncertain about what exactly the *ex ante* evidence was.)

There is much to be said for this objection. But before addressing it, let me make a dialectical point. Hindsight bias is widely taken to be an embarrassment, suggesting that humans are foolish, or at least less rational than we might have thought. But if some instances of hindsight bias are irrational only because it is irrational to be uncertain about complex relations of evidential support, then hindsight bias is far less troubling. After all, we already know that humans deviate from an ideal which includes, *inter alia*, logical omniscience! Moreover, even if ideal rationality requires omniscience about evidential support relations, hindsight bias would still be an appropriate response to our deviation from this ideal; it would thus be a requirement of ‘non-ideal’ rationality.

What about the objection itself? Would an ideally rational agent be certain of facts about evidential support relations? Here is why you might think the answer is ‘yes.’ First, these facts are *a priori*. This is most obvious for facts about logical entailment. But it is also plausible for fundamental facts about evidential support more generally, including facts about how competing theoretical virtues are to be traded off against each other. Second, it is plausible

⁸Note that if ideal rationality precludes uncertainty about the evidential support faces, then higher order evidence likewise has no epistemic significance for ideally rational agents, for they always know for certain the degree to which a given body of evidence supports a given hypothesis. How to respond to higher-order evidence is then a problem only for non-ideal theory.

that an ideally rational agent would be certain of all *a priori* facts. After all, *a priori* facts are knowable through reason alone, and ideally rational agents are perfect reasoners.

This claim is bolstered by models of rationality that build in logical omniscience, and perhaps *a priori* omniscience generally. In standard Bayesian models, the probability space consists only of propositions that are logically possible. So, if we take such models to characterise at least necessary conditions for ideal rationality, it follows that ideal rationality requires logical omniscience. It is also somewhat natural to use more restrictive models where the probability space consists includes only propositions that cannot be ruled out *a priori* (even if they are logically possible), leading to the conclusion that ideal rationality requires full *a priori* omniscience.

If these two claims are true—that fundamental facts about evidential support are *a priori*, and that ideal rationality requires *a priori* omniscience—it follows that ideal rationality precludes uncertainty about the degree to which a body of evidence supports a given hypothesis.

This is a powerful argument. **But while fully addressing it would go well beyond the scope of this paper, let me say briefly why I think it is mistaken.** First, it is less plausible that ideally rational agents would be omniscient about all *a priori* facts than that they would be omniscient about logic. Ideally rational agents may have infinite computational speed and make no inferential mistakes. But while this may suffice for them to know all logical truths, it is not clearly sufficient for them to know *a priori* facts more generally. There may be *a priori* facts whose truth does not follow deductively from obvious premises. Plausibly, fundamental facts about ethics are like this, and facts about abductive evidential support (e.g., about how to trade off competing theoretical virtues against each other) may be as well. Being perfect at logic will not suffice to arrive at their truth.

Second, there are good reasons for thinking that ideally rational agents will not be certain of even some logical truths. Ideally rational agents, by stipulation, make no mistakes in logical reasoning. But Christensen (2007) argues that even if such an agent makes no mistakes in proving some complex logical truth *T*, she should not be certain that she made no mistakes.

She should have at least some credence that she made an error somewhere. Ideally rational agents should not be certain of their own ideality.

To see this, consider Christensen's case in which an ideally rational agent comes up with a genuine proof of T but is then told that she was given a reason-distorting drug in her coffee which affects 99% of those given the drug, with 1% of the population immune. She is told that those affected don't notice any cognitive effects, but it causes them to make subtle logical mistakes. In such a case, it would be unreasonable for the ideally rational agent to conclude that she must be one of the 1% who are immune. And if it would be unreasonable to be certain that she was immune and made no mistakes in arriving at her proof of T , it seems unreasonable for her to be certain of T itself.

There is of course much more to say here. I do not claim to have a knock-down argument that uncertainty about evidential support relations can be ideally rational. If it can, then hindsight bias likewise can be ideally rational even when you know exactly what the *ex ante* evidence is. If not, then hindsight bias is still at least an appropriate response to this antecedent violation of the rational ideal.

