

# Biases in the production and reception of collective knowledge: the case of hindsight bias in Wikipedia

Aileen Oeberst<sup>1,2</sup> · Ina von der Beck<sup>2</sup> · Mitja D. Back<sup>3</sup> · Ulrike Cress<sup>2</sup> · Steffen Nestler<sup>4</sup>

Received: 18 November 2016 / Accepted: 5 April 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** The Web 2.0 enabled collaboration at an unprecedented level. In one of the flagships of mass collaboration—Wikipedia—a large number of authors socially negotiate the world’s largest compendium of knowledge. Several guidelines in Wikipedia restrict contributions to verifiable information from reliable sources to ensure recognized knowledge. Much psychological research demonstrates, however, that individual information processing is biased. This poses the question whether individual biases translate to Wikipedia articles or whether they are prevented by its guidelines. The present research makes use of hindsight bias to examine this question. To this end, we analyzed foresight and hindsight versions of Wikipedia articles regarding a broad variety of events (Study 1). We found the majority of articles *not* to contain traces of hindsight bias—contrary to prior individual research. However, for a particular category of events—disasters—we found robust evidence for hindsight bias. In a lab experiment (Study 2), we then examined whether individuals’ hindsight bias is translated into articles under controlled conditions and tested whether collaborative writing—as present in Wikipedia—affects the resultant bias (vs. individual writing). Finally, we investigated the impact of biased Wikipedia articles on readers (Study 3). As predicted, biased articles elicited a hindsight bias in

readers, who had not known of the event previously. Moreover, biased articles also affected individuals who knew about the event already, and who had already developed a hindsight bias: biased articles further increased their hindsight.

## Introduction

The World Wide Web has revolutionized our access to information. A myriad of even remote sources is immediately available at our fingertips. However, with the development of Web 2.0 technologies, the production of informational contents is no longer limited to professionals. Rather, any person with internet access can contribute to the informational web content. Laypersons are thus not only receivers of information but also its producers. Interestingly, a large number of these products are the result of collective actions as the Web 2.0 enabled people to collaborate at an unprecedented level.

By now, much research points to the benefits of mass collaboration. For instance, the collaboration among millions has led to the creation of the largest compendium of world knowledge: Wikipedia. Psychological research with individuals indicates, however, that human information processing is often biased (e.g., Pohl, 2017). For example, we falsely believe in hindsight that we had known all along in foresight what would happen (hindsight bias; Fischhoff, 1975; Roese & Vohs, 2012; Pohl & Erdfelder, 2017). Considering that many biases are robust and widespread, the question arises as to whether they extend to a collective level, that is, whether they are mirrored in products of “collective effort” such as mass collaboration. The aim of the present studies is thus to examine whether these collective representations contain traces of individual biases.

✉ Aileen Oeberst  
aoeberst@uni-mainz.de

<sup>1</sup> University of Mainz, Binger Str. 14-16, 55122 Mainz, Germany

<sup>2</sup> Leibniz-Institut für Wissensmedien, Tübingen, Germany

<sup>3</sup> University of Münster, Münster, Germany

<sup>4</sup> University of Leipzig, Leipzig, Germany

To investigate this question, we analyzed how representations of events in Wikipedia articles change over time and whether Wikipedia articles show evidence of hindsight bias. For instance, does the article about the nuclear power plant of Fukushima suggest—in hindsight, but not in foresight—that the nuclear disaster was likely? Moreover, we examined whether hindsight bias in Wikipedia articles may, in turn, affect readers' perceptions of events (i.e., their hindsight bias). That is, we tested whether hindsight bias is (1) transferred within the course of producing Wikipedia articles, as well as (2) when perceiving Wikipedia articles. We chose Wikipedia as it is one of the flagships of mass collaboration (it is among the ten most frequently retrieved pages of the Internet, <http://www.alexacom>, and is also increasingly discovered in and for academic circles, e.g., see <http://www.psychologicalscience.org/members/aps-wikipedia-initiative>) and thus likely shapes the representations of a broad audience. Furthermore, Wikipedia comes along with a number of guidelines that aim at preventing bias. Hindsight bias was chosen, because it is one of the most robust cognitive biases (see meta-analyses of Christensen-Szalanski & Wilham, 1991; Guilbault, Bryant, Brockway, & Posavac, 2004) and has a number of far reaching consequences such as effects on the attribution of responsibility and guilt (e.g., Carli, 1999; Hastie, Schkade, & Payne, 1999; LaBine & LaBine, 1996; Rachlinski, 1998).

The article is organized as follows. We first turn to Wikipedia and briefly outline its principles and its status as a repository of collective representations. Next, we define hindsight bias and elaborate on relevant research. Subsequently, we turn to the perception of biased Wikipedia articles. We then report two studies that examined whether Wikipedia articles show traces of hindsight bias (production; Studies 1 and 2) and one study that investigated how Wikipedia articles affect readers' biases in the representations of events (perception; Study 3). Finally, we summarize our results and discuss their implications.

## Mass collaboration and knowledge production in Wikipedia

Mass collaboration involves the activities of a large number of people. It is usually mediated by digital tools (e.g., Web 2.0) and results in (digital) products (Cress, Jeong, & Moskaliuk, 2016; e.g., mathematical solutions, Gowers & Nielsen, 2009; data for scientific research, Barron, Martin, Mertl, & Yassine, 2016; computer games, Fields, Kafai & Giang, 2016). One of the most prominent results of mass collaboration is the online encyclopedia Wikipedia. By now, it exists in more than 280 different languages. Its

largest version alone—the English language version—was authored by more than 26 million users and contains more than five million articles (<https://en.wikipedia.org/wiki/Wikipedia:Statistics>).

In contrast to traditional encyclopedias, Wikipedia is written exclusively by volunteers and is open to anyone. Therefore, the educational background of the authors is very diverse (e.g., Merz & Döring, 2010) implying that many authors do not have a professional education in the topics they write about (e.g., Oeberst, Halatchliyski, Kimmerle, & Cress, 2014a). Another difference to traditional encyclopedias is the number of authors: on average, articles in the English Wikipedia are written by 50 different authors (Kittur & Kraut, 2008). This number is easily multiplied when it comes to articles of broad importance and high topicality (e.g., the nuclear disaster in Fukushima, Keegan, Gergle, & Contractor, 2011; Oeberst et al., 2014a, b; the Arab spring, Ferron & Massa, 2011; Massa & Scrinzi, 2012), or highly controversial topics (e.g., Wilson & Likens, 2015).

