



Assessing the Evidence for Outcome Bias and Hindsight Bias

Mikkel Gerken¹

Accepted: 5 January 2023 / Published online: 31 January 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Outcome bias and hindsight bias are important in philosophical debates and have wide-ranging implications outside of philosophy. Recently, Hedden has articulated a novel line of argument that the empirical evidence for what he labels hindsight bias is largely misguided and that empirical researchers who postulate such a bias are engaged in a *fallacy fallacy*. In this paper, I articulate Hedden's core insights in terms of two principles and argue that in the relevant empirical research, these principles are often (i) recognized, (ii) invoked in theoretical explanation, and (iii) guiding study designs and interpretation of data. On this basis, I argue that there is ample empirical evidence for outcome and hindsight bias and that cognitive psychologists do not *generally* exhibit a fallacy fallacy of postulating such biases in cases where there are none. More constructively, the present paper articulates some general principles that help us to better understand when an empirical study provides evidence of these types of bias. Thus, the debate casts light on both the specific evidence for outcome and hindsight bias but also on the general division of labor between philosophy and psychology.

1 The Influence of Outcome

Outcome and hindsight bias are commonly regarded as ubiquitous biases which pertain to the influence of outcome on human cognition. Before setting sail, some conceptual clarification is required since the labels 'outcome bias' and 'hindsight bias' are both used in different senses in philosophy and psychology. Even within psychology, there are terminological variances (Blank et al. 2008).¹ Perhaps the most common use of 'hindsight bias' in psychology is a *narrow* sense that concerns the

¹ See also Fischhoff 1975, 2007; Hawkins and Hastie 1990; Christensen-Szalanski and Willham 1991; Mackie et al. 2001; Guilbault et al. 2004; Blank et al. 2007; Pohl 2007; Roesse and Vohs 2012. And, as mentioned, 'bias' is sometimes used descriptively to refer to outcome effects or outcome-influenced judgments.

✉ Mikkel Gerken
mikkel@sdu.dk

¹ University of Southern Denmark, Campusvej 55, 5230, Odense, Denmark

assessment of one's own past belief, credence, or epistemic situation. More specifically, this narrow sense of 'hindsight bias' denotes the phenomenon of (unconsciously) adjusting one's memory of one's previous estimates in the light of outcome information: "hindsight bias does not refer to all retrospective increases in the probabilities assigned to events. The hindsight bias is a projection of new knowledge into the past accompanied by a denial that the outcome information has influenced judgment." (Hawkins and Hastie 1990: 311. See also Fischhoff and Beyth 1975; Bradfield and Wells 2005; Blank et al. 2008). Thus, there is a metacognitive dimension to hindsight bias – sometimes called the *knew-it-all-along* effect. In contrast, outcome bias concerns the evaluation of an individual or procedure responsible for an outcome or the *ex ante* evidence for the outcome (Pezzo 2011). The evaluation is said to manifest outcome bias when the outcome information is given excessive epistemic weight.² So, outcome and hindsight bias share the feature that they concern the influence of outcome information on human cognition although they are different in other regards.³ Given the important shared feature between outcome and hindsight bias, I will discuss both but focus on outcome bias, which is often primary in the sense that hindsight bias consists in a subject's ignorance of his own outcome bias. That is, many cases in the empirical literature concern a judgment which is unduly influenced by outcome information (outcome bias) but in which the subject misrepresents the judgment as unaffected by the outcome information (hindsight bias).

Another aspect of outcome bias and hindsight bias that calls for some conceptual clarification is the relevant notion of *bias*. The term 'bias' is also polysemous across philosophical and psychological debates as well as within them. In consequence, conceptual equivocation is a source of dialectical and substantive complications (Gerken 2011a, 2013a). One use of 'bias' is a merely descriptive sense that indicates that a pattern of judgments is influenced by some factor but without any judgment on whether this influence is irrational. For example, one might say that judgments about the evidence for a hypothesis or the rationality of an individual are biased in the descriptive sense that they are influenced by information about an outcome such as the hypothesis' truth-value. The other, and perhaps more prominent, sense of 'bias' is normative in that it involves the idea that such judgments are suboptimal in some sense – for example, because they systematically deviate from ideal rationality. As a terminological guard against equivocation, I restrict my use of 'bias' to the normative sense. In this sense, biased judgments are, minimally, systematically deviating from *ideal* rationality. However, they may be *boundedly* rational – i.e., rational

² Edwards 1984; Baron and Hershey 1988; Stanovich and West 2008. Evidence of outcome bias has been found in diverse fields such as medical sciences (Berlin 2004), management studies (Marshall and Mowen 1993; Sezer et al. 2016), finance and business (Fisher and Selling 1993; Tan and Lipe 1997), legal studies (Clarkson et al. 2002), social cognition (Mackie et al. 2001; Mazzocco et al. 2004). Moreover, I have argued that some empirical findings (Turri 2016, 2017) are best understood as instances of epistemic outcome bias (Gerken 2017, 2020). If so, the empirical motivation for factive norms of action and assertion is as dubious as the philosophical motivation (Gerken 2011b, 2012).

³ For example, one could correctly adjust one's credence in the light of outcome information but misrepresent one's adjustment of credence. This would be hindsight bias without outcome bias. Conversely, one could overestimate the epistemic weight of outcome information without misrepresenting one's adjustment of it. This would be outcome bias without hindsight bias. So, while closely related, hindsight and outcome bias can be doubly dissociated.

given the cognitive capacity-limitations of the individual. In contrast, I will use the phrase ‘outcome-influenced judgments’ descriptively to denote the judgments that are influenced by outcome information.

