

What Determines Hindsight Bias in Written Work? One Field and Three Experimental Studies in the Context of Wikipedia

Marcel Meuer¹, Steffen Nestler², and Aileen Oeberst^{3, 4}

¹ Department of Psychology, University of Mainz

² Department of Psychology, University of Münster

³ Department of Psychology, University of Hagen

⁴ Knowledge Construction Lab, Leibniz-Institut für Wissensmedien, Tübingen, Germany

Hindsight bias not only occurs in individual perception but in written work (e.g., Wikipedia articles) as well. To avoid the possibility that biased written representations of events distort the views of broad audiences, one needs to understand the factors that determine hindsight bias in written work. Therefore, we tested the effect of three potential determinants: the extent to which an event evokes sense-making motivation, the availability of verifiable causal information regarding the event, and the provision of content policies. We conducted one field study examining real Wikipedia articles ($N = 40$) and three preregistered experimental studies in which participants wrote or edited articles based on different materials (total $N = 720$). In each experiment, we systematically varied one determinant. Findings provide further—and even more general—support that Wikipedia articles about various events contain hindsight bias. The magnitude of hindsight bias in written work was contingent on the sense-making motivation and the availability of causal information. We did not find support for the effect of content policies. Findings are in line with causal model theory and suggest that some types and topics of written work might be particularly biased by hindsight (e.g., coverage of disasters, research reports, written expert opinions).

Public Significance Statement


This research demonstrates that written work (e.g., Wikipedia articles) can be biased by hindsight: After an event happened, written work is more suggestive of the event, mistakenly describing it as more foreseeable and inevitable than it had been. The stronger the need for explanation and the more causal information available, the more biased the writing. These findings are important because biased writing can bias the views of many people.

Keywords: hindsight bias, text production, media, Wikipedia

“Why Trump’s Election Was Inevitable” (Marchetti, 2017), “Analysis: Pope’s resignation almost predictable” (Rocca, 2013), and “The Financial Crisis Was Foreseeable and Preventable” (Frieden, 2011). When news headlines suggest that an event was inevitable, predictable, foreseeable, or preventable, then, of course, said event might have been just that. However, when looking back at an event, people often succumb to hindsight bias, which is the

tendency to perceive events *after* their occurrence as more likely, more inevitable, and more foreseeable than they had actually been *before* they happened (see Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004, for meta-analyses; see Hawkins & Hastie, 1990; Pohl & Erdfelder, 2017; Roese & Vohs, 2012, for reviews). Besides the robust finding that people are subject to hindsight bias in their *individual perceptions*, recent research has demonstrated that

Marcel Meuer  <https://orcid.org/0000-0003-1648-5359>

Steffen Nestler  <https://orcid.org/0000-0001-9724-2441>

Aileen Oeberst  <https://orcid.org/0000-0002-1094-9610>

Marcel Meuer is now at the University of Hagen, Germany. This research was funded by the German Research Foundation to Aileen Oeberst (OE 604/3-1) and Steffen Nestler (NE 1485/8-1). All authors declare that they have no conflicts of interest. All procedures performed in the present study were in accordance with the ethical guidelines specified in the American Psychological Association (APA) Code of Conduct as well as the authors’ national ethics guidelines. Preregistrations, study materials, data, and analysis scripts are openly accessible at <https://osf.io/8dfuh/>. The authors would like to thank Marieke von Elert, Lena Scholz, Louisa Spielvogel, and Maria Zuleger for data collection.

Marcel Meuer played lead role in data curation, formal analysis, investigation, software, visualization, and writing of original draft, supporting role in resources, and equal role in methodology, validation, and writing of review and editing. Steffen Nestler played lead role in conceptualization, funding acquisition, project administration, and supervision, supporting role in data curation, formal analysis, investigation, resources, and writing of original draft, and equal role in methodology, validation, and writing of review and editing. Aileen Oeberst played lead role in conceptualization, funding acquisition, project administration, and supervision, supporting role in data curation, formal analysis, investigation, and writing of original draft and equal role in methodology, resources, validation, and writing of review and editing.

Correspondence concerning this article should be addressed to Marcel Meuer, Department of Psychology, University of Hagen, Universitätsstraße 47, Hagen 58097, Germany. Email: marcel.meuer@femuni-hagen.de

hindsight bias can even manifest in *written work* such as Wikipedia articles (Meuer, Nestler, & Oeberst, 2021; Oeberst et al., 2018).

In the present research, we aimed to examine under which circumstances hindsight bias enters written work. This is an important research question as it might not only inform our theorizing about the mechanisms underlying hindsight bias in written event representations, but it may also contribute to a more accurate representation of the world. Specifically, reading biased written event accounts can elicit or increase the already existing hindsight bias in readers (Meuer, von der Beck, et al., 2021; Oeberst et al., 2014, 2018; von der Beck et al., 2017). Believing that an event was more likely, more inevitable, and more foreseeable than it actually was can cause overconfidence concerning one's knowledge and abilities (Bradfield & Wells, 2005; Cassar & Craig, 2009; Granhag et al., 2000) and unwarranted attributions of responsibility and blame (Carli, 1999; Harley, 2007; Hastie et al., 1999). Knowing the factors that determine hindsight bias in written work, in turn, might help to find potential strategies to avoid or reduce article hindsight bias and its negative impact on the individual representation of events.

In addition, since hindsight bias in written work might also depend on the specific writing context of authors, the present research draws on one concrete type of writing for which hindsight bias has already been documented: Wikipedia articles (Oeberst et al., 2018). With more than 56 million articles in 321 language editions and 1.7 billion unique active user devices per month (Wikimedia Statistics, 2021; "Wikipedia", 2021), Wikipedia is the most comprehensive and most widely used online encyclopedia (Similarweb, n.d.). Therefore, biased Wikipedia articles might distort the views of broad audiences, and assessing the determinants of hindsight bias in Wikipedia articles is thus of practical relevance.

Hindsight Bias in Individuals and Written Work

Since Fischhoff's (1975) early demonstration that people in retrospect overestimate the likelihood of historical events, hindsight bias in individuals has been documented in many contexts such as legal (Giroux et al., 2016), medical (Arkes, 2013), economic decision-making (Biais & Weber, 2009), election outcomes (Blank et al., 2003), sporting events (Bonds-Raacke et al., 2001), and scientific findings (Slovic & Fischhoff, 1977). Hindsight bias is a robust and pervasive phenomenon (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004), and people are often unaware that they succumbed to hindsight bias (Pohl & Hell, 1996). In addition, people rarely succeed in avoiding hindsight bias even if they know about it and want to counteract it (Fischhoff, 1977; Guilbault et al., 2004). Therefore, one might assume that individuals also incorporate hindsight bias into their writing about events.

Oeberst and colleagues examined this assumption in the context of the online encyclopedia *Wikipedia* (Oeberst et al., 2018; Oeberst, von der Beck, Cress, & Nestler, 2020), where articles are collaboratively created and edited by anyone who wants to. The wiki technology stores every single version of each article, so that any revision made can be tracked or reversed (van Dijk, 2021). To test for hindsight bias, Oeberst et al. (2018) extracted Wikipedia articles for 33 specific events from six event categories (i.e., disasters, elections, official decisions, personal decisions, sports events, scientific findings). For each event, they retrieved three article versions from the revision history: the last article version that existed prior to the event (t_1), the first article version that mentioned the event (t_2), and the article

version that existed 8 weeks after the occurrence of the event (t_3).¹ They then had 10 independent, trained coders who were blind to the research question judge the extent to which each article version suggested the occurrence of the respective event. The analyses identified hindsight bias in the t_3 articles about disasters (i.e., a significant t_1 - t_3 increase in the suggestiveness ratings). For the remaining event categories, however, the authors found *no* evidence for hindsight bias in Wikipedia. This is a noteworthy finding because hindsight bias for events from these categories has been well documented on an individual level (i.e., elections, Blank et al., 2003; personal and official decisions, Louie et al., 2007; sporting events, Bonds-Raacke et al., 2001; scientific findings, Slovic & Fischhoff, 1977). In a similar vein, Oeberst, von der Beck, Cress, & Nestler (2020) directly compared hindsight bias in individuals and Wikipedia articles for a state election and found hindsight bias at an individual level but not in the Wikipedia articles. Evidently, hindsight bias can but does not have to manifest itself in written work such as Wikipedia articles. In the following, we elaborate on three factors that may determine whether written work comprises hindsight bias.

Potential Determinants of Hindsight Bias in Written Work

To understand which factors might determine the size of hindsight bias in *written work*, it is reasonable to first outline the presumed mechanisms behind hindsight bias in *individuals*. According to *causal model theory* (Blank & Nestler, 2007; Nestler et al., 2008), hindsight bias in individuals is the result of one-sided sense-making processes that take place after an event occurred (Pezzo, 2003; Roes & Olson, 1996; but see also Roes & Vohs, 2012, for an overview on metacognitive and motivational determinants of hindsight bias). To find an explanation for the event (i.e., a causal model), people engage in biased information selection, interpretation, and/or integration. That is, people search for relevant antecedents of the event (i.e., preceding circumstances that are perceived to be related to the event) by considering the available event information (e.g., official investigations, media coverage) and/or information in their long-term memory that is (stereotypically) related to the type of event. For instance, to explain the Fukushima Daiichi nuclear disaster, people might gather information about the preceding tsunami and the safety measures taken at the power plant. However, since the event knowledge serves as a retrieval cue, people tend to select predominantly *event-consistent* antecedents (e.g., that there were long-known construction errors and covered-up incidents at the power plant rather than information about existing safety measures; Carli, 1999; Roes & Olson, 1996; Slovic & Fischhoff, 1977). The selected information will then be interpreted and integrated into a causal model of the event. But again, striving to explain the event, this evaluation process can be biased in favor of affirmative rather than disproving antecedent-event relations (Blank & Nestler, 2007). For instance, ambiguous information can be interpreted in an event-consistent manner (e.g., considering a tsunami protection

¹ To obtain a normative measure of hindsight *bias* in Wikipedia articles, the authors needed to compare foresight and hindsight accounts on each event and were thus confined to topics for which there already existed a Wikipedia article before the respective event occurred (e.g., the article about the Fukushima Daiichi nuclear power plant for the Fukushima nuclear disaster).

wall as a safety concern due to its low height instead of citing its mere existence as a protection factor) and event-consistent antecedents can be given more weight than event-inconsistent information (e.g., emphasizing safety deficiencies instead of existing safety measures). As a result, people in hindsight unknowingly adopt a one-sided, event-consistent causal representation of the event.

If individuals, in turn, hold such biased representations, they might also introduce these representations into written work. However, hindsight bias was not observed in writings about several types of events for which the bias has been well documented in individuals (Oeberst et al., 2018; Oeberst, von der Beck, Cress, & Nestler, 2020), suggesting that biased individual views are not incorporated into written work in *any* case. Instead, the content (and hence accuracy) of written work is the result of a complex interaction of multiple factors, such as features of the topic, the writing context, and the authors (Hayes, 1989; Nestler et al., 2017). Therefore, we propose two event characteristics and one characteristic of the writing context that might determine whether written work contains hindsight bias. First, events differ in the extent to which they elicit the *motivation to find an explanation*, which might affect the sense-making processes that, according to causal model theory, result in biased causal representations. Second, events differ in how much *causal information* is available, providing people with a varying capability to construct an event-consistent (and thus biased) account. Finally, with regard to the writing context, some information outlets such as Wikipedia or news agencies require authors to adhere to *content policies*, which might heighten the threshold for hindsight bias to enter the writing. In the following, we describe in more detail how each of the three characteristics might determine the size of hindsight bias in written work.

Sense-Making Motivation

Causal model theory suggests that hindsight bias is based on effortful causal reasoning processes (Blank & Nestler, 2007). Although people are generally motivated to uncover and understand the causal relations of their world (Gopnik, 2000; Lombrozo, 2006; Schwitzgebel, 1999), certain situational factors foster these sense-making processes. First, sense-making is particularly triggered after *unexpected* events, that is, when there is a discrepancy between people's expectations and reality (Hastie, 1984; Meyer et al., 1997; Nestler & Egloff, 2009; Pezzo, 2003; Reizenzein et al., 2019; Weiner, 1985). If this discrepancy exceeds a certain threshold, people experience a feeling of surprise. As a consequence, people shift their attention to the unexpected event to make sense of the occurrence (Hagmayer & Sloman, 2009; Meyer et al., 1997; Reizenzein et al., 2019). Similarly, sense-making is particularly pronounced after *negative* events (Fischer-Preßler et al., 2019; Weiner, 1985): To quickly terminate potential negative consequences and to avoid negative events and their detriments in the future, people are motivated to identify the causes of negative events (Neuberg et al., 2011; Park, 2010).

Consequently, unexpected and negative events evoke particularly pronounced causal reasoning processes, which, however—according to causal model theory—are biased toward an event-consistent evaluation (Blank & Nestler, 2007). This is in line with findings that individuals' hindsight bias is particularly high after unexpected and negative events (Ash, 2009; Pezzo, 2003; Schkade & Kilbourne, 1991). Similarly, Oeberst et al. (2018) identified hindsight bias exclusively in Wikipedia articles about *disasters*, which combine

both characteristics (i.e., disasters are negative in that they cause deaths, injuries, and/or material damage, and they usually occur unexpectedly; Fischhoff et al., 2005; Verplanken & Pieters, 1988), and which might thus trigger a higher sense-making motivation than events from other event categories (i.e., scientific findings, results of elections or sports events, and personal decisions are not necessarily negative and unexpected). In the present research, we therefore examined whether the extent to which an event elicits sense-making processes affects the magnitude of hindsight bias in written work.

Availability of Causal Information

Another factor that might determine the extent to which written work contains hindsight bias is the availability of antecedents. A crucial assumption of causal model theory is that people use available antecedents to construct an event-consistent explanation (Blank & Nestler, 2007). Hence, the more one succeeds in establishing strong antecedent-event relations (i.e., attributions of the event to antecedents such as attributing the Fukushima nuclear disaster to severe safety deficiencies), the more likely, inevitable, and foreseeable should the event appear in retrospect (Pezzo, 2003). In line with this reasoning, hindsight bias increases when individuals have access to a greater number of antecedents (Nario & Branscombe, 1995; Nestler et al., 2008, 2010; Nestler & von Collani, 2008) and decreases when antecedent-event relations are weakened (e.g., by asking participants to use the antecedents to explain an *alternative* event; Arkes et al., 1988; Carli & Leonard, 1989; Nario & Branscombe, 1995; Wasserman et al., 1991). Also, if no causal information is available at all, people do not succumb to hindsight bias (Yopchick & Kim, 2012).