Wikipedia's content has thus been socially negotiated (by collaborative authoring) and is publicly available. It may, therefore, be conceived of as a repository for collective representations (Pentzold, 2009; Olick, 1999). Moreover, it is, indeed, frequently retrieved (<http://www.alexacom>). If Wikipedia articles were biased, they could thus likely shape the views of millions.

Unknown to many users, however, Wikipedia operates on a number of basic rules that aim at preventing bias. Most important for the present purpose is the following three rules: (1) verifiability (<http://en.wikipedia.org/wiki/Wikipedia:Verifiability>; [http://en.wikipedia.org/wiki/Wikipedia:Verifiability#What\\_counts\\_as\\_a\\_reliable\\_source](http://en.wikipedia.org/wiki/Wikipedia:Verifiability#What_counts_as_a_reliable_source)), (2) no original research ([http://en.wikipedia.org/wiki/Wikipedia:No\\_original\\_research](http://en.wikipedia.org/wiki/Wikipedia:No_original_research)), and (3) neutral point of view ([http://en.wikipedia.org/wiki/Wikipedia:Neutral\\_point\\_of\\_view](http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view)). These rules urge authors (1) to contribute only information that is verifiable and from reliable sources, (2) to contribute recognized knowledge (i.e., precluding novel thoughts and theories to be presented), and (3) to use an unbiased language and to include “all significant viewpoints that have been published by reliable sources, in proportion to the prominence of each viewpoint”.

Prior research has shown that norms may, indeed, decrease bias (Postmes, Spears, & Cihangir, 2001). Moreover, Wikipedia's guidelines, its overall goal to provide access to world knowledge, and the fact that the information is publicly available may effectively foster accuracy motivation (Chen, Shechter, & Chaiken, 1996) rather than motivated reasoning (Kunda, 1990). Could this effectively prevent biases to enter Wikipedia articles? Several studies document that Wikipedia's rules guide individual contributions effectively (Forte & Bruckman,

2008; Oeberst et al., 2014a, b; Viégas, Wattenberg, & Dave 2004; Viégas, Wattenberg, Kriss, & van Ham 2007). And even if Wikipedia may not prevent vandalism (e.g., inserting knowingly false information; e.g., Potthast, Stein, & Gerling, 2008), many authors and computer algorithms continuously check and implement Wikipedia's guidelines, which mostly leads to the fast correction of destructive edits (Adler, Alfaro, de Mola-Velasco, Rosso, & West, 2011; Potthast et al., 2008; Viégas et al., 2004, 2007). Similarly, errors often get corrected soon after they get published (e.g., Fallis, 2009). Even more importantly, errors do not necessarily occur more frequently than in traditional encyclopedias (e.g., Britannica, Giles, 2005; see also Fallis, 2008; Magnus, 2009) and a balanced presentation of highly political and emotionally laden events has been obtained as well (Oeberst et al., 2014a, b).

Despite these positive demonstrations of Wikipedia's quality, the encyclopedia is not free from errors. For instance, Wikipedia authors often share certain characteristics (e.g., interest in social media), which may result in an "imbalanced coverage of subjects on Wikipedia" ([http://en.wikipedia.org/wiki/Wikipedia:Systemic\\_bias](http://en.wikipedia.org/wiki/Wikipedia:Systemic_bias), e.g., Bellomi & Bonato, 2005; Callahan & Herring, 2011; Hecht & Gergle, 2009, 2010; Royal & Kapila, 2009). While Wikipedia's guidelines do not tackle topic coverage, this example perfectly demonstrates that there are biases that are not covered by Wikipedia's guidelines: an article may contain verifiable information from reliable sources and be presented neutrally, but may, nevertheless, contain bias—merely due to the selective presentation of information that results from the authors' perspective. Another bias that might not be prevented by Wikipedia's guidelines is hindsight bias.

### Hindsight bias as an individual bias

Hindsight bias is the tendency to overestimate in hindsight what one has known in foresight. Once an event occurred, people tend to perceive it as more likely, more inevitable, or more foreseeable than they had before its occurrence (see Roese & Vohs, 2012; Pohl & Erdfelder, 2017, for overviews). In his seminal study, Fischhoff (1975) presented participants with a historical event (e.g., the British–Gurkha War) and asked them to estimate the likelihood of possible outcomes (e.g., British victory, Gurkha victory). Crucially, some of his participants were informed about the alleged outcome of the war prior to making their likelihood judgments, while participants in a control group did not receive any information about the outcome of the war. Compared to this control group, participants with outcome knowledge systematically overestimated the likelihood of the "actual" event. This biased retrospective evaluation of

events even held when participants were urged to ignore outcome knowledge (Fischhoff, 1975) or when they were informed and warned about hindsight bias prior to the experiment (Fischhoff, 1977). Participants were thus unable to ignore outcome knowledge and to put themselves into the foresight perspective.

Since Fischhoff's experiments, a vast number of studies have investigated hindsight bias and demonstrated its robustness (see Christensen-Szalanski & Willham, 1991, and Guilbault et al., 2004, for meta-analyses) and pervasiveness (e.g., Pohl, Bender, & Lachmann, 2002). In addition, a number of explanations have been put forward (see Roese & Vohs, 2012 for an overview). Concerning events, several researchers identified causal reasoning as a crucial underlying process (e.g., Blank & Nestler, 2007; Louie, 2005; Jennings, Lowe, & Reckers, 1998; Nestler, Blank, & von Collani 2008; Nestler & Egloff, 2009; Pezzo, 2003; Roese & Olson, 1996; Yopchick & Kim, 2012). Based on the assumption that individuals are generally motivated to understand the world, it is proposed that they search for antecedents that are causally linked to the outcome and evaluate these antecedents regarding their suitability to explain the outcome's occurrence. Importantly, as the search process is biased towards seeking antecedents that may explain the occurred outcome, individuals place more weight on event-consistent antecedents than inconsistent ones (which would have spoken for a different outcome; Nestler et al., 2008; see also Carli, 1999; Fischhoff, 1975), suggesting that the event was more likely to happen.