Given this conceptual and terminological clarification, we can say that the most prominent view in both philosophical and psychological discussions of outcome-influenced judgments is that many of them manifest a bias. However, this view has recently been challenged by a novel line of argument (Hedden 2019). Hedden’s playful title – *Hindsight bias is not a bias* – plays on two noted senses of ‘bias’ where the first occurrence is descriptive, and the second occurrence is normative. Thus, the gist of Hedden’s title is that judgements that are influenced by outcome information are not manifesting a bias in the normative sense. That is, when Hedden defines what he calls ‘hindsight bias’ (Hedden 2019: 45), he is, in the present terminology, defining outcome-influenced judgments to argue that they are not biased in the normative sense. Another terminological issue is Hedden’s use of the term ‘*hindsight bias*.’ Both Hedden’s informal and formal definitions concern the updating of belief or credence given information on outcome. That is, the noted metacognitive dimension of hindsight bias (the *knew-it-all-along* effect) does not figure in Hedden’s informal or formal definitions or in his critical arguments. Rather, both Hedden’s definitions and substantive arguments concern the somewhat broader phenomenon which is more commonly labelled ‘*outcome bias*’ (Pezzo 2011). Of course, this terminological variance does not lessen the importance of Hedden’s substantive arguments. But since his paper aims to bring the formal and epistemological considerations in closer contact with the empirical research, I will seek to reflect the most established terminology. Consequently, I will primarily cast the discussion in terms of *outcome bias* which aligns with Hedden’s arguments that seek to rationalize the outcome-influenced judgments about some individual or procedure rather than the participants’ judgments about their own original (pre-outcome information) judgments. However, I will also draw on the empirical research on hindsight bias (in the narrow metacognitive sense) and consider whether this research is compromised by Hedden’s critical considerations regarding the epistemic value of outcome information.

Hedden’s challenge is part of a more general recent trend in philosophy that consists in applying epistemological and formal tools to rationalize patterns of judgment that psychologists typically take to manifest a cognitive bias. This type of criticism differs from existing criticisms of views that regard patterns of human judgment as biased. For example, Gigerenzer famously criticizes the postulation of bias in human cognition from several angles. For example, he provides empirical evidence of cases in which fast and frugal heuristics from an adaptive toolbox outperform more reflective types of deliberation (Gigerenzer 2008 *et passim*). Furthermore, Gigerenzer resists appeals to various notions of ideal rationality as significant for what he calls “rationality for mortals” (Gigerenzer 2008). Specifically, he resists that deviance from an ideal “gold standard” pattern of judgment is a good basis for postulating a cognitive bias for bounded agents. This is in part because he argues that cognitive responses should be evaluated in terms of their performance in real-life environments rather than in terms of their approximation to an optimal response relative to some ideal of rationality (Gigerenzer 2008 see, e.g., 8 and 80–82). The adaptive toolbox framework does not entail a wholesale rejection of bias in human cognition.

In the case of hindsight bias, Gigerenzer's nuanced proposal emerges in joint work with Hoffrage and Hertwig (Hoffrage et al. 2000).⁴ More generally, Gigerenzer's critical stance toward ideal rationality differs from Hedden's arguments, which rest on the idea that the phenomena in question – i.e., the participants' response patterns – manifest (or approximate) ideal rationality. Consequently, Hedden's arguments form a novel challenge (but see Young et al. 2010).

By considering outcome-influenced judgments as an important instance of the more general debate about the evidence for postulating bias, I will argue for a two-fold methodological conclusion that concerns the division of labor between philosophy and the empirical sciences. The first conclusion is that epistemological and formal tools can provide vital insights that may constrain or even correct empirical research on cognitive biases. The second conclusion is that assessments about whether a pattern of human judgment manifests a bias must engage closely with empirical research on the matter. Specifically, I will argue that if both Hedden's structural insights and the empirical research are considered, it continues to be reasonable to postulate outcome and hindsight bias. In this manner, my aims are both methodological and substantive.

In sum, I will argue that empirical research provides evidence for outcome and hindsight bias and that cognitive psychologists are, therefore, not generally committing a fallacy fallacy. More constructively, I will articulate some principles concerning the relationship between *ex ante* evidence and outcome information. These principles constrain study designs and enable us to better adjudicate whether outcome-influenced judgments are biased. Finally, I will draw some methodological lessons about how philosophical theorizing may contribute to empirical work on biases.

2 Hedden's Arguments that Outcome-Influenced Judgments are not Biased

Hedden considers an epistemic variety of outcome-influenced judgments that does not concern judgments about probability but rather judgments about a participant's, S's, credence concerning how well someone's, A's, evidence supports a hypothesis, *H* (Hedden 2019: 45). On the basis of two formally precise characterizations of this phenomenon, he provides two related strands of argument that outcome-influenced judgments are unbiased since they align with norms of ideal epistemic rationality.

⁴ According to this account, hindsight bias is the product of a knowledge-updating process that is rational in Gigerenzer's preferred sense of being ecologically valid. Hoffrage et al. use 'hindsight bias' in a descriptive sense and reject that it is a bias in the normative sense because it is "a consequence of learning by feedback" (Hoffrage et al.: 579). In contrast, psychologists less skeptical of postulating bias will not take the fact that systematically erroneous misjudgments are the byproduct of an ecologically valid or boundedly rational process as a reason against classifying them as biased. This grand debate concerning the nature of bias is central to the so-called *rationality wars* and unlikely to be settled anytime soon (see Samuels et al. 2012; Wallin 2013). Gigerenzer's view is discussed *et passim* in Gigerenzer 2008. But, for the present purpose, the key thing to note is that Gigerenzer's general framework and the specific account in Hoffrage et al. differ from any claim that systematically erroneous recall of one's past assessment is ideally rational.