Events, in turn, vary in the extent to which causal information is available. For instance, in the aftermath of a disaster, it is important to understand the exact course of events to determine who should be held responsible and to learn from the events and take action to prevent future disasters. To this end, detailed investigations are carried out and final reports (or at least parts of them) are made public. Furthermore, disasters are extensively covered by the media (Gaddy & Tanjong, 1986; Joye, 2010; Simon, 1997), providing people with a vast amount of information from various sources. On the other hand, there are events for which strong antecedent-event linkages are rare: For instance, for personal or public decisions such as the engagement of Prince William and Kate Middleton or the secretly held election of Cardinal Joseph Ratzinger as the new pope, causal information from verifiable sources might not be accessible at all. Importantly, however, many information outlets (e.g., Wikipedia, news media) demand that all information in their publications is verifiable (see also below). Consequently, the more *verifiable* causal (event-consistent and event-inconsistent) information is available in the aftermath of an event, the more capacities authors have to explain the respective event. As suggested by causal model theory, authors might then primarily incorporate the available event-consistent information into their articles, fostering a biased written event representation (Blank & Nestler, 2007).

Content Policies

Not only features of the event but also features of the writing context affect the production of texts (Hayes, 1989; Nestler et al., 2017), and might thus also influence the magnitude of hindsight bias

in written work. One feature of the writing context that might *reduce* hindsight bias is the existence of writing policies or guidelines. Since prior research on hindsight bias in written work was set in the writing context “Wikipedia” (Oeberst et al., 2018; Oeberst, von der Beck, Cress, & Nestler, 2020), we will address three content policies that guide the collaborative writing in Wikipedia (“Wikipedia: Core content policies”, 2021). However, very similar guidelines apply in other writing contexts (e.g., news outlets, lexica; Reuters News Agency, n.d.), so our considerations are also informative for these types of written work.

First, all contributions to Wikipedia need to be written from a *neutral point of view* (NPOV) by “representing fairly, proportionately, and (. . .) without editorial bias, all the significant views that have been published by reliable sources on a topic” (“Wikipedia: Neutral point of view”, 2021). Second, information within Wikipedia articles needs to be *verifiable*, that is, attributable to reliable sources (“Wikipedia: Verifiability”, 2021). Finally, articles must *not include original research*—content such as ideas, allegations, or even facts that have not been published before (“Wikipedia: No original research”, 2021). Together, the three policies specify the type and quality of content that is acceptable for Wikipedia, which is intended to prevent the inclusion of any kind of bias.

Based on the assumptions of causal model theory (Blank & Nestler, 2007), these content policies might reduce hindsight bias in written work in at least two ways. First, people’s biased causal model of an event might rely not only on verifiable evidence but also on personal speculation about alleged antecedents (Carli, 1999). The policies to use no original research and only verifiable information from reliable sources, however, prompt authors to refrain from such unverifiable speculation. Take the engagement of Prince William and Kate Middleton again as an example: After the engagement was announced, some people might have believed that the couple’s body language during interviews had always indicated a happy, stable relationship, which might have led them to the conclusion that the engagement was foreseeable all along. However, these people could not have included this unverifiable interpretation in the respective Wikipedia article. Second, the policy to write from an NPOV asks authors to proportionately present *all* the significant published perspectives on a topic. This demand might encourage people to consider not only event-consistent but also event-*inconsistent* information, which might foster a more balanced event representation. Adhering to this policy might trigger similar cognitive processes as the consider-the-opposite strategy, which prompts people to find reasons for why the event was *not* going to happen (Lord et al., 1984). This strategy is the most promising intervention to reduce hindsight bias in individuals (Guilbault et al., 2004).

Yet, although such content policies might help to *reduce* hindsight bias, they cannot entirely *preclude* it, as demonstrated by the finding that Wikipedia articles about disasters contained hindsight bias (Oeberst et al., 2018). For one thing, people are usually unaware that they succumbed to hindsight bias and might thus present an unbalanced written event account without noticing (Pohl & Hell, 1996). Even if multiple authors collaborated (which is the case in Wikipedia), such biased views might remain unnoticed as hindsight bias, given its robustness and pervasiveness (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004), is likely shared among all authors of an article. Furthermore, under certain circumstances, the policy to write from an NPOV might even encourage to put comparably *more* focus on event-consistent information, thereby facilitating

hindsight bias: The policy demands authors to present all significant viewpoints on the topic “in proportion to the prominence of each viewpoint in the published, reliable sources” (“Wikipedia: Neutral point of view”, 2021). Consequently, if reliable sources are already biased and predominantly refer to event-consistent information, this should be also reflected in written work that adheres to the NPOV policy.

In sum, then, the three content policies might heighten the threshold for hindsight bias to enter written work such as Wikipedia articles. As a result, hindsight bias in writings might be prevented in some instances even though the respective events elicit hindsight bias in individuals (e.g., elections; Oeberst et al., 2018; Oeberst, von der Beck, Cress, & Nestler, 2020), but might still be evident in written accounts of events for which people feature a particularly high hindsight bias (e.g., disasters).

The Present Research

We conducted one exploratory field study and three preregistered experimental studies to examine whether each of the three proposed determinants affects the size of hindsight bias in written work. In Study 1, we built on prior research on hindsight bias in written work (Oeberst et al., 2018) and used real Wikipedia articles for an exploration of the proposed determinants. Specifically, we obtained measures reflecting the sense-making motivation and the availability of causal information for the specific events represented in the selected Wikipedia articles and performed a preliminary analysis of whether these two determinants predicted the magnitude of hindsight bias in the articles (since all articles were written based on Wikipedia’s content policies, we were not able to explore the effect of this determinant in this study). Additionally, by comparing the size of hindsight bias in articles referring to disasters versus non-disastrous events, the study allowed us to examine whether we could replicate the finding that particular articles about disasters contained hindsight bias (Oeberst et al., 2018). To systematically test for an effect of each of the three proposed determinants under controlled conditions, we then ran one preregistered lab study per determinant. The general procedure for these three studies was that participants should write or edit an article about a specific topic based on fictive, causally relevant source material, mirroring real-world conditions (i.e., high sense-making motivation, availability of causal information, adherence to content policies). In each of the three studies, however, we systematically varied one of the determinants and tested whether this manipulation affected the size of hindsight bias in the produced articles.

Specifically, in Study 2, participants learned about a specific event that elicited either a low or high sense-making motivation (or received no event knowledge, representing the foresight perspective), and we expected hindsight bias to be smaller in articles written by participants who learned about the event triggering only a low (vs. high) sense-making motivation (H_1). In Study 3, all participants learned about a specific event that elicited a high sense-making motivation (or received no event knowledge), but we manipulated whether they received source information that was of high or only low causal relevance. Here, we expected hindsight bias to be smaller in articles based on the material with low (vs. high) causal relevance (H_2). In Study 4, participants again learned about a specific event (or received no event knowledge) and received an article with both verifiable and unverifiable information concerning

the event. All participants should then edit and finalize the article, but they either were asked to adhere to specific content policies or received no further writing guidance while editing. We expected hindsight bias in the edited articles to be greater when content policies were provided (vs. not provided; H_3).

Study 1—Pilot Study in Wikipedia

In the first study, we used real Wikipedia articles to explore whether the extent to which the presented events elicit sense-making motivation—operationalized as the increase in article views after the event had happened—and the availability of causal information—assessed by the extent to which each article added explanatory content—are related to the magnitude of hindsight bias in the respective articles. Another aim of this field study was to examine whether only Wikipedia articles addressing disasters comprise hindsight bias (as suggested by Oeberst et al., 2018) or whether hindsight bias is also present in Wikipedia articles about other events.

Method

Transparency and Openness

For all the studies that follow, we report planned sample sizes, all data exclusions, all manipulations, and all measures, following the Journal Article Reporting Standards (Appelbaum et al., 2018). Data were analyzed using *SPSS*, Version 23.0 (IBM Corp., 2015), and *R*, Version 4.0.0 (R Core Team, 2020). All materials, data, analysis scripts, and results of supplemental analyses are openly accessible on the Open Science Framework (OSF; see <https://osf.io/8dfuh/>). Furthermore, Studies 2–4 were preregistered; the preregistrations can also be found in the aforementioned OSF project.

Selection of Events and Article Versions

We selected 20 specific events that qualified as a disaster and 20 specific events from other event categories (henceforth referred to as “nondisasters”). As disasters, we chose incidents with serious negative consequences such as fatalities, many injured/endangered persons, or enormous material damage (e.g., Fukushima Daiichi nuclear disaster; capsizing of the cruise ship “Costa Concordia”; collapse of a bridge in Italy). As nondisasters, we chose events from various event categories of different valence that are commonly used in hindsight bias research (e.g., elections such as the Irish abortion referendum; scientific discoveries such as the direct observation of gravitational waves; personal and official decisions such as the resignation of Pope Benedict XVI). To obtain unbiased foresight representations of each event, event selection was confined to topics (a) for which a Wikipedia article already existed before the occurrence of the event (e.g., the article about the Fukushima Daiichi power plant, the article about Pope Benedict XVI) and (b) which were thematically specific enough to potentially suggest the event in question (e.g., no global articles about power plants or popes in general; see also Oeberst et al., 2018 and Footnote 1). For each of the selected 40 events, we then extracted three Wikipedia article versions from the revision history: the last article version that existed prior to the event (t_1), the first article version that mentioned the event (t_2), and the article version that existed

8 weeks after the occurrence of the event (t_3). In total, we thus retrieved $3 \times 40 = 120$ article versions, comprising a 2 (event category; between-events) \times 3 (article version; within-event) mixed design.

Since the original articles were up to 20 pages long, posing the risk of coder fatigue, we decided to shorten the articles by deleting text passages containing background information that was irrelevant to the event in question (e.g., *Abba*’s discography for their reunion in 2018, the band history of *Goodbye to Gravity* for the nightclub fire in 2015). Importantly, such passages were only deleted if they appeared in all three article versions of the respective event to avoid a systematic impact of the deletion on the magnitude of hindsight bias (see <https://osf.io/8dfuh/>, for the full list of events and all original and edited German article versions).

Article Coding

We assigned all articles to fixed blocks of six articles. Each block consisted of articles about six different events from one event category (two t_1 versions, two t_2 versions, and two t_3 versions). We then asked 124 participants who were blind to the research questions to code one randomly selected block of articles. This assignment procedure ensured that participants never coded multiple article versions of the same event. By this means, each article was rated by five to eight independent coders ($M = 6.20$), whose ratings were averaged for the main analyses.

The six articles were presented in random order, and participants rated each article on five dimensions related to hindsight bias, each with respect to the *specific* event: First, they judged whether the articles suggested the occurrence of the respective events (e.g., “To what extent does the article suggest a *collapse of the bridge/a resignation of the Pope/ . . . ?*”) and whether the articles implied that the events were inevitable or foreseeable (e.g., “To what extent does the article suggest that a *collapse of the bridge/a resignation of the Pope/ . . . was inevitable/foreseeable?*,” all on scales from 1 = *not at all* to 5 = *very much*). These three measures are common operationalizations to assess hindsight bias.² Next, they coded whether the articles contained explicit statements that are typical for hindsight bias (e.g., “It was obvious that the event would happen”; 0 = *no explicit statements*, 1 = *explicit statements*) and whether the articles provided an explanation for the event (on a scale from 1 = *no explanation provided* to 4 = *extensive explanation provided*), both measures addressing potential means by which hindsight bias might be introduced into written work. We used the latter as a proxy for the availability of causal information. In addition, participants judged the quality of the articles (on a scale from 1 = *low quality* to 5 = *high quality*). Importantly, participants were instructed to base their evaluations only on the content of each article, irrespective of whether the event in question was actually mentioned in the article

² Recent studies differentiated hindsight bias in three distinct phenomenological manifestations (Blank et al., 2008; Nestler et al., 2010). To obtain an exhaustive representation of hindsight bias, we thus used three operationalizations: (a) the general suggestiveness of the specific incident (i.e., as a global hindsight bias measure, resembling the likelihood measure utilized in classic work on hindsight bias), (b) the suggested inevitability (i.e., whether the specific incident was described to be objectively predictable, “It had to happen”), and (c) the suggested foreseeability (i.e., whether the specific incident was described to be subjectively predictable, “It was obvious that it would happen”).

(which was rarely the case in foresight articles; see <https://osf.io/8dfuh/>, for a survey export).

As the coding comprised an ill-structured measurement design (i.e., raters and articles were neither fully crossed nor nested), we computed $G(q, k)$ according to Putka et al. (2008) to estimate the reliability of the ratings. The analysis yielded moderate-to-good interrater reliabilities for all variables, $G_{\text{suggestiveness}}(0.16, 6.02) = .78$, $G_{\text{inevitability}}(0.16, 6.02) = .70$, $G_{\text{foreseeability}}(0.16, 6.02) = .76$, $G_{\text{explanation}}(0.16, 6.02) = .83$, $G_{\text{quality}}(0.16, 6.02) = .65$, except for the coding of explicit statements, which was not reliable and thus excluded from further analyses, $G_{\text{statements}}(0.16, 6.02) = .37$.

In addition to the coding, we used the online tool *Wikishark Trends* (Vardi et al., 2021) to extract a measure that reflects the public interest in each event, which might, in turn, serve as a proxy for people's sense-making motivation. The tool provides daily page views for Wikipedia articles of the English Wikipedia as of January 01, 2008.³ For all events that occurred after 2007 and for which an equivalent English-language version of the German Wikipedia article was available ($n = 31$; 19 disasters, 12 nondisasters), we used the tool to extract the number of article views per day for the 3 days before the respective event occurred, the event date itself, and the 2 following days. As a measure of public interest, we then computed the increase in article views after the event occurrence by subtracting the average number of views on the day of the event and the 2 subsequent days from the average number of views on the 3 days prior to the event (which reflects the baseline interest in the topic of each article).