### Hindsight bias in the production of Wikipedia articles

To date, the vast majority of studies on hindsight bias examined individuals' personal perceptions: participants received background information as well as outcome information and were asked for their personal perceptions regarding the likelihood, inevitability, or foreseeability of the outcome. In other words, hindsight research focused on the reception of information and how this information is evaluated. The question of whether Wikipedia articles contain hindsight bias thus differs substantially from the previous studies as it involves the production of information. Information production, however, comes along with a number of processes that go beyond the reception of information and which are affected by additional factors (Hayes, 1989; see also Nestler et al., 2017). One of these factors is the context in which information is produced. As we have outlined above, Wikipedia provides a unique context: first, Wikipedia employs several guidelines that explicitly aim at preventing personal opinions and

subjective evaluations and demand verifiable contents from reliable sources instead. Second, Wikipedia articles are socially negotiated by many authors—on the basis of Wikipedia’s guidelines. Therefore, analyzing Wikipedia articles also differs substantially from the previous group studies on hindsight bias. There, individuals or small groups received information (e.g., statements describing psychological research, Stahlberg et al., 1995, Study 1), half of them also learned about the actual outcome (e.g., that this finding was actually confirmed/falsified by research), and all participants were asked to estimate how likely they would have thought this statement to be true (disregarding outcome knowledge when provided). Participants in the group conditions typically have to discuss the matter for a limited amount of time (e.g., 30–45 min) and have to come up with an unanimous judgment (see also Bukszar & Conolly, 1988; Choi et al., 2007; Yama et al., 2010). The differences to the context of Wikipedia are evident: first, Wikipedia articles involve many more people than lab groups (typically 2–4 people). Second, Wikipedia authors are much more heterogeneous. Third, social negotiation in Wikipedia takes place via the production of text rather than scales. Fourth, social negotiation in Wikipedia is guided by Wikipedia’s guidelines of verifiability, neutrality, and recognized knowledge. Fifth, Wikipedia authors are not limited in their amount of time for their social negotiation. Importantly, this may also mean that they do not achieve a consensus at a certain point in time (i.e., in an article version that we analyzed in Study 1).

Taken together, Wikipedia is a prominent example of knowledge production in the World Wide Web, which results from collaboration at an unprecedented level. Several studies have documented the positive effects of mass collaboration, and Wikipedia, in particular, has implemented several guidelines to foster the quality of the articles produced. In consideration of psychological research about biases in human information processing, however, the question arises whether the collective representations in Wikipedia, nevertheless, contain traces of individual biases—such as hindsight bias. After all, hindsight bias is likely shared among authors: much research has shown how difficult it is to overcome hindsight bias and that individuals do not spontaneously engage in strategies to reduce the bias (e.g., considering how the same circumstances could have led to a different outcome; see Roese & Vohs, 2012). Hence, it can be presumed that the same cognitive processes that underlie hindsight bias occur in all individuals who collaborate. Moreover, studies indicate that hindsight bias does not vanish in groups. It was rather obtained to the same (Bukszar & Conolly, 1988; Stahlberg, Eller, Maass, & Frey 1995, Exp. 1) or even to a greater extent than in individuals (Choi et al., 2007). This implies that any correction processes that one may assume

occurring at the collective or group level are, in fact, not taking place (or not so strong to significantly reduce the bias). On the contrary, people become even more extreme in their view when exchanging with others due to the exposure to novel arguments that are consistent with one’s own evaluation (i.e., event-consistent information, Isenberg, 1986) and due to the motivation to present oneself in a socially desirable way (e.g., as highly knowledgeable person who is able to foresee developments (Mark & Mellor, 1991)). Hence, hindsight bias is likely shared among all individual authors but unlikely detected and reduced by their collaboration.

A second major argument for why hindsight bias might enter Wikipedia articles is that hindsight bias might circumvent Wikipedia’s guidelines. Specifically, if an article contains event-consistent antecedents but not event-inconsistent ones (see above), this biased selection of information may entirely go unnoticed as long as the (outcome-consistent) information included is verifiable, from reliable sources and presented neutrally. Moreover, given the pervasiveness of hindsight bias (Guilbault et al., 2004; Pohl et al., 2002), and the fact that people are mostly not aware of it (Pohl & Hell, 1996) or unable to avoid it (Fischhoff, 1975, 1977), it is unlikely that a non-biased representation of the event is regarded as a “significant” viewpoint that should be included into the article. In sum then, when biased individuals collaboratively construct a representation of an event, this representation is likely biased as well (e.g., Cress & Kimmerle, 2008; Schulz-Hardt, Jochims, & Frey 2002).

### Hindsight bias in the perception of biased Wikipedia articles

If Wikipedia articles contain a hindsight bias, they would be highly suggestive of the occurrence of an event—after the fact. Interestingly, reading such highly suggestive articles may have the consequence that a participant’s perception of the likelihood of an event is even more increased. Note that this question extends prior research on hindsight bias, which was concerned with the elicitation of the bias. That is, all participants were presented with the same information and whether they receive outcome knowledge was varied. This proceeding allowed conclusions about how the same antecedents (i.e., the situation at foresight) are evaluated differently once the outcome is known (e.g., Carli, 1999; Fischhoff, 1975). What happens, however, when the information read already contains a hindsight bias? In this case, one would expect biased articles to (1) elicit a hindsight bias in individuals who have not heard of the event beforehand (i.e., as in the standard hindsight paradigm) and to (2) increase hindsight bias in



individuals who already knew the event outcome. We have obtained some preliminary evidence for the notion that reading biased articles increases hindsight distortions in comparison to when unbiased articles are read (Oeberst, von der Beck, & Nestler, 2014b). In this study, participants read either a foresight version of the article about the nuclear power plant in Fukushima (the last one that existed prior to the nuclear disaster) or a hindsight version of the article that existed 8 weeks after the catastrophe began. The hindsight article version had been rated as being highly suggestive of the disaster (i.e., to contain hindsight bias). Reading the hindsight article version increased participants' perceptions of the likelihood, inevitability, and foreseeability of the disaster. Since we had not obtained genuine foresight estimates for the nuclear disaster, however, it remains unclear, whether reading biased articles increases individuals' hindsight bias above and beyond a previously developed "classic" hindsight bias. There are reasons for such an additional effect: reading an article that is biased by hindsight, may, for instance, provide novel outcome-consistent arguments for the event's occurrence (see Isenberg, 1986, for a related effect). However, even if the information contained in the article was identical to participants' own information, reading the article may still increase their certainty concerning their perception and evaluation (e.g., Tesser, 1978).