First, Hedden argues that although outcome information is not first-order evidence regarding the hypothesis H , it is often higher-order evidence concerning A 's evidence for H . For example, Hedden suggests that “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” (Hedden 2019: 46). Hedden argues that, in such cases, the outcome-information, H , rationalizes an outcome-influenced judgment that A has strong evidence for H . Indeed, according to Hedden, such a judgment would be ideally rational, and, derivatively, a judgment that A is more rational than someone in a bad outcome (non- H) condition is also ideally rational *pace* those who postulate that such judgements exemplify a bias.

However, as Hedden recognizes, one may be aware that misleading evidence regarding H is prevalent. If I learn that a conspiracy, H , is true, I cannot infer that A 's evidence supports H . Likewise, I may be unaware whether A has any of the evidence regarding H . If I learn that it rains in Muscat, it should hardly strengthen my credence that the evidence of someone living in Managua strongly supports the hypothesis that it rains in Muscat. So, it is only in some cases that the H rationalizes the judgment that A 's evidence strongly supports H . It is presumably for this reason that Hedden only claims that outcome information is *fallibly* good higher-order evidence.

Hedden's second rationale for the ideal rationality of outcome-influenced judgments is that outcome information may serve as evidence about evidential support. Hedden notes that “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 ” (Hedden 2019: 47). Some of Hedden's cases are ingenious but fairly exotic in that they involve scenarios in which one knows that the evidence either entails or strongly supports exactly one of H or $\neg H$ but not which one. But other cases involve an abductive inference where the uncertainty concerns how one should weigh conflicting epistemic virtues, such as H 's simplicity versus H 's fit with the evidence. In such cases, Hedden suggests that outcome information provides evidence about the rational trade-off between those theoretical virtues and, hence, about how well the *ex ante* evidence supports H . Again, the suggested upshot is that the outcome information, H , makes it ideally rational to make fairly strong judgments about the strength of the *ex ante* evidence and derivative judgments about those who rely on it. If so, the outcome influence found in experimental work is not indicative of a bias but of ideally rational responses to evidence about evidential support. Importantly, however, the epistemic weight of outcome information tends to depend on the degree of uncertainty about the *ex ante* evidence's support. To illustrate this with a simplified case, assume that H and H^* are competing hypotheses in a domain, D , and that H is simple although it does not fit the evidence as well as the more complex H^* . Assume, furthermore that there is very little uncertainty that it is more truth-conducive to accept the hypothesis that fits the evidence than it is to accept the simple hypothesis in the domain D . In this case, learning the truth of H should not lead us to assume that the *ex ante* evidence supports H over H^* . Note that I am not suggesting that Hedden would reject this. My point is simply to illustrate that unless there is significant uncertainty about the *ex ante* evidence, the outcome information does not tend to provide strong evidence about the strength of the *ex ante* evidence.

Thus, it is an important feature of Hedden's cases in which outcome information is highly relevant to assessing the *ex ante* evidence that these are cases in which there is significant uncertainty about the *ex ante* evidence. This is an important point but, as I will argue, one that the study designs used to generate evidence for postulating outcome and hindsight bias are sensitive to.

That noted, Hedden convincingly argues that outcome information often, but not infallibly or invariably, serves as higher-order evidence about the evaluatee's *ex ante* evidence or provides evidence of evidential support. Moreover, Hedden's two strands of argument are not mere disjoint ones. Rather, they converge on a general unified claim that may be expressed as follows:

Frequent Rationality

In many cases, outcome-influenced judgments are ideally rational.

As I have argued, Hedden's cases also provide a valuable lesson about the relationship between uncertainty of *ex ante* evidence and outcome information. I will articulate it as the following principle:

Epistemic Weight

Outcome information tends to have higher epistemic weight when there is a high degree of uncertainty about the *ex ante* evidence.

The principle *Epistemic Weight* is not a universal law of epistemology but a *ceteris paribus* generalization motivated by reflection on cases such as Hedden's.⁵ I use the broad phrase 'uncertainty about the evidence' so that the principle captures both cases in which the first-order *ex ante* evidence is weak and cases in which the first-order *ex ante* evidence is strong but associated with peculiar higher-order uncertainty such as uncertainty as to whether it strongly supports *H* or its negation.

I regard both *Frequent Rationality* and *Epistemic Weight* as important principled insights. But I will argue that these theses do not provide reason to think that outcome and hindsight bias are not biases or that cognitive psychologists are generally engaged in a fallacy fallacy.

3 The Status of *Frequent Rationality* and *Epistemic Weight* in Empirical Research

Importantly, *Frequent Rationality* and *Epistemic Weight* are not foreign to the cognitive psychologists who conduct research on hindsight and outcome bias. On the contrary, these theses are integral to the research program. Sometimes this is fairly explicit but, naturally, cognitive scientists are more concerned with specific

⁵ It is not hard to find supplementary cases. For example, some types of outcome information may make it reasonable to suppose that the evaluatee, *A*, has more evidence than the description explicitly mentions or that *A* has assessed it better than someone in a negative outcome condition.

experiments than philosophical discussions. Consequently, the commitment to the theses must often be seen in the study designs, interpretations of data, debiasing manipulations etc.