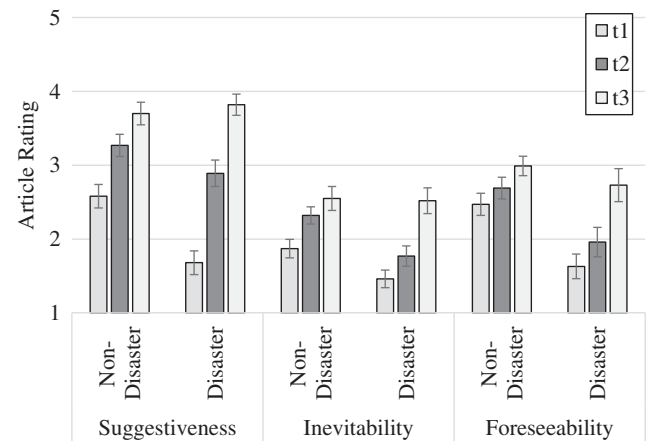
Results

Article Hindsight Bias

We first explored whether the selected hindsight articles (i.e., the t_2 and t_3 versions) about disasters and nondisasters contained hindsight bias, signified by an increased suggestiveness, inevitability, and foreseeability rating compared to the foresight article (i.e., the t_1 version). Starting with the analysis of hindsight bias in the t_2 articles, we computed separate 2 (event category; disaster, non-disaster; between-events factor) \times 2 (article version; t_1, t_2 ; within-event factor)⁴ mixed analyses of variance (ANOVAs) for each of the three dependent variables (averaged participant ratings of the suggestiveness, the inevitability, and the foreseeability of the respective event). The analysis yielded a significant main effect of article version on all three measures, $F_{\text{suggestiveness}}(1, 38) = 49.53, p < .001, \eta_p^2 = .57$, $F_{\text{inevitability}}(1, 38) = 16.74, p < .001, \eta_p^2 = .31$, and $F_{\text{foreseeability}}(1, 38) = 5.47, p = .025, \eta_p^2 = .13$. The t_2 hindsight articles were more suggestive of the event and implied a higher inevitability and foreseeability of the event than the t_1 foresight articles, signifying hindsight bias in the t_2 articles (see Figure 1). Also, there was a significant main effect of event category for all three measures, $F_{\text{suggestiveness}}(1, 38) = 11.87, p = .001, \eta_p^2 = .24$, $F_{\text{inevitability}}(1, 38) = 9.97, p = .003, \eta_p^2 = .21$, $F_{\text{foreseeability}}(1, 38) = 14.25, p = .001, \eta_p^2 = .27$. Articles about disasters were generally less suggestive of the respective incident and expressed a lower inevitability and foreseeability than articles about nondisasters. However, there was no significant interaction effect of event category and article version for any of the dependent variables, $F_{\text{suggestiveness}}(1, 38) = 3.75, p = .060, \eta_p^2 = .09$, $F_{\text{inevitability}}(1, 38) = 0.64, p = .427, \eta_p^2 = .02$, $F_{\text{foreseeability}}(1, 38) = 0.20, p = .659, \eta_p^2 = .01$.

Figure 1

Article Ratings as a Function of Article Version and Event Category



Note. Error bars are standard errors.

Thus, we found no support that hindsight bias in t_2 articles was greater for disasters than for nondisasters.

We applied the same test logic to examine hindsight bias in the t_3 articles, computing another set of 2 (event category; disaster, nondisaster; between-events factor) \times 2 (article version; t_1, t_3 ; within-events factor) mixed ANOVAs. We obtained a significant main effect of article version for all three dependent measures, $F_{\text{suggestiveness}}(1, 38) = 160.89, p < .001, \eta_p^2 = .81$, $F_{\text{inevitability}}(1, 38) = 53.69, p < .001, \eta_p^2 = .59$, and $F_{\text{foreseeability}}(1, 38) = 40.85, p < .001, \eta_p^2 = .52$. The suggestiveness, inevitability, and foreseeability ratings were higher for the t_3 hindsight articles than for the t_1 foresight articles. Hence, the t_3 articles also contained hindsight bias (see Figure 1). The analysis yielded a significant main effect of event category for the suggestiveness, $F(1, 38) = 4.94, p = .032, \eta_p^2 = .12$, and the foreseeability measure, $F(1, 38) = 6.99, p = .012, \eta_p^2 = .16$, but no effect for the inevitability measure, $F(1, 38) = 1.65, p = .207, \eta_p^2 = .04$. Compared to nondisasters, articles about disasters thus were less suggestive of the event and expressed a lower foreseeability of the event. Furthermore, there was a significant interaction effect of event category and article version for the suggestiveness, $F(1, 38) = 15.87, p < .001, \eta_p^2 = .30$, and the foreseeability measure, $F(1, 38) = 5.01, p = .031, \eta_p^2 = .12$, but no significant interaction for the inevitability measure, $F(1, 38) = 2.41, p = .129, \eta_p^2 = .06$.

To clarify the significant interaction between the suggestiveness and the foreseeability measures, we computed simple effects tests comparing the t_1 - t_3 differences (i.e., t_3 minus t_1) of articles about disasters and nondisasters. The analysis yielded a significant

³ Note that there is no tool that provides page views of German Wikipedia articles for a comparably large time frame. To retrieve an estimate of the public interest for as many of the selected events as possible, we therefore decided to use the pageview statistics of equivalent English Wikipedia articles (although the German article views would, of course, have provided a better approximation of the public interest in Germany).

⁴ We set *article version* as a within-factor (vs. between-factor) because the unit of analysis was *events* (and not *participants*), and the t_2 and t_3 articles of an event were revised versions of the respective t_1 article and were thus clearly dependent on each other, which is also signified by the predominantly moderate-to-large t_1 - t_2 and t_1 - t_3 Pearson correlation coefficients for all measures, $.39 < r_{112} < .63, .20 < r_{113} < .48$.

hindsight bias in articles about nondisasters in terms of both the suggestiveness and foreseeability measures, $\Delta M_{\text{suggestiveness}} = 1.11$, $SD = 0.73$, $F(1, 38) = 37.84$, $p < .001$, $\eta_p^2 = .50$, $\Delta M_{\text{foreseeability}} = 0.53$, $SD = 0.59$, $F(1, 38) = 8.62$, $p = .006$, $\eta_p^2 = .19$, but an even greater hindsight bias in articles about disasters, $\Delta M_{\text{suggestiveness}} = 2.14$, $SD = 0.89$, $F(1, 38) = 138.91$, $p < .001$, $\eta_p^2 = .79$, $\Delta M_{\text{foreseeability}} = 1.09$, $SD = 0.96$, $F(1, 38) = 37.23$, $p < .001$, $\eta_p^2 = .50$.⁵

In sum, we found hindsight bias in t_2 and t_3 articles of both event categories—consistently on all three operationalizations of hindsight bias. In addition, with regard to the suggestiveness and foreseeability measures, the t_3 articles about disasters were even more biased by hindsight than the t_3 articles about non-disastrous events.

Sense-Making Motivation

To examine whether people's sense-making motivation predicted the magnitude of hindsight bias in the articles, we correlated the t_1 – t_3 difference in the suggestiveness, inevitability, and foreseeability ratings with the increase in page views after the event occurrence. None of the correlations were significantly different from zero, $r_{s, \text{suggestiveness}}(29) = -.10$, $p = .600$, $r_{s, \text{inevitability}}(29) = .08$, $p = .688$, $r_{s, \text{foreseeability}}(29) = .02$, $p = .913$ (we computed Spearman correlation coefficients due to two extreme outliers in the increase in page views). Therefore, we obtained no support for a relationship between the magnitude of article hindsight bias and the sense-making motivation for the respective event.

Availability of Causal Information

To examine whether the availability of causal information is associated with the magnitude of hindsight bias in the articles, we correlated the t_1 – t_3 difference in the extent to which the articles provided an explanation for the respective event (i.e., addition of explanatory content) with the t_1 – t_3 difference in the suggestiveness, inevitability, and foreseeability ratings (i.e., hindsight bias). The analysis yielded a strong association of the addition of explanatory content and hindsight bias for all three operationalizations of hindsight bias, $r_{\text{suggestiveness}}(38) = .66$, $p < .001$, $r_{\text{inevitability}}(38) = .56$, $p < .001$, $r_{\text{foreseeability}}(38) = .50$, $p = .001$, suggesting a relationship between the availability of causal information and the magnitude of article hindsight bias.

Discussion

In this field study, we found consistent support for hindsight bias in real Wikipedia articles about both disasters and nondisasters (i.e., on all three operationalizations and for both t_2 and t_3 hindsight articles). Furthermore, hindsight bias in t_3 articles was greater for disasters than for nondisasters in terms of the suggestiveness and the foreseeability measures. Compared to the findings by Oeberst et al. (2018), who obtained hindsight bias only for t_3 articles about disasters, we thus obtained even more global evidence for the existence of hindsight bias in Wikipedia articles. Importantly, in contrast to the field study by Oeberst et al. (2018), the present research used substantially more events and relied on independent ratings of each article version, suggesting that the present results might be more reliable. Nevertheless, the varying extent to which hindsight bias was observed in the two studies makes it all the more

clear that it is a worthwhile endeavor to understand the mechanisms behind hindsight bias in written work. The present study already included a preliminary examination of the effect of two of the proposed determinants: the extent to which the event in question elicits the motivation to find an explanation (H_1) and the availability of causal information concerning the event (H_2).

The extent to which the respective events elicited sense-making motivation—operationalized as the average increase in page views of equivalent English-language Wikipedia articles after the event occurrence—was not associated with the magnitude of article hindsight bias. However, we found a strong association between the availability of causal information—assessed with the extent to which explanatory content was added to the articles—and the magnitude of article hindsight bias. This could be interpreted as the first tentative support for an effect of the availability of causal information. Importantly, however, the findings concerning these two determinants should not be given too much weight for at least two reasons. First, correlation estimates were based on a very small sample of 40 events (for the availability of causal information) or even only 31 events (for the sense-making motivation) and might thus be inaccurate (Schönbrodt & Perugini, 2013). Second, neither the sense-making motivation nor the availability of causal information was experimentally controlled, and the two operationalizations could also comprise other variables. For instance, the increase in page views might not only capture the sense-making motivation for the specific events but also a more general motivation to keep up with the news. Similarly, that no explanatory content was added to an article might not necessarily imply that there was no causal information available but could also signify a low motivation to include existing causal information. After all, the three proposed mechanisms occur in temporal sequence and are thus dependent on the preceding mechanisms (i.e., if there is no sense-making motivation, the availability of causal information is irrelevant; if there is no causal information available, content policies are irrelevant). Therefore, we followed up on this field study with three experimental studies, each systematically manipulating one of the three determinants.

Study 2—Sense-Making Motivation

The second study was done to test whether the extent to which an event motivates sense-making processes affects the magnitude of hindsight bias in written work about the respective event (H_1). To this end, we asked participants to write an article about a fictitious dam based on several fictitious newspaper articles. The sense-making motivation was manipulated by informing one group of participants about the occurrence of a *major* disaster at the dam (i.e., high sense-making motivation) and a second group learned about the occurrence of a *minor* disaster (i.e., low sense-making motivation). A third group

⁵ Additional simple effects analyses also signified that the significant main effect of event category for the suggestiveness and foreseeability measures was fully driven by the interaction effect: Only the t_1 articles about disasters were less suggestive of the incident, $M_{\text{disaster}} = 2.58$, $SD = 0.71$, $M_{\text{nondisaster}} = 1.68$, $SD = 0.72$, $F(1, 38) = 15.96$, $p < .001$, $\eta_p^2 = .30$, and expressed a lower foreseeability than articles about nondisasters, $M_{\text{disaster}} = 2.47$, $SD = 0.67$, $M_{\text{nondisaster}} = 1.64$, $SD = 0.75$, $F(1, 38) = 13.64$, $p = .001$, $\eta_p^2 = .26$. For the t_3 versions, however, neither the suggestiveness, $M_{\text{disaster}} = 3.70$, $SD = 0.68$, $M_{\text{nondisaster}} = 3.82$, $SD = 0.64$, $F(1, 38) = 0.34$, $p = .565$, $\eta_p^2 = .01$, nor the foreseeability measures differed, $M_{\text{disaster}} = 2.99$, $SD = 0.59$, $M_{\text{nondisaster}} = 2.73$, $SD = 1.00$, $F(1, 38) = 1.04$, $p = .314$, $\eta_p^2 = .03$.

learned that no disaster, but another event occurred (serving as an additional exploratory condition, see below), and a final, fourth group received no event knowledge (serving as an unbiased foresight group). We expected hindsight bias to be greater in articles of participants learning of a major disaster than in articles of participants learning of a minor disaster. We note that the newspaper articles contained sufficient causal information and participants were asked to adhere to content policies while writing the articles, so that this study allows us to examine the unique effect of the sense-making motivation.

Method

Participants and Design

Via an online recruitment system (called ORSEE; Greiner, 2015), we sampled 270 persons to participate in our study in the Mainz Behavioral and Experimental Laboratory in exchange for course credit or a salary of 9€/h (we preregistered $N \geq 260$; see <https://aspre-dicted.org/xe8au.pdf>).⁶ As preregistered, we excluded participants who did not complete the study ($n = 5$), did not consent to data usage ($n = 3$), or participated in prior studies that used the same material ($n = 2$), yielding a final sample size of $N = 260$ (183 women, 76 men, 1 nonbinary; $M_{\text{age}} = 21.97$, $SD = 2.99$; 256 undergraduates). Participants were randomly assigned to one of four experimental conditions (*foresight*, $n = 67$; *hindsight minor disaster*, $n = 66$; *hindsight major disaster*, $n = 63$; *hindsight no disaster*, $n = 64$; the latter for exploratory purposes, see <https://osf.io/8dfuh/>, for the additional exploratory analyses).

Materials and Pretest

As a source of information for participants' articles, we used a set of 11 bogus newspaper articles about a fictitious dam in Spain, which had successfully induced hindsight bias in previous studies (Meuer, Nestler, & Oeberst, 2021; Oeberst et al., 2018). The short, independent articles covered various news about the dam (e.g., construction process, safety precautions, public discourse on safety concerns) and provided arguments both for and against the occurrence of an incident at the dam. To manipulate participants' sense-making motivation, we constructed three additional articles that each informed about the occurrence of a different incident at the dam with varying degrees of negative consequences (see Heine et al., 2006; Park, 2010): (a) *no disaster* (partial collapse without any casualties; intended to serve as an additional exploratory condition), (b) *minor disaster* (partial collapse with two casualties; intended to serve as an event eliciting a low sense-making motivation), (c) *major disaster* (total collapse with over 80 casualties; intended to serve as an event eliciting a high sense-making motivation). A pretest with $N = 160$ participants (126 women, 34 men; $M_{\text{age}} = 25.74$, $SD = 7.76$; 157 undergraduates) showed that the three articles indeed elicited different levels of sense-making motivation in readers: Participants read one of the three articles and answered three questions constructed to assess their sense-making motivation for the incident (e.g., "How important is it to you to find out how this incident could have happened?," scale from 1 = *not at all* to 5 = *very*, Cronbach's $\alpha = .84$). As expected, the major disaster elicited a higher sense-making motivation, $M_{\text{major}} = 2.92$, $SD = 0.82$, than both the minor disaster, $M_{\text{minor}} = 2.53$, $SD = 0.76$, $t(105) = 2.52$, $p = .013$, $d = 0.48$, and the incident with no

disaster, $M_{\text{no}} = 2.45$, $SD = 0.87$, $t(102) = 2.87$, $p = .005$, $d = 0.56$. There was no significant difference between the minor disaster and the no disaster condition, $t(107) = 0.57$, $p = .569$, $d = 0.11$. In sum, the three incidents are suitable to examine two research questions: First, comparing hindsight bias in articles of authors who learned about the minor disaster versus major disaster is the most direct test to examine whether a higher sense-making motivation leads to a higher hindsight bias in written work (H_1) while controlling for the type of event (i.e., both incidents qualify as disasters but vary in the extent to which they elicit sense-making motivation). Second, comparing hindsight bias in articles of authors who learned about the incident with no disaster versus the minor disaster allows us to explore whether hindsight bias in written work is greater for the specific event category *disasters* while controlling for the sense-making motivation (i.e., whether other features of disasters such as the presence of human loss affects the magnitude of hindsight bias; see Study 1; Oeberst et al., 2018).