6 Conclusion

I have argued against the consensus that hindsight bias is irrational. The truth of a hypothesis often provides evidence about what the evidence available *ex ante* was, and also about what that *ex ante* evidence supports. So often, upon learning that the hypothesis is true, you should become more confident that the *ex ante* evidence strongly supports that hypothesis and also increase your expectation of the degree to which it does so.

My defence of hindsight bias is partial. I do not claim that it is always rational. One might err by going overboard and shifting one's credences about the import of the *ex ante* evidence more than is warranted, or by basing that shift not on

the evidential considerations emphasised above, but on evidentially irrelevant motivational factors like the need for closure or the preservation of one's self-esteem (Roese and Vohs 2012, 415-6).

Nonetheless, if hindsight bias is often perfectly rational, this has important practical upshots. First, scholars have been concerned that hindsight bias has harmful effects, especially in tort cases in which jurors must determine whether the defendant was negligent or otherwise liable, based on an evaluation of the import of the evidence available *ex ante*. As a result, researchers have studied various techniques for 'debiasing,' such as instructing subjects to 'consider the opposite,' that is, to think of reasons why it might have been rational to expect the opposite of what actually happened (Koriat et al 1980). And legal scholars have considered proposals to mitigate the effects of hindsight bias, in particular by blinding jurors to the facts of the outcome as much as possible, or even taking power out of the hands of jurors and having judges or experts determine liability and damages (Hastie et al 1999; Harley 2007). But if hindsight bias often yields a more rational, and hence presumably more accurate, assessment of the significance of the *ex ante* evidence, then these debiasing and mitigation efforts may be a step in the wrong direction.

Second, if I am right about the epistemic significance of lower order evidence, this suggests that hindsight can play an important role in our evaluation of competing theories about evidential support. Track record matters. It is not the case that the truth value of a hypothesis affects how well it is supported by some body of evidence. I take facts about fundamental evidential support relations be necessary. Nonetheless, for those of us who are uncertain of the facts about evidential support (whether or not such uncertainty is consistent with ideal rationality), learning the truth of a hypothesis can provide evidence about those facts which can then guide us going forward.

References

- Bodenhause, G. 1990. 'Second-Guessing the Jury: Stereotypic and Hindsight Biases in Perceptions of Court Cases.' *Journal of Applied Social Psychology* 20 (13): 1112-21.
- Christensen, D. 2007. 'Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals.' *Oxford Studies in Epistemology* Vol. 2, pp. 3-31.
- Christensen, D. 2010. 'Higher Order Evidence.' *Philosophy and Phenomenological Research* 81 (1):185-215.
- Elga, A. 2007. 'Reflection and Disagreement.' *Noûs* 41 (3):478-502.
- Feldman, R. 2005. 'Respecting the Evidence.' *Philosophical Perspectives* 19 (1): 95-119.
- Fischhoff, B. 1975. 'Hindsight \neq Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty.' *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.
- Gigerenzer, G. 2008. *Rationality for Mortals*. New York: Oxford University Press.
- Harley, E. 2007. 'Hindsight Bias in Legal Decision Making.' *Social Cognition* 25 (1): 48-63.
- Hastie, R., Schkade, D., and Payne, J. 1999. 'Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages.' *Law and Human Behavior*, 23 (5): 597-614.
- Kelly, T. 2004. 'Sunk Costs, Rationality, and Acting for the Sake of the Past.' *Noûs* 38 (1): 60-85.
- Kelly, T. 2010. 'Peer Disagreement and Higher Order Evidence.' In Alvin I. Goldman and Dennis Whitcomb (eds.), *Social Epistemology: Essential Readings*. Oxford University Press. pp. 183-217.
- Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). Reasons for confidence. *Journal of*

Experimental Psychology: Human Learning and Memory, 6, 107118.

LaBine, S. and LaBine, G. 1996. 'Determinations of Negligence and the Hindsight Bias.' *Law and Human Behavior* 20 (5): 501-516.

Littlejohn, C. 2013. 'No Evidence is False.' *Acta Analytica* 28 (2):145-159.

Roese, N. and Vohs, K. 2012. 'Hindsight Bias.' *Perspectives on Psychological Science* 7(5) 411-426.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.