However, it is reasonably clear that empirical researchers on outcome and hindsight bias recognize that outcome information has epistemic weight. In particular, outcome bias is only postulated when participants are taken to give *too much* weight to outcome information. For example, Sezer and colleagues characterize outcome bias as “paying too much attention to the outcomes of others’ actions while neglecting information about the original intentions leading to those outcomes” (Sezer et al. 2016). Sezer and colleagues accurately summarize the literature on outcome bias as concluding “that individuals overweight outcome favorability when making evaluations” (Sezer et al. 2016: 13). Similarly, Baron and Hershey remark: “While it is not in general unreasonable to use outcome as information bearing on the assessment of the basis of the outcome, this information may be overused” (Baron and Hershey 1988: 570).

Epistemic Weight has also been recognized: “A judge who does not know the decision maker’s probabilities may assume that the probability was higher for an outcome that occurred than for the same outcome had it not occurred... ..In the extreme, if we have no information except outcome, it is a reasonable prima facie hypothesis that bad outcomes... ..result from badly made decisions” (Baron and Hershey 1988: 569–570). This point is very similar to Hedden’s (Hedden 2019: 46). So, apart from aligning with *Frequent Rationality*, the quote indicates appreciation of *Epistemic Weight*. Thus, the initial reason for taking the outcome-influenced judgments to be biased is that they clash with considered judgments about how much weight the outcome information should have. Indeed, it is hard to see how a single evidence point of outcome information can rationalize the judgment that a procedure with 2% risk of mortality is superior to a procedure with 8% risk of mortality (Stanovich and West 2008). The *additional* reasons in a larger abductive case for postulating a bias should be evaluated against this basis.

Importantly, the recognition of *Frequent Rationality* is not a detached, dispensable presupposition in the empirical work on outcome and hindsight bias. Rather, it is used to characterize the nature and causes of these biases. For example, Mackie et al. take the source of bias to be due to “...the heuristic property of allowing inferences to be drawn even with a minimum of cognitive investment” (Mackie et al. 2001: 72). Others note that they “...expected to find outcome bias because the generally useful heuristic of evaluating decisions according to their outcome may be overgeneralized to situations in which it is inappropriate” (Baron and Hershey 1988: 571). Such remarks do not merely acknowledge *Frequent Rationality*; they invoke it as a key part of a principled explanation of why outcome-influenced judgements may be biased. While the precise explanation depends on the experimental design and the specific type of bias under investigation, the following *explanatory scheme* is a standard one: the reason for ascribing a bias is that participants *overestimate* the epistemic weight of outcome information in part because it is salient and fluently processed in a cognitively cheap manner. Thus, reliance on outcome information is generally a boundedly rational practice of deploying a cognitively low-cost, reasonably reliable heuristic.

However, in specific conditions, this heuristic is systematically unreliable since it gives too much weight to outcome information. Consequently, the outcome-influenced judgments are biased although they may be boundedly rational insofar as the outcome heuristic that produces them is ecologically valid.

In general, it is an additional reason in favor of a claim that a pattern of judgment is biased that its existence can be explained as ecologically valid for bounded agents. In general, biases are postulated as the consequences of cognitive heuristics that are boundedly rational. Often the bounded rationality is explained in terms of the heuristics being reasonably reliable in the environments in which they are evolved to operate in and, therefore, systematically unreliable in other environments. What the empirical researchers quoted above point out is that outcome bias fits this model extremely well. While this does not establish that it is a bias, it is a significant additional reason in the larger abductive case for this conclusion.

This explanatory scheme is reinforced in explanations of outcome and hindsight bias that integrate them with accounts of related biases (Hawkins and Hastie 1990: 319). For example, independent evidence indicates that we give excessive weight to information that is salient or accessible in the sense that it is fluidly processed (Harley et al. 2004; Oppenheimer 2008; Stanovich 2011). So, in conditions in which outcome information is in focus, it is likely to be given more weight than its proper epistemic weight (Gerken 2013b, 2017).

So, a third but related reason to take the relevant outcome-influenced judgments to be biased is that they may be explained in terms of or integrated with other biases or well-established assumptions about bounded cognition. An example is the assumption that cognition tends to privilege focal information or easily available information in a manner that may give it excessive weight. A different but related rationale for postulating bias derives from the fact that both outcome bias and hindsight bias have been found to correlate with cognitive ability (Stanovich and West 2008 summarized in Stanovich 2012 Table 7.2). Correlation with cognitive ability is not a necessary condition on a response pattern being biased. Many biases do not correlate with cognitive ability. But in cases where debiasing manipulations are found to be effective, correlation with cognitive ability provides a further reason to postulate a bias (Stanovich 2012: 149).

A fourth reason for upholding the bias account of outcome effects comes from experimental manipulations that are designed to diminish the effect of outcome information. For example, if participants in conditions that trigger more reflective processing do not exhibit the same degree of outcome sensitivity, this augments the reason for taking their heuristically based judgment to be biased. Of course, such debiasing manipulations are not akin to “crucial experiments” that independently and conclusively establish that a pattern of judgment manifests a cognitive bias. Rather, they are independent considerations in a broader abductive justification for postulating a bias. If one has reason to postulate a bias, it is an additional point in favor of the account if the relevant patterns of judgments may be debiased by standard debiasing manipulations. The accounts that postulate outcome and hindsight bias pass this test, and this augments the bias account.