Procedure

After providing informed consent, participants received a printed booklet containing the 11 newspaper articles about the dam in nonchronological order. Participants in the three hindsight conditions additionally received one of the pretested articles informing about the occurrence of an incident with no (*hindsight no disaster*), two (*hindsight minor disaster*), or over 80 casualties (*hindsight major disaster*). Participants had about 15 min to read the material ($M = 17.21$ min, $SD = 8.22$). After reading, participants learned about two common content policies (i.e., NPOV and verifiability; "Wikipedia: Neutral point of view", 2021; "Wikipedia: Verifiability", 2021)⁷ and needed to correctly answer one test question about each policy before proceeding. Next, participants were asked to write a "Wikipedia-like" article about the dam using an online writing tool (www.publishwith.me) that provided basic text editing options. Participants were asked to complete their articles within 45 min (but they could take a few minutes longer if necessary; actual time taken: $M = 43.01$ min, $SD = 12.55$). While writing, participants could always consult the printed booklet with all newspaper articles. Subsequently, we assessed several additional measures for secondary analyses and exploratory purposes (e.g., participants' individual hindsight bias; see <https://osf.io/8dfuh/>, for a full survey export) as well as demographic data. Finally, after a full debriefing, participants stated whether they had participated in prior studies using the same material and whether they wanted to retract their data (both serving as exclusion criteria).

⁶ Sample size was determined based on an a priori power analysis with G*Power (Faul et al., 2007). For the preregistered independent two-tailed t tests, a minimum of 64 participants per condition (total $N \geq 256$) was needed to detect moderate effects with sufficient power, $d = .50$, $\alpha = .05$, $1 - \beta = .80$. Taking potential data exclusions into account, we preregistered $N = 260$. Note, that we decided to analyze the data using linear mixed models to account for the pseudoreplication introduced by the multiple foresight ratings, which we oversaw during the preregistration (see Footnote 11). However, the mixed model implements a type of dependent test, so we do not believe that this decision negatively affects our power calculations.

⁷ We did not explicitly introduce the content policy to include no original research ("Wikipedia: No original research", 2021), but the policy to include only verifiable information already excluded opinions and allegations, and since the event was fictive, the only reliable source of information was the provided material.

Article Coding

Ten trained, independent coders, who were blind to the experimental conditions and the research question, rated all written articles on several indicators for hindsight bias. To obtain a normative measure of hindsight *bias* (i.e., erroneous deviations of the hindsight perspective from the foresight perspective), one needs to compare the ratings of the foresight and hindsight articles. However, the articles in the three hindsight conditions were based on event knowledge about three different incidents that likely vary in their perceived a priori probability (e.g., a partial collapse of a dam might generally be considered more likely than a total collapse). Therefore, the foresight–hindsight comparisons had to be incident-specific and coders needed to rate each article with respect to the specific incident in question. Specifically, for articles in the three hindsight conditions, coders gave ratings with regard to the specific incident that the authors were informed about (e.g., an article written by a participant who learned about a major disaster with over 80 casualties was rated with respect to the extent to which it suggested this specific incident; but to ensure coder blindness, coders were told that the to-be-rated incidents were assigned randomly). Since no specific incident was presented in the foresight condition, articles in the foresight condition needed to be judged once for each of the three incidents. Therefore, coders rated each foresight article three separate times.

The resulting set of $N = 394$ article evaluations (i.e., all hindsight articles once, all foresight articles thrice) was provided in a partially random order that made it extremely unlikely that the same foresight article was presented multiple times in a row, which should reduce dependencies between the three ratings. Specifically, all article evaluations were randomly assigned to blocks of 15 evaluations. We then manually checked whether a block contained multiple evaluations of the same foresight article, which was the case eight times. For these cases, we randomly selected one of the duplicate foresight articles and swapped it with a randomly selected article from the other blocks (with the exception that the swapped article could not be an article that was already in this block, as this would not have solved the issue). Thus, none of the final blocks contained the same foresight articles multiple times. The blocks were then provided in fixed order (which eased the administration of the coding process), and coders had to take a 30-min break after each block. The articles per block were presented in random order.⁸ Coding was implemented using the *formr* survey framework (Arslan et al., 2018, 2020).

All article evaluations comprised four dependent variables: (a) the general suggestiveness of the specific incident (“To what extent does the article suggest this incident?”), whether the articles implied that the specific incident was (b) inevitable and (c) foreseeable (“To what extent does the article suggest that the incident was inevitable/foreseeable?”; all on scales from 1 = *not at all* to 5 = *very much*), and (d) the extent to which the articles suggested that the specific incident was likely to happen. The latter rating was done by assigning a total of 100% to the three pretested incidents plus a scenario without the occurrence of an incident.⁹ Ratings of all 10 coders were averaged for the main analyses. Two-way random effects intraclass correlation coefficients (ICC) indicated moderate-to-good interrater consistencies for all four hindsight measures, $ICC_{\text{suggestiveness}}(2, 10) = .84$, $ICC_{\text{inevitability}}(2, 10) = .68$, $ICC_{\text{foreseeability}}(2, 10) = .81$, and $ICC_{\text{likelihood}}(2, 10) = .90$.

Results

To test our hypothesis, we used the ratings of articles written in the *hindsight minor disaster* and the *hindsight major disaster* conditions, as well as the two corresponding incident-specific ratings of articles written in the *foresight condition* (i.e., the judgments concerning the minor disaster and the major disaster), hence yielding a 2 (judged incident: minor vs. major disaster) \times 2 (author perspective: foresight vs. hindsight) design.¹⁰ We analyzed the data using a linear mixed model because these models allowed us to account for the dependency of the multiple foresight ratings for the same articles (i.e., article is the Level 2 variable). Specifically, for each of the four dependent variables (suggestiveness, inevitability, foreseeability, likelihood), we estimated a random intercept-mixed model in which the judged incident (minor vs. major disaster) and the author perspective (foresight vs. hindsight), as well as their interaction, were included as fixed effects, and article ID was included as a random intercept effect. All model parameters were estimated with a restricted maximum likelihood approach using the *lme4* package (Bates et al., 2015) in *R*. Estimates of degrees of freedom were obtained with Satterthwaite’s method implemented in the *lmerTest* package (Kuznetsova et al., 2017).¹¹

The analysis yielded a significant main effect of author perspective for all four dependent measures, $F_{\text{suggestiveness}}(1, 182.83) = 29.97$, $p < .001$, $F_{\text{inevitability}}(1, 180.44) = 35.37$, $p < .001$, $F_{\text{foreseeability}}(1, 177.78) = 34.58$, $p < .001$, and $F_{\text{likelihood}}(1, 179.25) = 35.00$, $p < .001$. Articles written in hindsight (i.e., with event knowledge) were more suggestive of the presented incident than articles written in foresight, signifying classic hindsight bias (see Figure 2). There was no significant main effect of the judged incident for any of the four hindsight measures, all $F_s(1, 233.86) < 1.79$, $p > .183$, signifying that there were no differences in the judgments regarding the minor and the major disasters. Most importantly, for the proposed hypothesis, the analysis yielded a significant interaction effect of author perspective and judged incident for the articles’ general suggestiveness of the given incident, $F(1, 235.78) = 7.93$, $p = .005$, and for

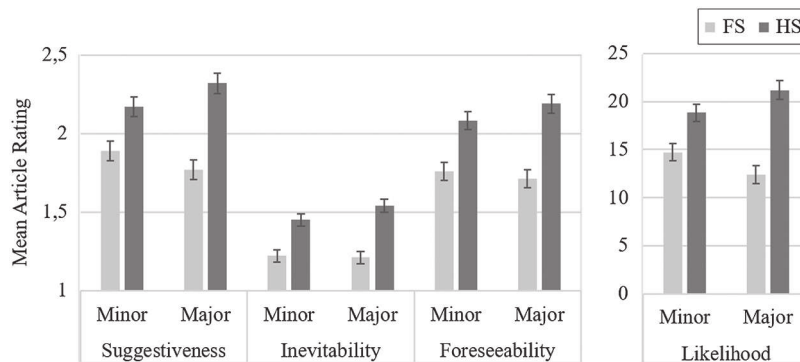
⁸ Note that this procedure does not entirely rule out that the same foresight article was evaluated two times in a row (i.e., in nine cases, the same foresight article was in consecutive blocks, and, for these cases, the random order within each block might have assigned the respective article to the last position in the first block and the first position in the second block). However, given our assignment procedure, this event was *extremely* unlikely to occur.

⁹ Besides the suggestiveness, inevitability, and foreseeability measures we already used in Study 1, we took advantage of the design entailing the three different incidents and additionally included a likelihood measure that closely resembles the hindsight bias operationalization utilized in classic work on hindsight bias (e.g., Fischhoff, 1975).

¹⁰ The results of the pretest suggested that a comparison of hindsight bias after learning about the minor versus the major disaster is the most suitable test to examine whether the elicitation of sense-making motivation affects the magnitude of article hindsight bias. In the Supplemental Analysis on OSF (see <https://osf.io/8dfuh/>), we additionally compare hindsight bias after learning about the no disaster versus the minor disaster incident to test whether other features of the event category *disasters* affect the size of article hindsight bias.

¹¹ We preregistered to first compute independent *t* tests comparing the ratings of foresight versus hindsight articles for each incident separately and to then test for differences in effect sizes for both incidents. However, this procedure would have introduced pseudoreplication as the foresight ratings for both the minor and major disasters refer to the same set of foresight articles (i.e., written by participants in the foresight condition). We thus decided to account for the partially dependent data by estimating mixed models.

Figure 2
Hindsight Measures as a Function of Author Perspective and Judged Incident



Note. Error bars are standard errors. Minor = minor disaster incident; Major = major disaster incident; FS = foresight perspective; HS = hindsight perspective.

the extent to which the articles suggested that the incident was likely, $F(1, 243.20) = 11.30, p < .001$. For the inevitability and foreseeability measures, the interaction effect did not reach statistical significance according to the preregistered level of $\alpha = .05$, $F_{\text{inevitability}}(1, 233.86) = 3.61, p = .059$, $F_{\text{foreseeability}}(1, 249.92) = 3.55, p = .061$.

To elucidate the significant interaction effects for the suggestiveness and likelihood measures, we applied simple effects tests using the *emmeans* package in *R* (Russel et al., 2021) and examined the foresight–hindsight differences (i.e., hindsight bias) separately for each incident. For both hindsight measures, we found a significant hindsight bias in articles written by authors who were informed about the occurrence of a minor disaster, $\Delta M_{\text{suggestiveness}} = 0.28, t(199) = 3.15, p = .002$, $\Delta M_{\text{likelihood}} = 4.11, t(199) = 3.19, p = .002$, but an even greater hindsight bias in articles written by authors who read about a major disaster, $\Delta M_{\text{suggestiveness}} = 0.55, t(199) = 6.10, p < .001$, $\Delta M_{\text{likelihood}} = 8.82, t(199) = 6.78, p < .001$. Thus, the extent to which a particular event motivates sense-making affected the magnitude of hindsight bias in terms of the general suggestiveness and the likelihood measure. We did not obtain such an effect for the inevitability and foreseeability measures when applying the preregistered significance level, but we note that, descriptively, the effect was in the same direction.¹²

Discussion

In this experimental study, we examined whether the magnitude of hindsight bias in written work is contingent on the extent to which the respective event motivates sense-making processes. First of all, articles written with event knowledge (i.e., the partial or total collapse of the dam) contained hindsight bias, that is, they were more suggestive of the event and implied to a greater extent that the event was inevitable, foreseeable, and likely than articles written without event knowledge. This is consistent with the findings of Study 1 and the studies by Oeberst et al. (2018) and provides further support for the existence of hindsight bias in written work. More importantly, the extent to which the presented event evoked sense-making processes affected the magnitude of article hindsight bias with regard to two of the four measures: For the suggestiveness and the likelihood measures, the deviation from the foresight articles was

higher in hindsight articles written by participants who learned about the total collapse (i.e., the event that elicited a high sense-making motivation) than in hindsight articles about the partial collapse (i.e., the event eliciting a low sense-making motivation). This finding is in line with causal model theory of hindsight bias (Blank & Nestler, 2007): If the sense-making motivation for an event is high, people will engage in pronounced evidence sampling and integration, which, however, will be biased by the salient event knowledge (Pezzo, 2003; Roese & Olson, 1996; Slovic & Fischhoff, 1977). Once established, such event-consistent causal representations will likely be also transferred to written work.

Also, descriptively, the effects on the inevitability and foreseeability measures were in the same direction, and the study thus yielded an overall consistent pattern of results for all four measures of hindsight bias. One potential reason for why the manipulation might have affected the inevitability and foreseeability measures to a lesser extent is that these measures are conceptually more specific than the suggestiveness and the likelihood measures: While only certain types of information might affect the extent to which an article expresses the inevitability (e.g., event-consistent antecedents) and foreseeability of an event (e.g., warnings from experts), a much wider scope of information can influence the general suggestiveness and the expressed likelihood of the event. Coders may thus have been more reluctant to take a given piece of information as an indication of increased inevitability or foreseeability.

Yet, the coding procedure we applied might have theoretically introduced demand characteristics that could have led to the interaction effect we obtained. Specifically, while hindsight articles were only coded once (with regard to the specific incident that the authors were informed about), foresight articles were coded thrice (once for each incident), and it was, thus, possible that the three foresight ratings per article were rated relative to one another. For instance, let us assume that coders initially rated a foresight article with respect to

¹² To test for the robustness of our findings, we repeated the main analyses of all three experimental studies (i.e., Studies 2–4) while controlling for the length of the produced articles (as coders might have unwarrantedly used article length as a heuristic for their ratings). However, for all three studies, the pattern of results was identical when including article length as a covariate, rendering a confounding effect of article length unlikely. A detailed report of this exploratory analysis can be found on OSF (see <https://osf.io/8dfuh/>).

the minor disaster and then were later asked to rate the same article again with respect to the major disaster. In such a case, if the coders recalled their first article evaluation, they might feel compelled to rate the major disaster as comparably less likely. Our finding that the foresight–hindsight differences (i.e., hindsight bias) were greater for the articles of the major (vs. minor) disaster condition might then be also driven by artificially decreased foresight ratings (rather than by a greater manifestation of hindsight bias in the respective hindsight articles). However, we deem it very unlikely that such demand characteristics contributed to our results. After all, we arranged the order of the article evaluations in a way that the three evaluations of the same foresight articles were scattered across the total set of 394 evaluations. Since coders took several weeks to perform all ratings, it is thus likely that the three evaluations were made—on average—several days apart. In addition, all coded articles were based on a nearly identical set of newspaper articles and were thus very similar, which makes it even more difficult to reliably recognize previously rated articles, let alone remember the specific rating and what incident the previous rating was about. Therefore, even though demand effects would be *theoretically* possible, we do not think that they *actually* affected the coding to such an extent that they alone could explain the large interaction effects we obtained for the suggestiveness and the likelihood measures.