## The present research

Taken together, the present paper investigates hindsight bias in the production and reception of Wikipedia articles. With regard to production, we examine, whether we find traces of hindsight bias in Wikipedia articles (Hypothesis 1), or whether Wikipedia's guidelines prevent hindsight bias to enter the articles (Alternative Hypothesis 1). A unique feature of Wikipedia allows us to investigate these hypotheses empirically: as Wikipedia is based on wiki technology, every article version (i.e., every edit) is saved separately, which enables comparisons of event representations over time (e.g., foresight vs. hindsight article versions). Studies 1 and 2 examine this question in the field and under controlled conditions, respectively.

Concerning reception, we expect biased articles to elicit a hindsight bias in readers who were unfamiliar with the event beforehand—which is consistent with prior research on hindsight bias (Hypothesis 2). Above and beyond, we propose that reading has an additional effect on readers' hindsight bias beyond their classic individual hindsight biases (Hypothesis 3). Study 3 tests both, Hypotheses 2 and 3.

## Study 1

This study investigated whether events in Wikipedia articles are represented as more likely in retrospect. For a total of 33 events, we retrieved article versions from the German Wikipedia that existed prior to the event (foresight) or after the event had happened (hindsight) and assessed indicators of hindsight bias in those articles. By comparing foresight and hindsight versions of articles, we were able to examine whether there is evidence for hindsight distortions in Wikipedia articles.

## Method

### *Selection of events and article versions*

We made use of 33 events from six different event categories: (1) elections (e.g., President election in Russia, 2008), (2) public/official decisions (e.g., the declaration of independence of Kosovo), (3) personal decisions (e.g., the suicide of Robert Enke), (4) disasters and calamities (e.g., the nuclear disaster in Fukushima), (5) sports events (e.g., winner of the European soccer championship in 2012), and (6) scientific discoveries (e.g., evidence for the Higgs Boson; see <https://osf.io/vsryp/> for the full list of events as well as the data for all studies). Every event category contained five to six events, whereby half of the events in each category were known by the raters and the other half was not. We selected popular and unpopular events to exclude the possibility that the assessments of hindsight indicators are influenced by coders knowing the event. As expected, knowledge of the event outcome did not influence any of the dependent variables. Hence, we omitted this variable in all further analyses.

For each event, we retrieved three article versions from the revision history: (1) the last article version that existed prior to the event ( $t_1$  version), (2) the first article version immediately after the event happened, which already contained outcome information ( $t_2$  version) as well as (3) the article version that existed eight weeks after the event had happened ( $t_3$  version). Overall, we retrieved  $3 \times 33 = 99$  article versions. The study thus comprised a  $6$  (event category)  $\times$   $3$  (article version) mixed design with article version varying within events and event category varying between events.

### *Linguistic hindsight indicators*

To analyze hindsight bias with a quantitative measure, we assessed linguistic markers that reflect indicators of hindsight bias. To this end, we conducted automatic text analyses with the Linguistic Inquiry and Word Count (LIWC;

Tausczik & Pennebaker, 2010). LIWC counts words that belong to different categories (e.g., positive emotions, cognitive mechanisms) and provides the percentage of words in the whole text that fall into this category. It has extensively been validated (see Tausczik & Pennebaker, 2010) and successfully used in various research contexts (e.g., Kufner, Nestler, Back, & Egloff, 2010; Robinson, Cassidy, Boyd, & Fetterman, 2015; Rodriguez, Holleran, & Mehl, 2010; Schultheiss, 2013). Here, we determined the number of words of the categories “cause” (containing words such as “hence”), “certainty” (e.g., “always”), tentativeness (e.g., “maybe”), “insight” (e.g., “consider”), and “discrepancy” (e.g., “should”), because the hindsight perspective is assumed to be the result of successful causal modeling (cf., Nestler et al., 2008) and thus is characterized by more certainty and insight and perceptions that those, who are responsible should have foreseen the event (Pezzo, 2003). For the analysis used the sum of all categories (Cronbach’s  $\alpha$  were 0.664, 0.725, and 0.714, for the  $t_1$ ,  $t_2$ , and  $t_3$  article versions, respectively).

#### *Coded hindsight indicators*

Furthermore, we had all Wikipedia article versions rated by ten coders each, who were blind to the specific research questions. The coding scheme developed for this study contained the following main variables: First, raters’ evaluation of whether the article version suggested that a particular event was likely to happen was assessed on a 5-point scale (1 = no particular event is suggested, 5 = a particular event is highly suggested). Second, the number of explicit phrases that are typical for a hindsight bias (e.g., “It was not surprising that [the event] took place.”) was scored for each article version. Finally, a number of other ratings were also assessed, including, for example, the number of explanations contained in the article or the valence of the outcome. However, these ratings are not analyzed here (see <https://osf.io/vsryp/> for the entire coding scheme).

All raters were trained with three extra events for which they coded each of the three article versions ( $t_1$ ,  $t_2$ , and  $t_3$ ). For each event, raters first coded the  $t_1$  version before receiving article versions  $t_2$  and  $t_3$ . In addition, they were urged not to search for further information before having coded  $t_1$ . We calculated intra-class correlation coefficients to determine consistency among coders regarding all metric ratings. Inter-rater agreement was  $ICC_{t_1} = 0.71$  for the hindsight rating of the first article version,  $ICC_{t_2} = 0.73$  for the hindsight of the second article version, and for the last version, it was  $ICC_{t_3} = 0.71$ . For the number of explicit phrases referring to hindsight bias, we found agreement values of  $ICC_{t_1} = 0.71$  for the first version,  $ICC_{t_2} = 0.54$  for the second, and  $ICC_{t_3} = 0.79$  for the final version.

All coded hindsight indicators concern the event that actually occurred. In case of the unknown events, we, therefore, recoded raters’ evaluation when they had evaluated the  $t_1$  article version to be highly suggestive of another event (which did not occur). That is, if a rater found an article about an election to be suggestive of a victory of party X and choosing a rating of “5”, but actually party Y won the election the rating was recoded to “1” as this indicated that the  $t_1$  article was not at all suggestive of the actual event—the victory of party Y.