A wide variety of experimental manipulations have been shown to minimize or eliminate outcome-influenced judgments. For example, it has consistently been

found that when participants are forced to consider both reasons for and reasons against their chosen answer or a range of alternative outcomes, the effect of outcome information tends to decrease (Koriat et al. 1980; Arkes et al. 1988; Lowe and Reckers 1994). This way of taking a bias of an intuitive judgment to be reduced by triggering more reflective processing or processing that is more likely to include relevant information is a staple in work on cognitive biases (Evans 2010; Stanovich 2011). Furthermore, a multitude of other debiasing manipulations have yielded significantly smaller effect sizes compared to the standard experimental paradigms (Wood 1978; Hawkins and Hastie 1990; Clarkson et al. 2002; Guilbault et al. 2004; Pezzo 2011; Bernstein et al 2016; Sezer et al. 2016).

Another type of finding which supports this pattern is that the effect of outcome information is often stronger in between-subject study designs than in within-subject study designs, and this suggests that a more reflective mode of processing minimizes the bias (Kneer and Machery 2019). Kneer and Machery went one step further and devised a “contrastive” study design in which participants compared agents in good and bad outcome scenarios by assessing comparative probes, such as “It was just as wrong for Anna to leave her son as it was for Beth” (Kneer and Machery 2019: 336). For a range of deontic notions, the contrastive manipulation eliminated the outcome effect found in between-subject study designs. So, the general finding is that in low comparison studies, such as between-subject studies, there is a strong effect of outcome information. However, this effect is minimized or eliminated in moderate or high comparison studies, such as within-subject or contrastive study designs, which prompt more reflective processing (Kneer and Machery 2019). These findings support the idea that the impact of outcome is the product of a generally useful heuristic processing that overestimates the epistemic weight of outcome information in some conditions, although this faulty reasoning may be corrected in more reflective conditions. In recent work on outcome effects in legal contexts, Kneer and Skoczen found three manipulations to reduce, and in some cases eliminate, the outcome effect sizes. The first is *probability stabilizing* via expert testimony, the second is *anchoring* by consulting participants on the probabilities prior to revealing the outcome information, and the third is *counterfactual priming* by having participants consider alternative outcomes (Kneer and Skoczen forthcoming). Meta-analyses found that counterfactual priming and probability stabilizing were especially effective debiasing manipulations (Kneer and Skoczen forthcoming). As with other debiasing manipulations, these findings are not crucial experiments that *establish* the bias account but reasons that contribute to an abductive case for it. To reject this abductive case, it is important to engage with the full set of empirical research that it is based on and not just a single study design.

The ability to debias is a natural rationale for postulating a bias in the standard experiment. In contrast, it is unclear how someone who takes the judgments in the original experiment to be ideally rational can account for the decreases in effect size in conditions that are designed to render the participants’ cognitive processes more reliable. Perhaps, it may be argued that the allegedly debiasing manipulations had the opposite effect of rendering unbiased judgments biased. As noted, there are cases in which primitive heuristics outperform reflective cognition (Gigerenzer

2008).⁶ But arguing that this is *generally* the case in the relevant empirical work would require detailed assessment of specific studies. To reject the general explanatory schema used in research on hindsight and outcome bias would be to reject core assumptions about cognitive biases and how to research them.⁷

Finally, *Frequent Rationality* and *Epistemic Weight* are not inert theoretical presuppositions of the empirical research. Rather, they guide study designs and constrain interpretations of findings. For example, Hawkins and Hastie note that “A tricky set of methodological problems is raised by the fact that the hindsight bias is difficult to distinguish from simple, “honest” learning from experience” (Hawkins and Hastie 1990: 311). Likewise, Baron and Hershey are explicit that “This sort of bias is not established by showing that people take outcomes into account. As we argued earlier, outcomes are relevant when they can inform us about actor information. One way to show an outcome bias is to give the judge all relevant information about outcome probabilities known to the decision maker, plus the outcome.” (Baron and Hershey 1988: 570). The key point is that researchers on outcome and hindsight bias recognize the evidential value of outcome information and design their studies accordingly. In accordance with *Epistemic Weight*, researchers seek to minimize uncertainty about the evidence that the evaluatee possesses. This may be done by giving participants specific descriptions of the *ex ante* evidence or explicating its strength numerically. It is important to note that more evidence can sometimes increase uncertainty about whether the evidence supports a hypothesis. A clear case of this is one in which an increased pool of evidence includes more diverging evidence than the original smaller pool of evidence. However, the study designs that are usually invoked in experiments that explore outcome and hindsight bias tend to give full information by a simple description of the strength of non-diverging *ex ante* evidence – for example, in numerical terms. For example, in one study concerning heart surgery, participants in the positive outcome condition were told that the medical procedure involved an 8% risk of death compared to a 2% risk in the negative outcome condition. Nevertheless “the decision with the positive outcome was rated as a better decision than the decision with the negative outcome, despite the fact that the latter was objectively better.” (Stanovich and West 2008: 675).

Generally, the case for outcome and hindsight bias rests on study designs that put the participants in very different positions vis-à-vis the *ex ante* evidence than the ingenious cases that Hedden sets forth. Specifically, the study designs generally reflect *Epistemic Weight*, by minimizing uncertainty about the *ex ante* evidence available to the evaluatee. For example, I do not know of a single study design that resembles Hedden’s case in which S knows that A’s evidence entails or strongly supports either *H* or $\neg H$, but not which one. Moreover, the studies in question do not tend to feature abductive inferences involving trade-offs between theoretical virtues

⁶ It is notable that even Hoffrage et al., who are skeptical of casting outcome effects as a bias, have pursued ways to reduce the effect of outcome information and have found several such manipulations to be effective (Hoffrage et al. 2000: 571ff).