Study 3—Availability of Causal Information

In the next experimental study, we examined whether the magnitude of hindsight bias in written work is contingent on the availability of causal information (H_2). Therefore, we again asked (novel) participants to write an article about the fictitious dam we used in Study 2, but this time, we manipulated the information provided in the source newspaper articles. Participants received a set of source articles with either low or high causal relevance for a disaster at the dam. We also manipulated whether or not participants were informed about the occurrence of a disaster (foresight vs. hindsight). If the availability of causal information determined the magnitude of hindsight bias in written work, articles based on information with high causal relevance should be more biased than articles based on information with low causal relevance. Again, to extract the unique effect of the availability of causal information, all participants were asked to adhere to content policies that authors need to follow on Wikipedia and other information outlets.

Method

Participants and Design

A total of 206 persons participated in our study in exchange for course credit or a salary of 8€/h (we preregistered $N \geq 160$; see <https://aspredicted.org/an2te.pdf>).¹³ As preregistered, we excluded participants who did not complete the study ($n = 1$), did not consent to data usage ($n = 1$), or participated in prior studies that used the same material ($n = 10$), resulting in a final sample size of $N = 194$ (150 women, 43 men, 1 nonbinary; $M_{\text{age}} = 24.89$, $SD = 6.31$; 191 undergraduates). Participants were randomly assigned to one of four experimental conditions receiving different sets of newspaper articles as a source of information for their article, yielding a 2 (perspective: foresight vs. hindsight) \times 2 (causal relevance: low vs. high) between-subjects design (*foresight, low causal relevance*, $n = 49$; *foresight,*

high causal relevance, $n = 50$; *hindsight, low causal relevance*, $n = 47$; *hindsight, high causal relevance*, $n = 48$).

Materials and Pretest

For the high causal relevance conditions, we again used the set of 11 bogus newspaper articles about the fictitious dam in Spain from Study 2, but we needed to construct an additional version of this material set with less causal information to manipulate the availability of causal information. To this end, in each newspaper article, we systematically replaced (presumably) causal information with similar but more neutral text passages while adhering to the original material as much as possible. For instance, while one original article informed about public protests against the dam due to *safety concerns*, the adapted article reported protests due to *nature protection concerns* (see <https://osf.io/8dfuh/>, for a side-by-side comparison). A pretest with $N = 152$ separate participants (122, women, 30 men; $M_{\text{age}} = 24.87$, $SD_{\text{age}} = 8.84$; 133 undergraduates) showed that the two sets of material indeed differed in their amount of available causal information: Participants first read an article informing about the collapse of the dam with over 80 casualties (i.e., the *major disaster incident* of Study 2) and then received one of the two newspaper article sets divided into a total of 23 text passages. Participants judged the extent to which each passage was causally relevant for the explanation of the presented incident (on a 7-point scale; $-3 = \text{strong argument against the occurrence of the incident}$, $0 = \text{not relevant at all}$, $3 = \text{strong argument for the occurrence of the incident}$). We averaged the absolute values of the 23 passages (Cronbach's $\alpha = .79$) and computed an independent t test to test for a difference between the two material sets. As expected, the adapted material set was rated to be causally less relevant than the original material set, $M_{\text{adapted}} = 0.68$, $SD = 0.31$, $M_{\text{original}} = 0.98$, $SD = 0.28$, $t(150) = 6.32$, $p < .001$, $d = 1.03$. Thus, the two material sets were suitable to manipulate the amount of accessible causal information.

Procedure

After providing informed consent, participants were randomly assigned to receive a printed booklet containing either the original (i.e., high causal relevance) or the adapted (i.e., low causal relevance) set of 11 newspaper articles about the dam. In each group, half of the participants additionally received the article informing about the collapse of the dam, leading to over 80 casualties, the other half was not informed about this event (i.e., hindsight vs. foresight perspective). As in Study 2, participants then had about 15 min to read the material ($M = 14.73$ min, $SD = 6.14$), learned the same two content policies via test questions, and had about 45 min to write a “Wikipedia-like” article (but could take a few minutes longer if needed; actual time taken: $M = 35.91$ min, $SD = 14.42$). While writing, participants could always consult the printed source material. Subsequently, we assessed several additional exploratory measures and demographic data (see <https://osf.io/8dfuh/>, for a full survey export). Finally, we debriefed all participants and obtained two

¹³ Based on an a priori power analysis with G^* power, the preregistered 2×2 between-subjects ANOVA required $N \geq 128$ participants to detect moderate effects with sufficient power, $\alpha = .05$, $1 - \beta = .80$, $f = .25$ (which equals $\eta_p^2 = .06$). However, to allow for stable correlation coefficients for additional analyses, we preregistered $N = 160$.

exclusion criteria (i.e., participation in prior studies with similar material, refusal to use their data).

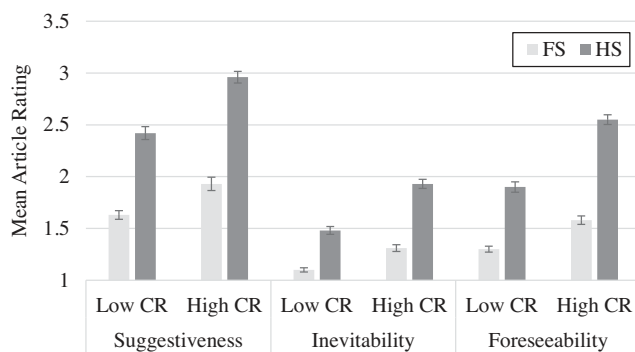
Article Coding

Ten trained, independent coders, who were blind to the experimental conditions and the research question, judged all written articles on three dependent measures: They rated the general suggestiveness of a disaster (“To which extent does the article suggest a disaster at the dam?”), the displayed inevitability, and the displayed foreseeability of the disaster (“To which extent does the article suggest that a disaster at the dam was inevitable/foreseeable?”).¹⁴ All ratings were made on 5-point scales (1 = *not at all* to 5 = *very much*), and for the main analyses, the ratings of all coders were averaged for each measure, which featured good interrater consistencies, $ICC_{\text{suggestiveness}}(2, 10) = .88$, $ICC_{\text{inevitability}}(2, 10) = .77$, $ICC_{\text{foreseeability}}(2, 10) = .86$. Coding was implemented using the *formr* survey framework (Arslan et al., 2018, 2020). All articles were randomly assigned to blocks of 15 articles. The blocks were then provided in fixed order and the articles per block were presented in random order. Coders had to take a 30-min break after each block to avoid fatigue.

Results

We computed a separate 2 (perspective: foresight, hindsight) \times 2 (causal relevance: low, high) between-subjects ANOVA for each of the three hindsight measures. For all three dependent variables, the analysis yielded a significant main effect of perspective, $F_{\text{suggestiveness}}(1, 190) = 256.20$, $p < .001$, $\eta_p^2 = .57$, $F_{\text{inevitability}}(1, 190) = 202.95$, $p < .001$, $\eta_p^2 = .52$, $F_{\text{foreseeability}}(1, 190) = 346.61$, $p < .001$, $\eta_p^2 = .65$. Articles written in hindsight were more suggestive of the event than articles written in foresight (i.e., classic hindsight bias; see Figure 3). Also, there was a significant main effect of causal relevance for all three hindsight measures, $F_{\text{suggestiveness}}(1, 190) = 53.81$, $p < .001$, $\eta_p^2 = .22$, $F_{\text{inevitability}}(1, 190) = 89.46$, $p < .001$, $\eta_p^2 = .32$, $F_{\text{foreseeability}}(1, 190) = 122.32$, $p < .001$, $\eta_p^2 = .39$. Articles based on the material with high causal relevance were more suggestive of the event and expressed a higher inevitability and foreseeability of the event than articles based on the material with low causal relevance. Furthermore, we obtained a

Figure 3
Hindsight Measures as a Function of Perspective and Causal Relevance



Note. Error bars are standard errors. CR = causal relevance; FS = foresight perspective; HS = hindsight perspective.

significant interaction effect of perspective and causal relevance for all three measures, $F_{\text{suggestiveness}}(1, 190) = 4.24$, $p = .041$, $\eta_p^2 = .02$, $F_{\text{inevitability}}(1, 190) = 11.96$, $p = .001$, $\eta_p^2 = .06$, and $F_{\text{foreseeability}}(1, 190) = 20.07$, $p < .001$, $\eta_p^2 = .10$.

Simple effects tests comparing the foresight–hindsight differences separate for both levels of causal relevance yielded a significant hindsight bias for all three measures in articles based on information with only low causal relevance, $\Delta M_{\text{suggestiveness}} = 0.80$, $F(1, 190) = 96.28$, $p < .001$, $\eta_p^2 = .34$, $\Delta M_{\text{inevitability}} = 0.38$, $F(1, 190) = 57.59$, $p < .001$, $\eta_p^2 = .23$, $\Delta M_{\text{foreseeability}} = 0.59$, $F(1, 190) = 98.91$, $p < .001$, $\eta_p^2 = .34$, but an even greater hindsight bias in articles based on information with high causal relevance, $\Delta M_{\text{suggestiveness}} = 1.03$, $F(1, 190) = 164.86$, $p < .001$, $\eta_p^2 = .47$, $\Delta M_{\text{inevitability}} = 0.62$, $F(1, 190) = 158.35$, $p < .001$, $\eta_p^2 = .46$, $\Delta M_{\text{foreseeability}} = 0.97$, $F(1, 190) = 269.54$, $p < .001$, $\eta_p^2 = .59$. Therefore, the magnitude of article hindsight bias was contingent on the availability of causal information—consistently for all three hindsight measures.

Discussion

We conducted this study to test whether the availability of causal information concerning an event affects the magnitude of hindsight bias in written work about the event. As in Study 2, we obtained strong evidence for hindsight bias in articles written with event knowledge: Both the hindsight articles based on source information with high causal relevance and the hindsight articles based on information with low causal relevance were more suggestive of the event and expressed a higher inevitability and foreseeability of the event than the respective foresight articles. Most importantly, however, consistently for all three operationalizations of hindsight bias, the bias was higher for articles based on information with high causal relevance than for articles based on information with only low causal relevance.

This finding is consistent with numerous demonstrations that people’s hindsight bias is dependent on the availability of causal information (e.g., Arkes et al., 1988; Carli & Leonard, 1989; Nario & Branscombe, 1995; Nestler et al., 2008; Nestler & von Collani, 2008; Wasserman et al., 1991; Yopchick & Kim, 2012) and provides first support that this contingency is also evident in written work. The latter is particularly noteworthy given that all our participants had to adhere to content policies, which precluded a straightforward generalization from individual hindsight bias to article hindsight bias (Nestler et al., 2017). Seemingly, the highly causally relevant newspaper articles equipped authors with more or stronger *verifiable* event-consistent antecedents that comply with the content policies. Thus, according to causal model theory (Blank & Nestler, 2007), this material set offered more opportunities to construct an unbalanced event-consistent written representation of the event. Importantly, compared to the material set with low causal relevance, the set with high causal relevance not only contained more and stronger event-consistent information but also (equally) more and stronger event-inconsistent information, so that the effect of causal relevance cannot be attributed to more biased source information that is simply mirrored in the articles.

¹⁴ In contrast to Study 2, raters were not asked to assign likelihoods to four mutually exclusive events as there was only one incident presented (i.e., total collapse of the dam).

Study 4—Content Policies

So far, we have examined the magnitude of hindsight bias in written work in relation to two characteristics of the event in question. In the final experimental study, we turned to the potential effect of a characteristic of the writing context. Specifically, we tested whether the provision of content policies that call for an NPOV and the verifiability of included information can mitigate hindsight bias in written work (H_3). Recall that, in Studies 2 and 3, we had provided instructions to all participants to adhere to the NPOV and the verifiability policy. In this study, we experimentally varied the presence of the policies to examine their effect on article hindsight bias. Also, we decided to adapt the experimental setting used in Studies 2 and 3 with regard to two important aspects. First, to allow for the assessment of policy-consistent behavior, we gave all participants an article draft that contained various violations of the policies (i.e., unverifiable speculations) and asked them to *edit and finalize* this article (rather than writing it from scratch). Second, to further encourage adherence to the content policies, we asked participants to edit the article in *pairs* (rather than alone), so that each pair of participants could discuss whether (parts of) the article violated the policies. Importantly, these two adaptations were also made to more closely mimic the writing context in Wikipedia, where the lion's share of work consists of editing already existing articles and where compliance with the content policies is collaboratively negotiated. Thus, we asked dyads of participants to edit the article draft about the fictitious dam, which contained information, some of which could not be verified based on the given set of source newspaper articles. We manipulated whether or not participants received event knowledge and whether or not they learned the content policies. We expected hindsight bias in the edited articles to be greater when content policies were not presented (and thus there was no need to delete unverifiable information and write with an NPOV). Since discussions about adding or removing content often relate to specific pieces of information, we preregistered to test this hypothesis not only on a global article level (i.e., coding of each article as a whole, as in the prior studies) but also on a more fine-grained information level (i.e., coding of the inclusion of certain information in each article).

Method

Participants and Design

We sampled 290 individuals via the online recruitment system ORSEE (Greiner, 2015) to participate in our study. Participants were invited to the Mainz Behavioral and Experimental Laboratory and participated in exchange for course credit or a salary of 9€/h (we preregistered $N \geq 270$; see <https://aspredicted.org/mv48y.pdf>).¹⁵ As preregistered, we excluded participants who did not consent to data usage ($n = 7$) or participated in prior studies that used the same material ($n = 5$), as well as their respective writing partners ($n = 12$), yielding a final sample size of $N = 266$ (170 women, 94 men, 2 nonbinary; $M_{\text{age}} = 22.56$, $SD = 3.59$; 260 undergraduates). Participants were allocated a writing partner, and each dyad was randomly assigned as to whether or not they received event knowledge and whether or not they learned about the content policies. They were asked to edit and finalize an article draft that contained both event-consistent and event-inconsistent information, some of which was

verifiable and some was not, resulting in a 2 (perspective: foresight, hindsight; between-subjects) \times 2 (policies: not presented, presented; between-subjects) \times 2 (event-consistency of information: inconsistent, consistent; within-subjects) \times 2 (verifiability of information: unverifiable, verifiable; within-subjects) mixed design. Since editing was implemented in dyads, the analyses were based on $N = 133$ articles (*foresight, no policies*, $n = 34$; *foresight, policies*, $n = 33$; *hindsight, no policies*, $n = 32$; *hindsight, policies*, $n = 34$).