## Results

### *Linguistic hindsight indicators*

We first ran a mixed measures analysis of variance with event category (elections, official decisions, personal decisions, disasters, sports events, and scientific discoveries) as between-event factor and article version ( $t_1$ ,  $t_2$ , and  $t_3$ ) as within-event factor. It yielded a significant main effect of article version,  $F(2, 54) = 3.52$ ,  $p = .037$ ,  $\eta_p^2 = 0.12$ , a significant main effect of event category,  $F(5, 27) = 4.36$ ,  $p = .005$ ,  $\eta_p^2 = 0.45$ , but no significant interaction,  $F < 1$  (see Table 1). There was an increase in the proportion of hindsight-related words across article versions. Specifically, version 3 contained a significantly higher percentage of hindsight-related words ( $M = 3.49$ ,  $SD = 1.61$ ) than article version 1 ( $M = 3.15$ ,  $SD = 1.68$ ),  $t(32) = 2.07$ ,  $p = .046$ , and article version 2 ( $M = 3.11$ ,  $SD = 1.86$ ),  $t(32) = 2.17$ ,  $p = .038$ . The main effect of event category was due to some categories containing a generally higher proportion of hindsight-related words (e.g., disasters:  $M = 2.87$ ,  $SD = 1.33$ , scientific events:  $M = 5.65$ ,  $SD = 1.33$ ).

### *Coded hindsight indicators*

We ran the same mixed measures analysis of variance as above with the rating whether the article was suggestive of a particular event (averaged across raters) as dependent variable. It revealed a significant main effect of article version,  $F(2, 54) = 3.74$ ,  $p = .030$ ,  $\eta_p^2 = 0.12$ , a significant main effect of event category,  $F(5, 27) = 3.77$ ,  $p = .010$ ,  $\eta_p^2 = 0.41$ , as well as a significant interaction of article version and event category,  $F(10, 54) = 2.62$ ,  $p = .011$ ,  $\eta_p^2 = 0.33$ . As can be seen in Fig. 1, it was the disaster category that showed a distinct pattern of results over time. A separate repeated-measures analyses of variance with article version ( $t_1$ ,  $t_2$ , and  $t_3$ ) of the disasters category as within-event variable yielded a significant main effect of article version,  $F(2, 10) = 5.92$ ,  $p = .02$ ,  $\eta_p^2 = 0.54$ . For none of the other event categories, we obtained significant differences in this hindsight indicator

**Table 1** Means and standard deviations (in parentheses) of the linguistic hindsight indicator in Study 1 depending on the article version ( $t_1$ ,  $t_2$ , and  $t_3$ ) and the event category

Event category	LIWC $t_1$	LIWC $t_2$	LIWC $t_3$
Elections	2.71 (1.76)	2.72 (1.84)	3.47 (1.25)
Public/official decisions	2.90 (1.37)	2.48 (0.93)	3.17 (0.88)
Personal decisions	3.32 (2.09)	3.30 (2.08)	3.60 (2.40)
Disasters	2.88 (1.56)	2.60 (1.17)	3.12 (1.10)
Sports events	2.18 (0.91)	2.08 (0.79)	2.11 (0.60)
Scientific discoveries	5.25 (0.99)	5.89 (1.81)	5.81 (0.66)

between-article versions,  $F_s < 1$ . As displayed in Fig. 1, the main effect of article version in the disasters category was entirely driven by the  $t_3$  ratings, which were higher than the  $t_2$ ,  $F(1,5) = 6.42, p = .05, \eta_p^2 = 0.56$ , and the  $t_1$  ratings,  $F(1,5) = 7.89, p = .04, \eta_p^2 = 0.61$ , which, in contrast, did not differ from one another,  $F(1,5) = .05, p = .83$ .

The number of explicit phrases expressing hindsight bias (averaged across raters) was analyzed the same way, but revealed no significant effects, all  $F_s < 1.933$ . Descriptively, the number of explicit phrases was low for all three article versions, version 1:  $M = 0.68, SD = 0.84$ , version 2:  $M = 0.49, SD = 0.54$ , version 3:  $M = 0.51, SD = 0.77$ .

*Relation between linguistic and coded hindsight indicators*

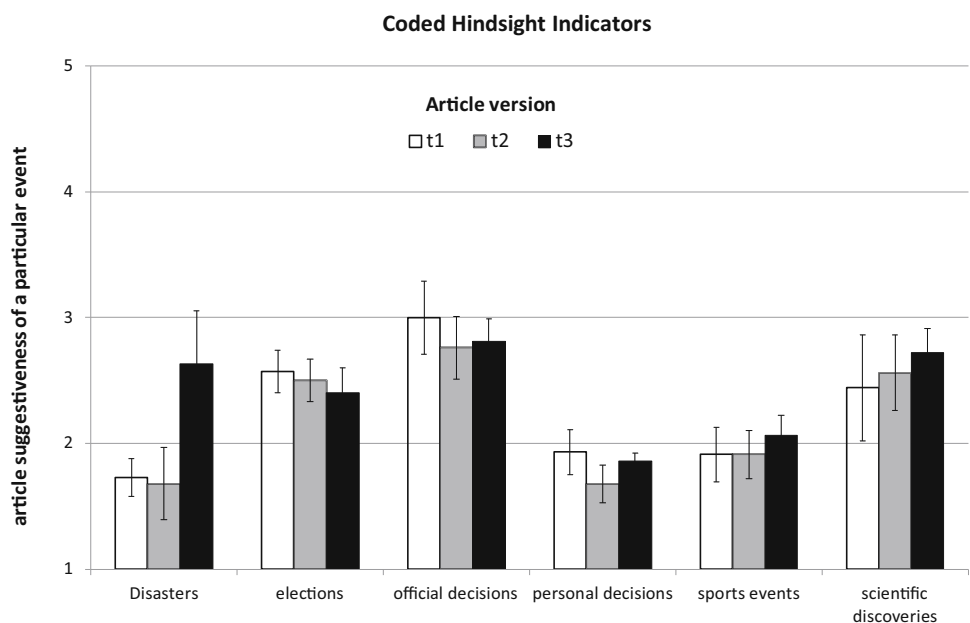
Furthermore, we tested whether the linguistic hindsight indicators (i.e., the proportion variable from the automatic text analyses) were predictive of the coded hindsight indicators (i.e., the ratings). We used a multilevel