⁷ As noted, such questions are worth raising, but I take Hedden’s arguments to be situated within this broad tradition.

such as simplicity and fit with evidence. To his credit, Hedden recognizes that in the relevant experiments “study participants are told a great deal about the evidence available *ex ante*.” (Hedden 2019: 46). But he continues, “third parties rarely have all of the evidence possessed by those in the *ex ante* situation” and concludes that “the fact that the event in question occurred still provides some evidence about what further evidence was possessed *ex ante*” (Hedden 2019: 46).

However, since research on hindsight bias is often concerned with measuring how well participants represent *their own* past judgments or hypothetical judgments, the evaluatee is often identical to the evaluator and, hence, presented with identical evidence. Even when the evaluatee is an imagined protagonist in a case, participants are often simply informed what evidence this protagonist has. This makes for a stark contrast with the cases involving a high degree of uncertainty about A’s evidence that Hedden reasonably take to rationalize outcome-influenced judgments. The claim here is not that identity between the evaluatee and the evaluator guarantees low uncertainty about the *ex ante* evidence but only that it generally does so in the operative study designs. For example, I may forget my past evidence about some outcome, and learning the outcome may then teach me a lot about what the past evidence supported. But, again, this is not the structure of typical study designs in which participants simultaneously get the *ex ante* evidence and outcome information and then make a judgment. In such cases, the evaluator/evaluatee identity contributes to minimizing uncertainty about what the evaluatee’s *ex ante* evidence supports. In general, experiments on outcome and hindsight bias research tend to provide the participants with fairly precise information about the *ex ante* evidence. This, in turn, tends to render the epistemic weight of the outcome information rather minimal. So, methodologically speaking, such study designs indicate sensitivity to *Epistemic Weight*.

In sum, in the empirical research on outcome and hindsight bias, the principles motivated by Hedden’s discussion – i.e., *Frequent Rationality* and *Epistemic Weight* – are (i) recognized, (ii) invoked in theoretical explanation, and (iii) guiding study designs and interpretation of data.

4 Response to Hedden’s Arguments and Related Arguments

It is important to reemphasize that the case for postulating outcome bias and hindsight bias is abductive. As I have outlined above, there are several distinct strands of converging reasons for the bias account. First, because the study designs seek to minimize the epistemic weight of the outcome information, it is implausible to assume that its weight rationalizes the effect sizes. This is particularly so in cases in which the *ex ante* reliability of the bad outcome procedure is significantly lower than the *ex ante* reliability of the good outcome procedure (e.g., Stanovich and West 2008). Second, the presence of outcome-influenced judgments fits into an explanatory scheme for postulating biases that appeals to ecological validity and bounded rationality. Third, the bias account is augmented by the fact that the effect sizes can be minimized by way of a wide variety of stock debiasing manipulations and that

it correlates with cognitive ability. Fourth, the bias account fits with other bounded cognition accounts that appeal to salience, availability, and focal bias.

While these points do not establish the ubiquity of outcome or hindsight bias, they jointly provide a strong abductive case that such biases are present in human cognition. To rebut this case, it is not enough to identify cases in which outcome-influenced judgments may be ideally rational. Rather, it must also be argued that such cases play a central role in the abductive motivation for postulating outcome and hindsight bias. It is an important and largely a priori task to identify informational structures in which outcome-influenced judgments are ideally rational. But the assessment of whether the abductive empirically motivated account is at fault in virtue of hinging on such informational structures is an a posteriori task that requires close engagement with the relevant empirical research.

In responding to Hedden's arguments, it is important to get clear on the conclusions he draws. The title "Hindsight bias is not a bias" and suggestion that cognitive psychologists who postulate it are committing a "fallacy fallacy" suggests a strong conclusion. But subsequently Hedden provides the more qualified conclusion: "I do not claim that hindsight bias 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..." (Hedden 2019: 50).

This qualified conclusion is puzzling given the noted point that cognitive psychologists only postulate outcome and hindsight bias when participants "go overboard" and give the outcome information *too much* epistemic weight. As highlighted, studies supporting this claim are deliberately designed to capture the *specific conditions* in which outcome information is given too much weight. Recall that the presence of bias may be explained by taking it to be the product of a generally reasonable heuristic that is relied on in specific circumstances where it leads to systematically unreliable judgments.

Consequently, Hedden's qualified conclusion appears to be entirely compatible with the central claims in the empirical literatures on both outcome and hindsight bias. Of course, one might argue that in specific studies bias is mistakenly postulated because the experimenters underestimate the epistemic weight of the outcome information. Let us call such putative cases '*underestimation cases*.' To argue that a case is an underestimation case requires detailed discussion of its study design and stimulus materials. I do not deny that the literature on outcome and hindsight bias contains underestimation cases. But since cognitive psychologists recognize the epistemic weight of outcome information and design studies accordingly, a charge that underestimation cases are the norm in the empirical literature is unsubstantiated. Thus, it would be mistaken to suggest that the existing body of empirical evidence does not provide evidence of outcome and hindsight bias. Likewise, it would be mistaken to claim that empirical researchers are generally prone to a fallacy fallacy. However, Hedden is not merely arguing against specific studies. In fact, his arguments engage minimally with specific study designs. Without further engagement with actual studies, the case has not been made that *any* actual studies fail to provide evidence of outcome or hindsight bias or that any actual researcher exhibits a fallacy fallacy.