Materials and Procedure

After providing informed consent, participants received a printed booklet with 10 of the original newspaper articles we used in Studies 2 and 3 (we needed to omit one source to construct a balanced article for participants to work on, see below). Participants in the *hindsight conditions* additionally received the article informing about the collapse of the dam. We asked participants to read the material within about 15 min. After reading ($M = 9.99$ min, $SD = 5.25$), participants in the *policies presented conditions* learned about two common content policies (i.e., NPOV and verifiability; “Wikipedia: Neutral point of view”, 2021; “Wikipedia: Verifiability”, 2021) and could only proceed if they passed a test question about each policy. We then assigned all participants to a writing partner within the same experimental condition and instructed them to collaboratively edit a given article draft so that it could be published. Since all participants took part in the study on separate computers, writing partners had no direct personal contact while editing. Instead, editing took place in an online writing tool that allowed participants to chat with their partners (www.publishwith.me). Their task was to finish the article in about 30 min (but they could take a few minutes longer if necessary; actual time taken: $M = 37.47$ min, $SD = 4.90$). While editing, participants could always consult the printed source material.

All participants started with the same initial article draft containing an equal number of event-consistent, event-inconsistent, verifiable, and unverifiable information, which we had constructed based on the results of the pretest for Study 3 (see <https://osf.io/8dfuh/>, for the article). Specifically, the initial article contained three pieces of information that were rated to be *event-consistent* (i.e., arguments for the occurrence of a dam collapse) and three pieces of information that were rated to be *event-inconsistent* (i.e., speaking against a collapse). These six arguments could be found in the provided newspaper articles and were thus *verifiable*. On top of that, we added three more event-consistent and three more event-inconsistent pieces of information that could not be found in any of the sources. Specifically, they went beyond the verifiable part of each argument by containing *unverifiable* speculation (e.g., that cracks on the dam wall might indicate that it is not stable enough). In sum, the initial article thus contained 12 relevant pieces of information (i.e., three verifiable event-consistent, three verifiable event-inconsistent, three *unverifiable* event-consistent, three *unverifiable* event-inconsistent).

After editing, we obtained several additional exploratory measures and demographic data (see <https://osf.io/8dfuh/>, for a full survey

¹⁵ Based on an a priori power analysis with G^* power, the preregistered 2 \times 2 between-subjects ANOVA required $N \geq 128$ articles ($N \geq 256$ participants) to detect moderate effects with sufficient power, $\alpha = .05$, $1 - \beta = .80$, $f = .25$ (which equals $\eta_p^2 = .06$). Taking potential data exclusions into account, we preregistered $N = 270$.

export). We then debriefed all participants and assessed two exclusion criteria (i.e., participation in prior studies, refusal to use their data).

Article Coding

Ten trained, independent coders, who were blind to the experimental conditions and the research question, rated all edited articles in two consecutive rounds. In the first round, coders judged each article on the same three dependent measures as in Study 3 (general suggestiveness, inevitability, foreseeability). For the main analyses, we aggregated the ratings of all 10 coders for each measure, which featured good interrater consistencies, $ICC_{\text{suggestiveness}}(2, 10) = .88$, $ICC_{\text{inevitability}}(2, 10) = .86$, $ICC_{\text{foreseeability}}(2, 10) = .88$. Again, coding was implemented using the *formr* survey framework (Arslan et al., 2018, 2020). All articles were randomly assigned to blocks of 20 articles. The blocks were then provided in fixed order and the articles per block were presented in random order. Coders had to take a 30-min break after each block.

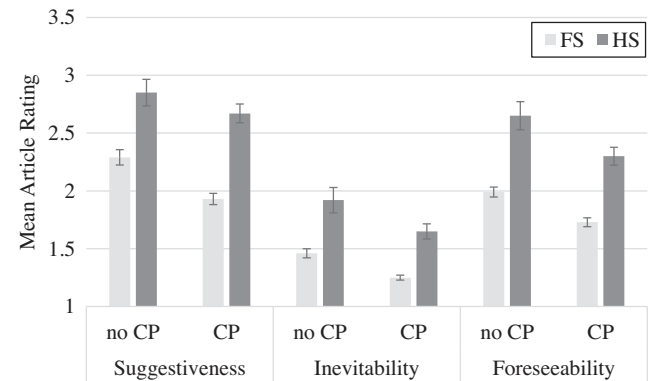
In the second round, coders were asked to indicate the extent to which each of the 12 pieces of information that were included in the initial article draft was still present in the final article (on a 4-point scale; 0 = *not included*, 1 = *included but attenuated*, 2 = *included without alteration*, 3 = *included and expanded*). For the main analyses, we then dichotomized these ordinal-scaled ratings to allow for the preregistered metric analysis of the number of deleted pieces of information (i.e., values “1”–“3” were merged into the category *included*, value “0” remained *not included*)¹⁶ and registered each information as included if five or more coders had rated it as such, interrater agreements were very high; 93.92% of all classifications were based on an agreement of 9 or 10 coders; average interrater consistency: $ICC(2, 10) = .96$. For each of the four types of information (e.g., verifiable outcome-consistent), we summed up the corresponding pieces of information registered as *included*, resulting in a score of 0–3. For this coding round, all articles were randomly assigned to blocks of 10 articles, which were provided in fixed order. As in the first round, the articles per block were presented in random order, and coders had to take a 30-min break after each block.

Results

First, we tested for an effect of the content policies at the article level, drawing on the global judgment dimensions that we already used in the previous studies. If content policies determined the magnitude of hindsight bias in written work, hindsight bias should be greater in articles of participant dyads who did not receive these policies. We tested this assumption with a separate 2 (perspective: foresight, hindsight) \times 2 (policies: not presented, presented) between-subjects ANOVA for each of the three dependent measures. The analysis yielded a significant main effect of perspective for all three hindsight measures, $F_{\text{suggestiveness}}(1, 129) = 64.26, p < .001, \eta_p^2 = .33$, $F_{\text{inevitability}}(1, 129) = 41.46, p < .001, \eta_p^2 = .24$, $F_{\text{foreseeability}}(1, 129) = 64.43, p < .001, \eta_p^2 = .33$, signifying hindsight bias in all articles written with event knowledge (see Figure 4). Furthermore, we obtained a significant main effect of policies for all dependent variables, $F_{\text{suggestiveness}}(1, 129) = 11.01, p = .001, \eta_p^2 = .08$, $F_{\text{inevitability}}(1, 129) = 13.47, p < .001, \eta_p^2 = .10$, $F_{\text{foreseeability}}(1, 129) = 15.79, p < .001, \eta_p^2 = .11$. Articles written

Figure 4

Hindsight Measures as a Function of Perspective and Content Policies



Note. Error bars are standard errors. no CP = content policies not presented; CP = content policies presented; FS = foresight perspective; HS = hindsight perspective.

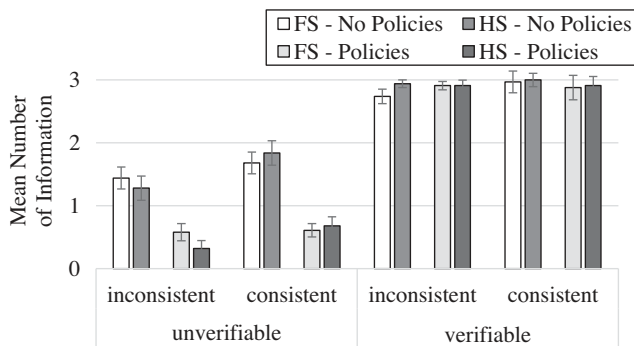
by dyads who received the policies were less suggestive of the incident and expressed a lower inevitability and foreseeability than articles written without knowledge about the policies. However, there was no significant interaction effect of perspective and policies for any of the hindsight measures, $F_{\text{suggestiveness}}(1, 129) = 1.25, p = .266, \eta_p^2 = .01$, $F_{\text{inevitability}}(1, 129) = 0.27, p = .606, \eta_p^2 = .01$, $F_{\text{foreseeability}}(1, 129) = 0.30, p = .583, \eta_p^2 < .01$. Therefore, we found no support that the provision of content policies affected the magnitude of hindsight bias in the articles.

To examine whether the provision of content policies affected the writing process and hindsight bias at the information level, we computed a 2 (perspective: foresight, hindsight; between-subjects) \times 2 (policies: not presented, presented; between-subjects) \times 2 (event-consistency of information: inconsistent, consistent; within-subjects) \times 2 (verifiability of information: unverifiable, verifiable; within-subjects) ANOVA with the number of included pieces of information as the dependent variable.

First of all, if author dyads who received the policies adhered to the policies correctly, they should delete more information in general (i.e., main effect of policies), and more unverifiable information in particular (i.e., Policies \times Verifiability interaction), compared to dyads who did not know the policies. The analyses indeed yielded a significant main effect of policies, $F(1, 129) = 40.92, p < .001, \eta_p^2 = .24$, and a significant interaction effect of policies and the verifiability of the information, $F(1, 129) = 51.78, p < .001, \eta_p^2 = .29$. More information was deleted when policies were presented, and according to simple effects tests, policies did only foster the deletion of unverifiable information, $\Delta M = 1.02, F(1, 129) = 55.93, p < .001, \eta_p^2 = .30$, but not the deletion of verifiable information, $\Delta M = 0.01, F(1, 129) = 0.02, p = .902, \eta_p^2 < .01$ (see Figure 5). Thus, the provision of content policies effectively reduced unverifiable content in the articles.

¹⁶ Since our main research question was to determine whether or not certain information was deleted from the articles, we preregistered to analyze the number of deleted pieces of information for each of the four types of information. However, we decided to let coders use a more fine-grained 4-point scale and dichotomize the scale for the main analysis. This allowed for potential exploratory follow-up analysis of the editing process besides the mere deletion of information (e.g., rephrasing or expansion of arguments).

Figure 5
Mean Number of Information Included in the Articles



Note. Error bars are standard errors. FS = foresight perspective; HS = hindsight perspective.

If the content policies also mitigated hindsight bias, this should be reflected in a significant four-way interaction of perspective, policies, verifiability, and consistency. The logic behind this is that article hindsight bias results from an imbalance of event-consistent and event-inconsistent information and that the policies prompt participants in the hindsight condition to settle this imbalance (e.g., by deleting unverifiable event-consistent information). However, the four-way interaction was not significant, $F(1, 129) = 0.34, p = .560, \eta_p^2 < .01$. This corresponds to the finding at the article level that the provision of content policies did not affect the magnitude of article hindsight bias.

Discussion

In this experimental study, we examined whether the provision of content policies can mitigate hindsight bias in collaboratively edited articles. Consistent with the previous studies, we found hindsight bias in all hindsight articles. Furthermore, author dyads who received the content policies produced articles that were less suggestive of the event and expressed a lower inevitability and foreseeability of the event than dyads who did not receive the policies. Yet, the provision of content policies did not affect the magnitude of hindsight bias as they equally affected the articles of dyads with and without event knowledge. This is also reflected in the findings at the information level: While the provision of content policies led to a deletion of most of the unverifiable information, we found no evidence that participants with and without event knowledge differed in how much unverifiable event-consistent and event-inconsistent information they deleted.

One possible explanation for why we could not find an effect of the content policies on the magnitude of article hindsight bias might be that certain violations of the policies might go unnoticed. Recall that hindsight bias, according to causal model theory (Blank & Nestler, 2007), can be attributed to a process of information selection and evaluation that is biased by the event knowledge. We had argued that the policies to include only verifiable information and to present information with an NPOV might mitigate hindsight bias because they forbid the inclusion of (event-consistent) information that is unverifiable and foster a more neutral evaluation of both event-consistent and event-inconsistent information, respectively. Yet,

successful compliance with the policies requires people to be aware of violations in the first place. With regard to the verifiability policy, verifiable information can be easily and objectively identified by checking the sources, which is supported by the finding that participants were able to detect and delete both the event-consistent and event-inconsistent unverifiable information in the article. What constitutes a balanced, neutral representation of an event, however, leaves much room for interpretation. For instance, recall that the source articles provided a balanced set of verifiable event-consistent and event-inconsistent information. Based on this source information, however, authors might still have given event-consistent information more weight, which is in line with the verifiability policy but violates the NPOV policy. As people are usually unaware that they are subject to hindsight bias and hold a biased event representation (Pohl & Hell, 1996), this violation of the NPOV policy might have gone unnoticed.

Consequently, findings suggest that article hindsight bias in the present study was not driven by the inclusion of unverifiable (event-consistent) information (i.e., violations of the verifiability policy) but rather by the preference for verifiable event-consistent content (i.e., violations of the NPOV policy). Interestingly, although the provision of the policies did not affect the magnitude of article hindsight bias in the present study, they could still prevent hindsight bias for certain events: If there is no verifiable information available for an event (e.g., the conclave to elect a new pope), authors might be tempted to engage in unverifiable speculations. In line with the finding of the present study that authors complied with the verifiability policy, article hindsight bias might be effectively prevented in such instances. As the source material in the present study contained substantial verifiable information, future research needs to address this question.

However, another explanation for the finding that the provision of the policies did not significantly reduce the magnitude of hindsight bias in the revised articles could be that our design did not offer sufficient potential to apply the policies. After all, for the reason of coding feasibility, the initial article draft only contained three unverifiable event-consistent and three unverifiable event-inconsistent pieces of information. In fact, descriptively, the revised articles of authors of both hindsight conditions (i.e., with and without policies provided) contained more unverifiable event-consistent than event-inconsistent pieces of information (see Figure 5), as one would expect based on causal model theory of hindsight bias (Blank & Nestler, 2007; Nestler et al., 2008). It is thus possible that this overweighing of event-consistent information would be even more pronounced if overall more unverifiable information was available in the initial article draft, offering more potential to apply the policies if they are provided. In addition, we designed the article draft to contain an equal number of unverifiable event-consistent and event-inconsistent pieces of information, but the application of the policies might be particularly relevant if unverifiable event-consistent information is already overrepresented. Importantly, in a real-world context, it is reasonable to assume that both information conditions are likely: The amount of unverifiable information on an event is theoretically unlimited, and authors might be tempted to include primarily event-consistent information in their article (even if it is unverifiable), considering that they are motivated to find an explanation for the event (Blank & Nestler, 2007; Nestler et al., 2008). Therefore, the finding that the provision of content policies did not affect the magnitude of hindsight bias in our experimental setting

might be the result of constraints in the information economy, and the policies might nevertheless be effective in reducing hindsight bias in a more realistic setting (as suggested, for instance, by findings that hindsight bias was absent in Wikipedia articles about various topics for which individuals succumb to hindsight bias; Oeberst et al., 2018; Oeberst, von der Beck, Cress, & Nestler, 2020).