regression approach for this purpose as the hindsight ratings and the proportions of hindsight words (Level-1) are nested within a single article (Level-2). Specifically, we computed a random intercept-random slope model in which the grand-mean centered linguistic hindsight indicators were used to predict coded hindsight indicators. The results of this model showed that higher values in the linguistic hindsight indicators go along with higher values in the coded hindsight indicators,  $b = 0.20, t(16.55) = 2.322, p = .033$ . However, this relationship differed considerably between articles, as indicated by a significant between-article slope variance:  $0.15, \Delta\chi^2 = 6.16, df = 1, p = .013$ . Further analysis showed that part of this variability could be explained by the event category the article belonged to: The relationship between the linguistic and the coded hindsight indicators was marginally significantly different from zero for the disaster category,  $b = 0.43, t(6.98) = 2.27, p = .058$  but not for any other event category, all  $t_s < 1$ . Therefore, the automatic text analysis indicators are related to the ratings obtained by the ten coders and the relation seems to be stronger for the disaster category.

**Discussion**

We investigated whether there is evidence for hindsight distortions in Wikipedia articles or whether Wikipedia’s guidelines effectively prevent hindsight bias to occur. Our study provides empirical evidence for both. On the one hand, we found evidence for a hindsight bias on two hindsight indicators we made use of. Hindsight articles—particularly later ones ( $t_3$ )—contained a greater percentage

**Fig. 1** Average coded hindsight indicators (error bars are SEs) as a function of article version and event category



of linguistic markers of hindsight bias and were also rated as more suggestive of the event than foresight articles. In other words, they implied to a systematically greater extent that the event was likely to happen. On the other hand, our coded hindsight indicator, which was a more thorough and fine-grained content analysis, revealed that this increase was mainly due to one particular event category: disasters. In other words, the overwhelming majority of articles did not show traces of hindsight bias. This is noteworthy in consideration of the fact that hindsight bias has been documented as a robust, widespread, and difficult to overcome bias (Guilbault et al., 2004; Roese & Vohs, 2012). Moreover, in the case of elections, it is of particular interest as there are numerous demonstrations of hindsight bias in individuals (e.g., Blank, Fischer, & Erdfelder, 2003; Blank & Nestler, 2006; Fischer & Budescu, 1995; Leary, 1982; Powell, 1988). In other words, our findings substantially deviate from prior research on hindsight bias and despite the fact that null effects should not be over-interpreted one may question whether research on individual biases may be generalized to Wikipedia articles. We have argued above that Wikipedia differs substantially from usual lab research in that it conceives itself as an encyclopedia, which may potentially prime accuracy motivation in its contributors, and that there are a number of guidelines that aim at preventing bias. In Wikipedia, authors are not asked to freely express their personal evaluations and the demand to insert verifiable information from reliable sources obviously raises the threshold to obtain hindsight bias—given that we did not find strong evidence for hindsight bias for the majority of articles. We will return to this issue—and the question of why articles about disasters contained a hindsight bias nevertheless—in the “[General Discussion](#)”.

In line with this reasoning, the evidence for hindsight bias we found was indirect rather than explicit, which would be typical for hindsight bias (e.g., “It was no surprise that...”). There were hardly any such phrases found and, more importantly, we did not obtain any increase over article versions. Instead, hindsight bias was evident more subtly by the more frequent use of hindsight-related words as well as by causal elaborations: The fact that a significant increase was obtained only for  $t_3$  article versions is consistent with prior research showing that outcome knowledge alone (here in  $t_2$  article version) does not elicit hindsight bias (Nestler & Egloff, 2009; Yopchick & Kim, 2012). Rather, (one-sided) causal elaborations are a necessary precondition (Nestler et al., 2008; Yopchick & Kim, 2012).

In sum then, our linguistic indicators suggest a hindsight bias, in general, whereas our coded indicators argue for hindsight bias only in the case of disasters. This is an interesting issue and we will get back to it in the “[General discussion](#)”. One might question, however, whether the

coded indicators truly reflect a hindsight bias in the article. After all, one could argue that it is the raters’ own individual biases that are reflected in the ratings rather than the article contents themselves. If our coded hindsight indicators were related to the authors’ own hindsight bias, however, it would provide a validation of our measure and ensure that the coded hindsight indicators actually reflect article contents. It is impossible to track this information in Wikipedia. Therefore, we conducted a lab experiment. In addition, this experiment allowed us to examine the effect of collaboration on the magnitude of hindsight bias in the produced articles. Although the number of authors was not predictive of hindsight bias in Study 1, there was, in fact, no article in which collaboration did not take place. In Study 2, we had participants write articles either individually or collaboratively to assess whether collaboration moderates the magnitude of hindsight bias in the resultant articles.

## Study 2

In this lab study, we presented all participants with identical information about a fictitious dam in Spain. Participants in the hindsight condition additionally learned that the dam collapsed. All of them were then asked to produce a “Wikipedia-like” article about the dam. Prior to this, we thoroughly informed them about the guidelines operating in Wikipedia. Article writing took place either individually or collaboratively. In addition, we assessed participants’ personal hindsight biases and let blind observers code the resulting articles for hindsight bias.

### Participants and design

One hundred and seventy-six participants (141 female,  $M_{\text{age}} = 22.80$ ,  $SD = 5.32$ ; range = 18–68) were invited to participate in a lab experiment by personal e-mail for monetary reward. Participants were randomly assigned to one of the four experimental conditions that resulted from our  $2 \times 2$  between-subjects design. Participants either received no outcome information (foresight condition) or learned about the disaster (hindsight condition). In addition, participants authored the article either in groups of three (collaborative writing condition) or alone (individual writing condition).

### Materials

To ensure that none of the participants had prior knowledge of the event in question, we developed fictitious material about an alleged dam in a touristic region of the Pyrenees, Spain. Participants received 11 bogus articles from Spanish



and German newspapers containing information (all in German) of varying relevance to the subject and arguments for and against the alleged event outcome. For example, one article described the state-of-the-art construction of the dam, whereas another one reported public protests against the dam due to safety concerns. Participants in the hindsight conditions were additionally informed that this dam collapsed in 1993 and the consecutive flooding caused severe devastation in neighboring villages.