Of course, this point cuts both ways: I do not claim that all study designs are flawless, and I agree that underestimation cases can be found in the empirical literature. But it is an intricate question whether a given study is an underestimation case. An important methodological upshot is that progress requires that philosophical theorizing about the epistemic strength of outcome information goes hand in hand with analyses of specific study designs. There is an interesting corollary to this point. Recall that an additional reason for taking the studies to provide evidence of bias is that researchers have been able to minimize the effect by way of stock debiasing manipulations. However, it might be objected that such debiasing manipulations are in effect changing the evidential situation in a manner that makes it ideally rational for the participants to change their view.⁸ Just as Hedden's point that outcome information has some epistemic weight should be recognized, it should also be recognized that debiasing manipulations may constitute changes in first- or higher-order evidence. This is something that empirical researchers should observe and something that they are generally sensitive to when devising debiasing strategies. For example, debiasing strategies are typically designed such that they do not introduce novel first-order evidence or overt defeaters to it. Rather, they seek to prompt the participants to process the original evidence in a cognitively different manner that is typically more reflective. As in the case of outcome information, there can be underestimation cases in which researchers underestimate the intrinsic epistemic weight of a debiasing manipulation. In such cases, this intrinsic epistemic weight, rather than an altered mode of processing, may be what causes a rational change in judgment. But although these types of underestimation cases may occur, the fact that researchers are sensitive to the issue and the fact that multiple types of debiasing manipulations minimize the effect of outcome information provide additional reasons to regard the judgments in question as biased.

So, given a partial survey of the empirical work, it is reasonable to conclude, *pace* Hedden, that (a) that there is ample empirical evidence taking outcome and hindsight bias to be important features of human cognition and (b) that cognitive psychologists do not *generally* exhibit a fallacy fallacy of postulating such biases in cases where there are none.

5 Conclusion

Assessing the evidence for postulating biases is a difficult affair requiring both reflection on epistemic norms of rationality and evaluation of the empirical research. While Hedden's arguments do not establish the conclusion that his title suggests or the charge that cognitive psychologists are generally committing a fallacy fallacy, his discussion produces insightful contributions. These include a formally clear characterization of the phenomena in question and clear examples of conditions in which outcome-influenced judgments are not merely boundedly but ideally rational. Although some of these case-types are recognized in the literature, it is a valuable

⁸ Thanks to Referee #3 for pressing this point.

contribution to integrate them in a formal framework. Likewise, reflection on the relation between the nature of the *ex ante* evidence and outcome information may result in principles that improve our understanding of when outcome-influenced judgments are and are not manifesting a bias. In this paper, I have sought to make such contributions by articulating the principle *Epistemic Weight* and the notion of an *underestimation case*. But in pursuing such contributions, there is a risk of philosophical hubris in drawing substantive conclusions about human rationality solely on the basis of, largely a priori, structural insights. Such insights are critical to an assessment of human rationality. Yet drawing substantive conclusions about whether types of patterns of judgments are biased also requires careful engagement with the relevant empirical research on the human judgments under discussion (Gerken 2017, 2022). So, from a methodological point of view, it is important to note that philosophical work is most fruitful in conjunction with detailed analysis of specific study designs. Taken in this more constructive manner, Hedden's challenge as well as the present response and development illustrate how philosophical reflection may contribute positively to empirical research on outcome and hindsight bias. In hindsight, this is a good outcome.

Acknowledgements I am grateful to referees 1, 3, and 4 for very helpful comments as well as to the editors for sensible advice. Thanks also to Kenneth Boyd, Brian Hedden, Markus Kneer, Uwe Peters, and Izabela Skoczen.

Funding Det Frie Forskningsråd Grant nummer: 8018-00053B.

Data availability No available data.

Declarations

The article complies with the journal's ethical standards and there are no conflicts of interest.

Conflict of interest statement No conflicts of interest.

References

- Arkes, H.R., D. Faust, T.J. Guilmette, and K. Hunt. 1988. Eliminating the hindsight bias. *Journal of Applied Psychology* 73 (2): 305–307.
- Baron, J., and J.C. Hershey. 1988. Outcome bias in decision evaluation. *Journal of Personality and Social Psychology* 54: 569–579.
- Berlin, L. 2004. Malpractice issues in radiology: Outcome bias. *American Journal of Roentgenology* (1976) 183 (3):557–560.
- Bernstein, D. M., Aßfalg, A., Kumar, R., & Ackerman, R. 2016. Looking backward and forward on hindsight bias. In *Oxford library of psychology. The Oxford handbook of metamemory*, eds. J. Dunlosky & S. K. Tauber, p. 289–304. Oxford University Press.
- Blank, H., J. Musch, and R.F. Pohl. 2007. Hindsight bias: On being wise after the event. *Social Cognition* 25 (1): 1–9.
- Blank, H., S. Nestler, G. von Collani, and V. Fischer. 2008. How many hindsight biases are there? *Cognition* 106 (3): 1408–1440.
- Bradfield, A., and G.L. Wells. 2005. Not the same old hindsight bias: Outcome information distorts a broad range of retrospective judgments. *Memory & Cognition* 33 (1): 120–130.