General Discussion

We conducted four studies to examine what determines the magnitude of hindsight bias in written work such as Wikipedia articles. First of all, we obtained further strong support that written work can contain hindsight bias—both in the field when examining real Wikipedia articles (Study 1) and under high experimental control (Studies 2–4). In all four studies and consistently across all manipulations of the type of event, the available source information, and the writing context, hindsight articles suggested the given event to a greater extent and implied a greater inevitability and foreseeability of the event than foresight articles. Most importantly, the experimental studies provided support that the magnitude of article hindsight bias was affected by characteristics of the specific event in question: Article hindsight bias was greater when the event elicited high (vs. low) sense-making motivation (H_1 , Study 2) and when the available sources provided information with high (vs. low) causal relevance (H_2 , Study 3). With regard to the effect of the writing context, however, we obtained no support that providing authors with content policies effectively reduced article hindsight bias (H_3 , Study 4). In addition to the findings of the experimental studies, the preliminary correlation analysis of real Wikipedia articles in Study 1 yielded a significant association of the magnitude of article hindsight bias with the availability of causal information—assessed by the extent to which each article added explanatory content—(providing further support for H_2), but there was no significant correlation with the extent to which the event elicited sense-making motivation—operationalized as the increase in article views after the event had happened (no support for H_1). However, we consider the findings of the preregistered experimental studies with a controlled manipulation of the determinants (i.e., Studies 2 and 3) more reliable than the preliminary correlational exploration of the hypotheses (i.e., Study 1). After all, Study 1 was based on a very small sample ($N_{H1} = 31$; $N_{H2} = 40$), a rather indirect operationalization of the determinants, and a quasi-experimental design that cannot rule out potential confounds (see also the Discussion section of Study 1). Therefore, we base our interpretations regarding the effect of the three proposed determinants primarily on the findings of the three experimental studies.

The findings of the experimental studies demonstrate that causal model theory, which was postulated to explain hindsight bias in individuals (Blank & Nestler, 2007), also applies to hindsight bias in written work: As suggested by the findings of Study 2, the higher the authors' motivation to find an explanation for the event in question, the more they will engage in one-sided event-consistent evidence sampling and integration for their article (Pezzo, 2003; Roese & Olson, 1996; Slovic & Fischhoff, 1977), resulting in a biased, *written* representation of the event. The more (event-consistent) causal information is available, in turn, the more opportunities authors have to construct such a biased account (Nario & Branscombe, 1995; Yopchick & Kim, 2012), as supported by the findings of Study 3. Evidently, this reasoning also suggests that it needs both the

elicitation of sense-making motivation and the actual sense-making process (i.e., evidence sampling and integration) for article hindsight bias to occur: If authors are not motivated at all to explain the given event, their articles will not contain hindsight bias even if sufficient causal information is available (Schkade & Kilbourne, 1991); and, likewise, if sources do not provide any causal information, there will be no article hindsight bias even if authors are highly motivated to find an explanation (Yopchick & Kim, 2012). We note, however, that our study designs did not allow us to test this multiplicative model, as we only included a *low* sense-making motivation condition (Study 2) and a *low* causal relevance condition (Study 3), but no condition in which one factor was entirely absent.

Importantly, the findings of Studies 2 and 3 do not constitute a mere replication of the validity of causal model theory as the specific writing context precludes its straightforward generalization from the individual perception to written work (Nestler et al., 2017). Specifically, participants of both studies had to adhere to content policies that require authors to include only verifiable information and write from an NPOV. Thus, authors could not simply transcribe their (biased) individual views into their written work, and it is noteworthy that the authors' sense-making motivation and the availability of causal information nevertheless affected the magnitude of article hindsight bias.

The main goal of the present research was to examine the effect of the three proposed determinants on hindsight bias during writing in general, but we planned the studies to resemble one specific writing context: Wikipedia. Besides Study 1's obvious focus on real Wikipedia articles, the three experimental studies were designed to be applicable to writing and editing in Wikipedia. Specifically, even though Studies 2 and 3 examined *individual* writing (which is a departure from the *collaborative* nature of Wikipedia), participants were asked to write a "Wikipedia-like" article while adhering to content policies that apply in Wikipedia. Also, individual writing is an essential part of collaboration, particularly in an asynchronous collaborative writing context such as Wikipedia, where each author can edit articles without prior discussion with other authors. In Study 4, we mirrored the writing context in Wikipedia even more closely by letting participants edit an article collaboratively. Therefore, our findings might be particularly relevant for writing in Wikipedia.

Specifically, the findings of Study 1 demonstrate that hindsight bias in Wikipedia is even more common than suggested by previous work (Oeberst et al., 2018): Not only Wikipedia articles about disasters but also articles about other event categories such as elections, scientific discoveries, and personal decisions consistently contained hindsight bias. Evidently, Wikipedia's content policies could not prevent hindsight bias from entering the articles (see also the results of Study 4). This is problematic as Wikipedia is among the 15 most frequently retrieved web pages on the Internet (Similarweb, n.d.) and thus likely shapes the views of broad audiences. Also, since the content of Wikipedia articles is socially constructed and might be interpreted as collective memory (Kanhabua et al., 2014; Oeberst, von der Beck, Matschke, et al., 2020), hindsight bias in Wikipedia constitutes a *collective* bias. Importantly, even though the average suggestiveness, inevitability, and foreseeability ratings of the articles in most of our studies were consistent on the lower half of the scale (i.e., articles were *slightly* to *moderately* suggestive of the events), an unwarranted increase in these measures due to hindsight bias opposes the goal of many information outlets to present accurate event accounts (e.g., "Wikipedia: Policies and guidelines", 2021)

and is thus equally problematic, regardless of where it occurs on the scale.

Furthermore, the findings of Studies 2 and 3 suggest that article hindsight bias is particularly pronounced for events that elicit high sense-making motivation and for events for which substantial causal information is available. This is also in line with the finding that Wikipedia articles about disasters are particularly biased by hindsight (Oeberst et al., 2018; Study 1): Disasters are negative and unexpected events, which evoke particularly pronounced sense-making processes (Weiner, 1985), and, in the aftermath of disasters, much information is acquired due to detailed investigations and comprehensive media coverage (Simon, 1997). Interestingly, in a real-world context, a high sense-making motivation might not only exert a *direct* effect on the magnitude of article hindsight bias but also an *indirect* effect through an increase in the availability of causal information. Specifically, events that trigger a high motivation to find an explanation (e.g., disasters) usually are extensively covered by the media, which likely yields a rich pool of causal information that can then be used to write up an event-consistent account. Also, if media coverage is guided by the search for an explanation, it will likely be also biased by hindsight, and it thus will bring up predominantly event-consistent information, providing even more potential to construct a one-sided written event representation.

Besides the practical implications for the specific writing context of Wikipedia, the findings may also be applicable to written work in other domains. Most relatedly, a very similar writing context applies to writing and editing newspaper articles, where verifiable information and an NPOV are often desired (e.g., Reuters News Agency, n.d.). But even for more remote writing contexts, our findings might be relevant. For instance, researchers write scientific articles after data have been analyzed and may thus succumb to hindsight bias by selecting, interpreting, and integrating information in accordance with the results (Nosek et al., 2018). Also, written expert opinions (e.g., culpability reports), which are commissioned by court to guide legal decision-making, are carried out after the occurrence of an event and can thus be subject to hindsight bias (Giroux et al., 2016; Harley, 2007; Oeberst, 2019). Both types of writings—research articles and expert opinions—are written with a high motivation to answer a specific question, and their authors acquire as much information as feasible to draw valid conclusions. Yet, our findings suggest that it might be this high sense-making motivation that puts such types of writing at an increased risk of hindsight bias. This is particularly problematic considering the consequences that biased scientific reports (e.g., recommendation of inappropriate actions) or expert opinions (e.g., unjust verdicts) may entail. It would thus be an exciting next step to examine the determinants of hindsight bias in written work in such applied fields.

Another interesting endeavor for future research might be to test how the proposed determinants affect article hindsight bias at different stages of the writing process. Specifically, initial writing and the revision of an already existing work are two distinct cognitive processes (Hayes, 1989; Hayes & Berninger, 2014; Hayes et al., 1987). For instance, writing from scratch entails greater degrees of freedom in terms of content, syntactic, and lexical decisions. Consequently, there is more potential to include hindsight bias when initially writing an article than when revising an existing foresight article. Also, article hindsight bias might be less pronounced after revision because the foresight article might assist authors to (re)adopt a

foresight perspective, for instance, by making the preevent knowledge more salient or by reactivating the foresight uncertainty regarding the occurrence of the respective event (see Davies, 1987; Hoffrage et al., 2000; Nestler et al., 2012). Admittedly, in the present research, the stage of the writing process at which our manipulations applied differed across studies. While the field study in Wikipedia (Study 1) and the experimental study that addressed the effect of content policies (Study 4) concerned the revision of preexisting foresight articles, the experimental studies that examined the effects of the sense-making motivation (Study 2) and the availability of causal information (Study 3) asked participants to write an article from scratch. Therefore, it might well be that an increased sense-making motivation or availability of causal information does *not* affect article hindsight bias in the context of an article revision (which, as a matter of fact, could also explain why we did not find support for an effect of the sense-making motivation in Study 1). Conversely, the provision of content policies *might* exert an effect during initial writing. Furthermore, in the context of Wikipedia, articles that were only created *after* the respective event occurred (e.g., “Financial crisis of 2007–2008”, 2008), might be even more biased than articles that already existed prior to the event (which were used in Study 1 and the research by Oeberst et al., 2018, to allow for a normative measure of hindsight bias, see also Footnote 1). Importantly, both—writing from scratch (e.g., newspaper articles) and the revision of preexisting work (e.g., Wikipedia articles)—are relevant processes, and it is thus interesting to examine the magnitude and determinants of article hindsight bias for both stages of the writing process.

In a similar vein, our study designs differed in their level of collaboration during writing or editing. Hence, it is unclear whether collaboration might moderate the effect of the sense-making motivation (Study 2) or the availability of causal information (Study 3), which we obtained during individual writing, or whether the provision of content policies might affect article hindsight bias during a different level of collaboration (Study 4). However, hindsight bias is a robust and pervasive phenomenon (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004) and is thus likely shared among all authors of a written work. This is in line with findings that hindsight bias in individuals did not disappear during group processing (Bukzar & Conolly, 1988; Choi & Choi, 2010) and that the magnitude of hindsight bias in written work did not differ between individual and group writing (Oeberst et al., 2018). It seems that a biased event representation is a shared “blind spot” among all individuals, and correction processes during collaboration might thus be ineffective (Oeberst et al., 2016). Therefore, we deem it unlikely that variations in the level of collaboration would yield diverging results, but, of course, follow-up studies with an experimental manipulation of collaboration are needed to rule out collaboration effects.

Another limitation is that we used very similar text material about one specific event (i.e., disaster at a dam) in all three experimental studies to test the unique effect of each proposed determinant under high experimental control. It is thus unclear whether these findings can be generalized to other events and event categories. We note, however, that research on hindsight bias typically concerns one specific practical context (e.g., legal decision-making, elections), and it therefore is common practice to focus on only one particular event or one event category. Also, in the field study, we included events from various event categories,

allowing for more generalizable conclusions. In sum, then, we used a combination of highly ecologically valid data (Study 1) and highly controlled data (Studies 2–4) to examine the magnitude and determinants of hindsight bias in written work.