### *Pilot study*

To test whether the material elicited a hindsight bias, we conducted a pilot study with  $N = 56$  people (40 female,  $M_{\text{age}} = 27.07$ ,  $SD = 9.16$ , range 19–60), who read the same materials. Some participants of the pilot study were informed about the collapse (i.e., hindsight condition) and some were not. All participants then estimated the likelihood of four different, mutually exclusive events (including the actual outcome) in percent (hindsight participants were urged to ignore their outcome knowledge), their impressions of foreseeability (seven items, e.g., “I would have foreseen that this accident was going to happen”, 1 = not at all to 5 = very much; Cronbach’s  $\alpha = 0.764$ ) and their impressions of inevitability (four items; e.g. “Sooner or later there had to be an accident”; 1 = not at all to 5 = very much; Cronbach’s  $\alpha = 0.815$ ). A MANOVA with all three dependent variables (likelihood of the actual event, foreseeability, and inevitability) yielded a significant effect of condition, Wilk’s  $\lambda = 0.718$ ,  $F(3, 52) = 6.801$ ,  $p = .001$ ,  $\eta^2 = 0.282$ . There was a significant hindsight bias with regard to likelihood ratings ( $M_{\text{hindsight}} = 15.59\%$ ,  $SD = 11.11$ ,  $M_{\text{foresight}} = 6.07\%$ ,  $SD = 6.48$ ),  $F(1, 54) = 15.012$ ,  $p < .001$ ,  $\eta^2 = 0.218$ , and with respect to impressions of inevitability ( $M_{\text{hindsight}} = 2.96$ ,  $SD = 1.03$ ,  $M_{\text{foresight}} = 2.13$ ,  $SD = 0.57$ ),  $F(1, 54) = 13.788$ ,  $p < .001$ ,  $\eta^2 = 0.203$ . Descriptively, differences in foreseeability impressions were into the same direction ( $M_{\text{hindsight}} = 2.95$ ,  $SD = 0.73$ ,  $M_{\text{foresight}} = 2.63$ ,  $SD = 0.74$ ), but failed to reach significance,  $F(1, 54) = 2.626$ ,  $p = .111$ . Note, however, that impressions of foreseeability and inevitability do not necessarily work in parallel (e.g., Blank et al., 2008; Nestler, Blank, & Egloff, 2010; Nestler & Egloff, 2009). Given the significant and large hindsight bias for likelihood and inevitability ratings, we regarded the materials as suitable for our main study.

### **Procedure**

After acknowledging legal information and agreeing to voluntarily participate, all participants received a booklet containing 11 alleged newspaper articles about a lake in the

Pyrenees. Information was given in short, stand-alone articles in non-chronological order, resembling a collection of cut out newspaper articles from various sources. Participants in the hindsight condition received an additional short article informing them about the outcome. Next, participants received detailed instructions about characteristics of Wikipedia articles and were asked to write such an article on the basis of the information from the newspaper articles. The writing process was realized with the online collaboration tool (<http://www.etherpad.com>) on laptop computers. The tool enabled simultaneous writing by participants in the group condition and contained basic text format editing options and a chat function. Writing time was set to 35–40 min. Afterwards, participants filled out an online questionnaire and were asked to rate the likelihood of four mutually exclusive events (same as in pretest), their impression of foreseeability, and their impression of inevitability (the same items as in the pretest with Cronbach’s  $\alpha = 0.761$  and Cronbach’s  $\alpha = 0.692$  for foreseeability and inevitability, respectively). Analogous to the pretest procedure, participants in the foresight condition gave these ratings for two different events, one being the actual outcome. Finally, there were some questions regarding prior knowledge of the event, general trust in Wikipedia, Wikipedia engagement, and basic demographic information. After debriefing, participants were paid and thanked.

### **Article analyses**

For analyzing the produced articles, we made use of the same hindsight indicators as in Study 1. First, we had three independent raters who were blind to the experimental conditions of the articles rate each article according to a shortened version of the coding scheme used in Study 1. The main dependent variable was—identical to Study 1—the extent to which the articles suggested that a disaster at the dam was likely, inevitable, and foreseeable (1 = not at all, 5 = very much). Coders were trained on a subset of 20 articles (ICC = 0.864) and yielded an acceptable agreement (ICC = 0.712). Second, we conducted an automatic text analysis of the articles with LIWC and measured the percentage of hindsight-related words of the LIWC categories “cause”, “insight”, “certainty”, “tentativeness”, and “discrepancy” as in Study 1.

### **Results**

#### *Individual perceptions*

We first analyzed individuals’ perceptions to ensure that they actually exhibited a hindsight bias—before analyzing whether their bias translated into articles. To this end, we

conducted a multivariate ANOVA with likelihood, inevitability, and foreseeability ratings as dependent variables and information condition (foresight and hindsight) and writing condition (individual and collaborative) as independent between-subjects factors. It yielded only a significant main effect of information condition, Wilk's  $\lambda = 0.631$ ,  $F(3, 84) = 16.393$ ,  $p < .001$ ,  $\eta^2 = 0.369$ . Neither writing condition, Wilk's  $\lambda = 0.975$ ,  $F(3, 84) = 0.709$ ,  $p = .549$ ,  $\eta^2 = 0.025$ , nor the interaction was significant, Wilk's  $\lambda = 0.995$ ,  $F(3, 84) = 0.151$ ,  $p = .929$ ,  $\eta^2 = 0.005$ . The effect of information condition was due to significant differences in all dependent variables (see Table 2 for means),  $F_{\text{likelihood}}(1, 86) = 28.651$ ,  $p < .001$ ,  $\eta^2 = 0.250$ ,  $F_{\text{inevitability}}(1, 86) = 40.476$ ,  $p < .001$ ,  $\eta^2 = 0.320$ ,  $F_{\text{foreseeability}}(1, 86) = 6.501$ ,  $p = .013$ ,  $\eta^2 = 0.070$ . Outcome knowledge thus had large effects on participants' individual perceptions. With the benefit of hindsight, they perceived the collapse of the dam to be more likely, more inevitable, and more foreseeable than participants without outcome knowledge (foresight condition). Did this translate into more biased articles?