- Christensen-Szalanski, J.J., and C.F. Willham. 1991. The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes* 48 (1): 147–168.
- Clarkson, P.M., C. Emby, and V.W.S. Watt. 2002. Debiasing the outcome effect: The role of instructions in an audit litigation setting. *Auditing: A Journal of Practice & Theory* 21 (2): 7–20.
- Edwards, W. 1984. How to make good decisions. *Acta Psychologica* 56: 5–27.
- Evans, J. 2010. *Thinking Twice – two minds in one brain*. Oxford University Press.
- Fischhoff, B. 1975. Hindsight foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1: 288–299.
- Fischhoff, B. 2007. An early history of hindsight research. *Social Cognition* 25: 10–13.
- Fischhoff, B., and R. Beyth. 1975. “I knew it would happen”— Remembered probabilities of once-future things. *Organizational Behavior & Human Performance* 13: 1–16.
- Fisher, J., and T.I. Selling. 1993. The outcome effect in performance evaluation: Decision process observability and consensus. *Behavioral Research in Accounting* 5: 58–77.
- Gerken, M. 2011a. Conceptual equivocation and warrant by reasoning. *Australasian Journal of Philosophy* 89 (3): 381–400.
- Gerken, M. 2011b. Warrant and action. *Synthese* 178 (3): 529–547.
- Gerken, M. 2012. Discursive justification and skepticism. *Synthese* 189 (2): 373–394.
- Gerken, M. (2013a). Epistemic reasoning and the mental. *Innovations in Philosophy*. Palgrave Macmillan.
- Gerken, M. 2013b. Epistemic focal bias. *Australasian Journal of Philosophy* 91 (1): 41–61.
- Gerken, M. 2017. *On folk epistemology. How we think and talk about knowledge*. Oxford University Press.
- Gerken, M. 2020. Truth-sensitivity and folk epistemology. *Philosophy and Phenomenological Research* 100 (1): 3–25.
- Gerken, M. 2022. *Scientific testimony. Its roles in science and society*. Oxford University Press.
- Gigerenzer, G. 2008. *Rationality for mortals: How people cope with uncertainty*. Oxford University Press.
- Guilbault, R.L., F.B. Bryant, J.H. Brockway, and E.J. Posavac. 2004. A metaanalysis of research on hindsight bias. *Basic and Applied Social Psychology* 26: 103–117.
- Harley, E.M., K.A. Carlsen, and G.R. Loftus. 2004. The ‘Saw-It-All-Along’ effect: Demonstrations of visual hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 960–968.
- Hawkins, S.A., and R. Hastie. 1990. Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin* 107: 311–327.
- Hedden, B. 2019. Hindsight bias is not a bias. *Analysis* 79 (1): 43–52.
- Hoffrage, U., R. Hertwig, and G. Gigerenzer. 2000. Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26 (3): 566.
- Kneer and Skoczen. forthcoming. Outcome Effects, Moral Luck and the Hindsight Bias *Cognition*.
- Kneer, M., and E. Machery. 2019. No luck for moral luck. *Cognition* 182: 331–348.
- Koriat, A., S. Lichtenstein, and B. Fischhoff. 1980. Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6 (2): 107–118.
- Lowe, D.J., and P.M. Reckers. 1994. The effects of hindsight bias on jurors’ evaluations of auditor decisions. *Decision Sciences* 25 (3): 401–426.
- Mackie, D.M., M.N. Ahn, A.G. Asuncion, and S.T. Allison. 2001. The impact of perceiver attitudes on outcome biased dispositional inferences. *Social Cognition* 19: 71–93.
- Marshall, G.W., and J.C. Mowen. 1993. An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling & Sales Management* 13 (3): 31–47.
- Mazzocco, P.J., M.D. Alicke, and T.L. Davis. 2004. On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology* 26: 131–146.
- Oppenheimer, D.M. 2008. The secret life of fluency. *Trends in Cognitive Sciences* 12 (6): 237–241.
- Pezzo, M.V. 2011. Hindsight bias: A primer for motivational researchers. *Social and Personality Psychology Compass* 5 (9): 665–678.
- Pohl, R.F. 2007. Ways to assess hindsight bias. *Social Cognition* 25 (1): 14–31.
- Roese, N., and K. Vohs. 2012. Hindsight bias. *Perspectives on Psychological Science* 7: 411–426.
- Samuels, R., Stich, S., & Bishop, M. 2012. Ending the rationality wars. *Collected Papers, Volume 2: Knowledge, Rationality, and Morality, 1978–2010*, 2, 191.
- Sezer, O., T. Zhang, F. Gino, and M.H. Bazerman. 2016. Overcoming the outcome bias: Making intentions matter. *Organizational Behavior and Human Decision Processes* 137: 13–26.
- Stanovich, K. 2011. *Rationality and the Reflective Mind*. Oxford University Press.

- Stanovich, K.E., and R.F. West. 2008. On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology* 94: 672–695.
- Tan, H.T., and M.G. Lipe. 1997. Outcome effects: The impact of decision process and outcome controllability. *Journal of Behavioral Decision Making* 10 (4): 315–325.
- Turri, John. 2016. The radicalism of truth-insensitive epistemology: Truth's profound effect on the evaluation of belief. *Philosophy and Phenomenological Research* 93(2):348–367.
- Wallin, A. 2013. A peace treaty for the rationality wars? External validity and its relation to normative and descriptive theories of rationality. *Theory & Psychology* 23 (4): 458–478.
- Wood, G. 1978. The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance* 4 (2): 345–353.
- Young, L., S. Nichols, and R. Saxe. 2010. Investigating the Neural and Cognitive Basis of Moral Luck: It's Not What You Do but What You Know. *Review of Philosophical Psychology* 1: 333–349.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.