References

- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Arkes, H. R. (2013). The consequences of the hindsight bias in medical decision making. *Current Directions in Psychological Science*, *22*(5), 356–360. <https://doi.org/10.1177/0963721413489988>
- Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305–307. <https://doi.org/10.1037/0021-9010.73.2.305>
- Arslan, R. C., Walther, M. P., & Tata, C. S. (2018). *formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R* (Version v0.18.3) [Computer software]. <https://doi.org/10.31234/osf.io/pjasu>
- Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, *52*(1), 376–387. <https://doi.org/10.3758/s13428-019-01236-y>
- Ash, I. K. (2009). Surprise, memory, and retrospective judgment making: Testing cognitive reconstruction theories of the hindsight bias effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 916–933. <https://doi.org/10.1037/a0015504>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Biais, B., & Weber, M. (2009). Hindsight bias, risk perception, and investment performance. *Management Science*, *55*(6), 1018–1029. <https://doi.org/10.1287/mnsc.1090.1000>
- Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, *11*(4–5), 491–504. <https://doi.org/10.1080/09658210244000513>
- Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition*, *25*(1), 132–146. <https://doi.org/10.1521/soco.2007.25.1.132>
- Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? *Cognition*, *106*(3), 1408–1440. <https://doi.org/10.1016/j.cognition.2007.07.007>
- Bonds-Raacke, J. M., Fryer, L. S., Nicks, S. D., & Durr, R. T. (2001). Hindsight bias demonstrated in the prediction of a sporting event. *The Journal of Social Psychology*, *141*(3), 349–352. <https://doi.org/10.1080/00224540109600556>
- Bradfield, A., & Wells, G. L. (2005). Not the same old hindsight bias: Outcome information distorts a broad range of retrospective judgments. *Memory and Cognition*, *33*(1), 120–130. <https://doi.org/10.3758/BF03195302>
- Bukszar, E., & Conolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, *31*(3), 628–641. <https://doi.org/10.2307/256462>
- Carli, L. L. (1999). Cognitive reconstruction, hindsight, and reactions to victims and perpetrators. *Personality and Social Psychology Bulletin*, *25*(8), 966–979. <https://doi.org/10.1177/01461672992511005>
- Carli, L. L., & Leonard, J. B. (1989). The effect of hindsight on victim derogation. *Journal of Social and Clinical Psychology*, *8*(3), 331–343. <https://doi.org/10.1521/jscp.1989.8.3.331>
- Cassar, G., & Craig, J. (2009). An investigation of hindsight bias in nascent venture activity. *Journal of Business Venturing*, *24*(2), 149–164. <https://doi.org/10.1016/j.jbusvent.2008.02.003>
- Choi, D.-W., & Choi, I. (2010). A comparison of hindsight bias in groups and individuals: The moderating role of plausibility. *Journal of Applied Social Psychology*, *40*(2), 325–343. <https://doi.org/10.1111/j.1559-1816.2009.00576.x>
- Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*(1), 147–168. [https://doi.org/10.1016/0749-5978\(91\)90010-Q](https://doi.org/10.1016/0749-5978(91)90010-Q)
- Davies, M. F. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight-retrieval strategies. *Organizational Behavior and Human Decision Processes*, *40*(1), 50–68. [https://doi.org/10.1016/0749-5978\(87\)90005-7](https://doi.org/10.1016/0749-5978(87)90005-7)
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Financial crisis of 2007–2008. (2008, September 17). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Financial_crisis_of_2007%E2%80%932008&oldid=239084999
- Fischer-Prefler, D., Schwemmer, C., & Fischbach, K. (2019). Collective sense-making in times of crisis: Connecting terror management theory with Twitter user reactions to the Berlin terrorist attack. *Computers in Human Behavior*, *100*, 138–151. <https://doi.org/10.1016/j.chb.2019.05.012>
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299. <https://doi.org/10.1037/0096-1523.1.3.288>
- Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(2), 349–358. <https://doi.org/10.1037/0096-1523.3.2.349>
- Fischhoff, B., Gonzalez, R. M., Lerner, J. S., & Small, D. A. (2005). Evolving judgments of terror risks: Foresight, hindsight, and emotion. *Journal of Experimental Psychology: Applied*, *11*(2), 124–139. <https://doi.org/10.1037/1076-898X.11.2.124>
- Frieden, J. A. (2011, January 30). The financial crisis was foreseeable and preventable. *The New York Times*. <https://www.nytimes.com/roomfordebate/2011/01/30/was-the-financial-crisis-avoidable/the-financial-crisis-was-foreseeable-and-preventable>
- Gaddy, G. D., & Tanjong, E. (1986). Earthquake coverage by the western press. *Journal of Communication*, *36*(2), 105–112. <https://doi.org/10.1111/j.1460-2466.1986.tb01428.x>
- Giroux, M. E., Coburn, P. I., Harley, E. M., Connolly, D. A., & Bernstein, D. M. (2016). Hindsight bias and law. *Zeitschrift für Psychologie*, *224*(3), 190–203. <https://doi.org/10.1027/2151-2604/a000253>
- Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299–323). MIT Press.
- Granhag, P., Strömwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology*, *14*(5), 397–420. [https://doi.org/10.1002/1099-0720\(200009\)14:5<397::AID-ACP658>3.0.CO;2-M](https://doi.org/10.1002/1099-0720(200009)14:5<397::AID-ACP658>3.0.CO;2-M)
- Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. <https://doi.org/10.1007/s40881-015-0004-4>
- Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology*, *26*(2–3), 103–117. <https://doi.org/10.1080/01973533.2004.9646399>
- Hagmayer, Y., & Sloman, S. A. (2009). Decision makers conceive of their choices as interventions. *Journal of Experimental Psychology: General*, *138*(1), 22–38. <https://doi.org/10.1037/a0014585>

- Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, 25(1), 48–63. <https://doi.org/10.1521/soco.2007.25.1.48>
- Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, 46(1), 44–56. <https://doi.org/10.1037/0022-3514.46.1.44>
- Hastie, R., Schkade, D. A., & Payne, J. W. (1999). Juror judgments in civil cases: Hindsight effects on judgments of liability for punitive damages. *Law and Human Behavior*, 23(5), 597–614. <https://doi.org/10.1023/A:1022352330466>
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107(3), 311–327. <https://doi.org/10.1037/0033-2909.107.3.311>
- Hayes, J. R. (1989). Writing research: The analysis of a very complex task. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 209–234). Lawrence Erlbaum.
- Hayes, J. R., & Berninger, V. W. (2014). Cognitive processes in writing: A framework. In B. Arfe, J. Dockrell, & V. Berninger (Eds.), *Writing Development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction* (pp. 3–15). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199827282.003.0001>
- Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Reading, writing, and language processes* (pp. 176–240). Cambridge University Press.
- Heine, S. J., Proulx, T., & Vohs, K. D. (2006). The meaning maintenance model: On the coherence of social motivations. *Personality and Social Psychology Review*, 10(2), 88–110. https://doi.org/10.1207/s15327957pspr1002_1
- Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 566–581. <https://doi.org/10.1037/0278-7393.26.3.566>
- IBM Corp. (2015). *IBM SPSS statistics for windows* (Version 23.0) [Computer software].
- Joye, S. (2010). News discourses on distant suffering: A critical discourse analysis of the 2003 SARS outbreak. *Discourse and Society*, 21(5), 586–601. <https://doi.org/10.1177/0957926510373988>
- Kanhabua, N., Nguyen, T. N., & Niederée, C. (2014). What triggers human remembering of events? A large-scale analysis of catalysts for collective memory in Wikipedia. In G. Buchanan, M. Klein, A. Rauber, & S. J. Cunningham (Eds.), *Proceedings of the 14th ACM/IEEE—CS Joint Conference on Digital Libraries* (pp. 341–350). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/JCDL.2014.6970189>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, 10(10), 464–470. <https://doi.org/10.1016/j.tics.2006.08.004>
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231–1243. <https://doi.org/10.1037/0022-3514.47.6.1231>
- Louie, T. A., Rajan, M. N., & Sibley, R. E. (2007). Tackling the Monday-morning quarterback: Applications of hindsight bias in decision-making settings. *Social Cognition*, 25(1), 32–47. <https://doi.org/10.1521/soco.2007.25.1.32>
- Marchetti, R. (2017, January 21). Why Trump's election was inevitable. *HuffPost*. https://www.huffpost.com/entry/why-trumps-election-was-inevitable_b_5882f582e4b0111ea60b9656
- Meuer, M., Nestler, S., & Oeberst, A. (2021). Debiasing media articles—Reducing hindsight bias in the production of written work. *Journal of Applied Research in Memory and Cognition*, 10(3), 435–443. <https://doi.org/10.1016/j.jarmac.2020.12.006>
- Meuer, M., von der Beck, I., Nestler, S., & Oeberst, A. (2021). What drives increases in hindsight impressions after the reception of biased media content? *Journal of Experimental Psychology: Applied*, 27(3), 461–472. <https://doi.org/10.1037/xap0000353>
- Meyer, W.-U., Reisenzein, R., & Schützwohl, A. (1997). Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21(3), 251–274. <https://doi.org/10.1023/A:1024422330338>
- Nario, M. R., & Branscombe, N. R. (1995). Comparison processes in hindsight and causal attribution. *Personality and Social Psychology Bulletin*, 21(12), 1244–1255. <https://doi.org/10.1177/01461672952112001>
- Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1399–1413. <https://doi.org/10.1037/a0020449>
- Nestler, S., Blank, H., & von Collani, G. (2008). Hindsight bias doesn't always come easy: Causal models, cognitive effort, and creeping determinism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1043–1054. <https://doi.org/10.1037/0278-7393.34.5.1043>
- Nestler, S., & Egloff, B. (2009). Increased or reversed? The effect of surprise on hindsight bias depends on the hindsight component. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1539–1544. <https://doi.org/10.1037/a0017006>
- Nestler, S., Egloff, B., Kűfner, A. C. P., & Back, M. D. (2012). An integrative lens model approach to bias and accuracy in human inferences: Hindsight effects and knowledge updating in personality judgments. *Journal of Personality and Social Psychology*, 103(4), 689–717. <https://doi.org/10.1037/a0029461>
- Nestler, S., Leckelt, M., Back, M. D., von der Beck, I., Cress, U., & Oeberst, A. (2017). Produktion von naturwissenschaftlichen Informationen im Internet am Beispiel von Wikipedia [Production of scientific information on the Internet using the example of Wikipedia]. *Psychologische Rundschau*, 68(3), 172–176. <https://doi.org/10.1026/0033-3042/a000360>
- Nestler, S., & von Collani, G. (2008). Hindsight bias and the activation of counterfactual mind-sets. *Experimental Psychology*, 55(5), 342–349. <https://doi.org/10.1027/1618-3169.55.5.342>
- Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience and Biobehavioral Reviews*, 35(4), 1042–1051. <https://doi.org/10.1016/j.neubiorev.2010.08.011>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Oeberst, A. (2019). Der Rückschaufehler im juristischen Kontext: Relevante psychologische Forschung, begründete Spekulationen und Schlussfolgerungen für die Praxis [Hindsight bias in a legal context: Relevant psychological research, reasoned speculations, and practical implications]. *Rechtswissenschaft*, 10(2), 180–203.
- Oeberst, A., Cress, U., Back, M., & Nestler, S. (2016). Individual versus collaborative information processing: The case of biases in Wikipedia. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education* (pp. 165–185). Springer International Publishing. https://doi.org/10.1007/978-3-319-13536-6_9
- Oeberst, A., von der Beck, I., Back, M. D., Cress, U., & Nestler, S. (2018). Biases in the production and reception of collective knowledge: The case of hindsight bias in Wikipedia. *Psychological Research*, 82(5), 1010–1026. <https://doi.org/10.1007/s00426-017-0865-7>
- Oeberst, A., von der Beck, I., Cress, U., & Nestler, S. (2020). Wikipedia outperforms individuals when it comes to hindsight bias. *Psychological Research*, 84(6), 1517–1527. <https://doi.org/10.1007/s00426-019-01165-7>
- Oeberst, A., von der Beck, I., Matschke, C., Ihme, T. A., & Cress, U. (2020). Collectively biased representations of the past: Ingroup Bias in Wikipedia articles about intergroup conflicts. *British Journal of Social Psychology*, 59(4), 791–818. <https://doi.org/10.1111/bjso.12356>

- Oeberst, A., von der Beck, I., & Nestler, S. (2014). Reading about explanations enhances perceptions of inevitability and foreseeability: A cross-cultural study with Wikipedia articles. *Cognitive Processing, 15*(3), 343–349. <https://doi.org/10.1007/s10339-014-0603-7>
- Park, C. L. (2010). Making sense of the meaning literature: An integrative review of meaning making and its effects on adjustment to stressful life events. *Psychological Bulletin, 136*(2), 257–301. <https://doi.org/10.1037/a0018301>
- Pezzo, M. V. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory, 11*(4–5), 421–441. <https://doi.org/10.1080/09658210244000603>
- Pohl, R. F., & Erdfelder, E. (2017). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (2nd ed., pp. 424–445). Routledge.
- Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes, 67*(1), 49–58. <https://doi.org/10.1006/obhd.1996.0064>
- Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology, 93*(5), 959–981. <https://doi.org/10.1037/0021-9010.93.5.959>
- R Core Team. (2020). *R: A language and environment for statistical computing*. (Version 4.0.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science, 11*(1), 50–74. <https://doi.org/10.1111/tops.12292>
- Reuters News Agency. (n.d.). *Standards and values*. Thomson Reuters. <https://www.reutersagency.com/en/about/standards-values/>
- Rocca, F. X. (2013, February 11). *Analysis: Pope's resignation almost predictable*. Arkansas Catholic. <https://www.arkansas-catholic.org/news/article/3335/Analysis-Popes-resignation-almost-predictable>
- Roese, N. J., & Olson, J. M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology, 32*(3), 197–227. <https://doi.org/10.1006/jesp.1996.0010>
- Roese, N. J., & Vohs, K. D. (2012). Hindsight Bias. *Perspectives on Psychological Science, 7*(5), 411–426. <https://doi.org/10.1177/1745691612454303>
- Russel, V. L., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *Estimated marginal means, aka least-squares means* (Version 1.6.1) [R package]. <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>
- Schkade, D. A., & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes, 49*(1), 105–123. [https://doi.org/10.1016/0749-5978\(91\)90044-T](https://doi.org/10.1016/0749-5978(91)90044-T)
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Schwitzgebel, E. (1999). Children's theories and the drive to explain. *Science and Education, 8*(5), 457–488. <https://doi.org/10.1023/A:1008633709447>
- Similarweb. (n.d.). *Top website-ranking*. Retrieved July 20, 2022, from <https://www.similarweb.com/de/top-websites/>
- Simon, A. F. (1997). Television news and international earthquake relief. *Journal of Communication, 47*(3), 82–93. <https://doi.org/10.1111/j.1460-2466.1997.tb02718.x>
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance, 3*(4), 544–551. <https://doi.org/10.1037/0096-1523.3.4.544>
- van Dijk, Z. (2021). *Wikis und die Wikipedia verstehen* [Understanding wikis and Wikipedia]. Transcript Verlag. <https://doi.org/10.14361/9783839456453>
- Vardi, E., Muchnik, L., Conway, A., & Breakstone, M. (2021). WikiShark: An online tool for analyzing Wikipedia traffic and trends. In J. Leskovec, M. Grobelnik, M. Najork, J. Tang, & L. Zia (Eds.), *WWW'21: Companion proceedings of the web conference 2021* (pp. 558–571). Association for Computer Machinery. <https://doi.org/10.1145/3442442.3452341>
- Verplanken, B., & Pieters, R. G. (1988). Individual differences in reverse hindsight bias: I never thought something like Chernobyl would happen. Did I? *Journal of Behavioral Decision Making, 1*(3), 131–147. <https://doi.org/10.1002/bdm.3960010302>
- von der Beck, I., Oeberst, A., Cress, U., & Nestler, S. (2017). Cultural interpretations of global information? Hindsight bias after reading Wikipedia articles across cultures. *Applied Cognitive Psychology, 31*(3), 315–325. <https://doi.org/10.1002/acp.3329>
- Wasserman, D., Lempert, R. O., & Hastie, R. (1991). Hindsight and causality. *Personality and Social Psychology Bulletin, 17*(1), 30–35. <https://doi.org/10.1177/0146167291171005>
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573. <https://doi.org/10.1037/0033-295X.92.4.548>
- Wikimedia Statistics. (2021, July 1). *Unique devices*. <https://stats.wikimedia.org/>
- Wikipedia: Core content policies. (2021, November 3). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Core_content_policies&oldid=1053277364
- Wikipedia: Neutral point of view. (2021, June 25). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Neutral_point_of_view&oldid=1030312226
- Wikipedia: No original research. (2021, June 24). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:No_original_research&oldid=1030211039
- Wikipedia: Policies and guidelines. (2021, June 17). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Wikipedia:Policies_and_guidelines&oldid=1028977013
- Wikipedia: Verifiability. (2021, June 28). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Wikipedia:Verifiability&oldid=1030951494>
- Wikipedia. (2021, June 28). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Wikipedia&oldid=1030922202>
- Yopchick, J. E., & Kim, N. S. (2012). Hindsight bias and causal reasoning: A minimalist approach. *Cognitive Processing, 13*(1), 63–72. <https://doi.org/10.1007/s10339-011-0414-z>

Received September 21, 2021

Revision received June 24, 2022

Accepted June 28, 2022 ■