## Article analyses

### *Coded hindsight indicator*

An ANOVA with the averaged hindsight ratings of all three raters as dependent variable and with information condition (foresight, hindsight) as well as writing condition (collaborative, individual) as between-article factors yielded only a significant main effect of information condition,  $F(1, 86) = 12.298$ ,  $p = .001$ ,  $\eta_p^2 = 0.125$ . The main effect of writing condition,  $F(1, 86) = 0.064$ ,  $p = .800$ , as well as the interaction,  $F(1, 86) = 0.002$ ,  $p = .967$ , were not significant. Trained observers thus judged the articles in the hindsight condition to be significantly more suggestive of a disaster ( $M = 2.49$ ,  $SD = 0.75$ ) than the articles from participants of the foresight condition ( $M = 1.92$ ,  $SD = 0.78$ ).

### *Linguistic hindsight indicator*

An analysis of the percentage of words related to the hindsight perspective with information condition

**Table 2** Mean personal evaluations (SDs) in Study 2 as a function of outcome knowledge

	Foresight	Hindsight
Likelihood	7.14 (8.36)	21.71 (16.57)
Inevitability	2.53 (0.45)	3.25 (0.61)
Foreseeability	2.56 (0.44)	2.84 (0.59)

Since the 11-point likelihood scale reflected percentages (0–100%), we present likelihood estimates as percentages

(foresight, hindsight) and writing condition (individual, collaborative) as between-article factors yielded no significant effects at all,  $F_s < 1.12$ ,  $p_s > .290$ .

### *Article ratings and individuals' evaluations*

In a further step, we examined directly, whether participants' individual biases translated into the article by examining whether individuals' evaluations regarding the likelihood, inevitability, and foreseeability of the disaster were related to the article ratings. To this end, we estimated a multilevel model accounting for the partial nesting of individuals in groups in the group condition (see Bauer, Sterba, & Hallfors, 2008, or Sterba, 2017, for a description of the adapted multilevel model for partially nested designs).

In the model, we used article ratings to predict each individual measure (likelihood, inevitability, and foreseeability). Results of the multilevel model showed that article ratings were significantly associated with individual likelihood ratings,  $b = 6.41$ ,  $t(96.09) = 4.69$ ,  $p < .001$ , inevitability ratings,  $b = 0.17$ ,  $t(110.08) = 2.80$ ,  $p = .006$ , but not foreseeability ratings,  $b = 0.06$ ,  $t(111.13) = 0.99$ ,  $p = .32$ . Finally, the LIWC measure of the article was also significantly predicted by the article ratings,  $b = 0.22$ ,  $t(154.89) = 2.42$ ,  $p = .017$ .

## Discussion

In this study, we sought to replicate the effect of Study 1 under controlled conditions and to validate our coded hindsight indicators. Having provided participants with the exact same information, we found the classic hindsight bias: participants with outcome knowledge perceived the event—the collapse of the dam in this case—as more likely, inevitable, and foreseeable than participants in the foresight condition. More importantly, however, individuals' hindsight bias entered their articles. Articles about the dam, which had been authored by participants with outcome knowledge, were significantly more suggestive of a disaster than were articles that had been written by participants without outcome knowledge. Furthermore, the hindsight bias present in the articles was clearly linked to the authors' individual biases, which does not only validate our coded hindsight indicators, but also provides direct evidence for the translation of individual biases into article biases. This is remarkable in consideration of the fact that we had urged participants to follow Wikipedia's guidelines and several indices, indeed, show their compliance: For instance, participants frequently inserted references to the sources of the information they contributed. In addition, they mostly used a very neutral language for their presentation. Interestingly, writing condition had no

substantial impact on this process—collaboration neither reduced nor increased the resultant hindsight bias in the articles. This is consistent with Study 1, as well. Recall, that Study 2 made use of a disaster—the only category of events, for which we had obtained evidence for a hindsight bias in Study 1.

In sum then, the findings of this study validate the hindsight effects we found in Study 1. The fact that we did not obtain any effects with our objective hindsight measure needs to be discussed, though. Due to the fact that participants in our lab study had much less time for their article construction than actual Wikipedia authors do, the articles produced in this study differ from actual Wikipedia articles in several dimensions: they are shorter, less elaborated, and of lower quality. Therefore, the chance to detect differences is lower than for actual Wikipedia articles. In addition, one might conclude from our findings that the coded hindsight indicators are actually the more sensitive ones. Having provided evidence for their validity in Study 2, the event-specific pattern of the coded hindsight indicators obtained in Study 1 argues for this notion. We will return to the event-specific hindsight bias in Wikipedia articles in the “[General Discussion](#)”. There is one aspect in which Study 1 and Study 2 differ from one another, which might argue for the notion that our lab results even underestimate the effects that might be obtained in Wikipedia—at least in the special case of disasters. As outlined above, disasters and calamities are characterized by their unexpectedness. At the same time, this sort of event often attracts a large number of authors—many of whom have not already previously contributed to the article or not even contributed to Wikipedia at all beforehand (Keegan et al., 2011; Oeberst et al., 2014a, b). In other words, these are—also—people who are attracted to the topic only after the fact and who, therefore, have retrieved and searched for information exclusively with the benefit of hindsight. In our lab experiment, we first presented all participants with the same information and then informed some of the collapse. The real-world equivalent, however, would be that they first hear of the disaster and then read the information that is available. It is reasonable to assume that hindsight bias would be much greater in this case, because all information is already perceived and evaluated in the light of the outcome.

### Study 3

Having provided evidence that Wikipedia articles about disasters contain a hindsight bias, we now turn to an article’s effects on readers’ personal perceptions regarding the event in question. It suggests itself that reading Wikipedia articles that contain a hindsight bias might increase

readers’ subjective perceptions of likelihood, inevitability, and/or foreseeability of past events, whereas the reception of unbiased Wikipedia articles should not. Specifically, we propose biased articles to (1) elicit a hindsight bias in people who have not heard of the event beforehand and to (2) increase hindsight bias in people who already knew of the event—and may even have already developed a hindsight bias prior to reading the biased article.

## Method

### *Participants and design*

Altogether, 135 participants (106 female,  $M_{\text{age}} = 24.99$ ,  $SD = 6.90$ ) completed our online experiment in return for the chance to win vouchers for online stores. None of them had heard of the unknown event prior to the study. They were randomly assigned to one of three experimental conditions—the foresight condition ( $N = 53$ ), the hindsight condition ( $N = 44$ ), and the hindsight plus article condition ( $N = 38$ ).

### *Materials and procedure*

Participants were invited via mailing lists to an online study on the perception of events. After agreeing to participate in the study and acknowledging legal and ethical information participants were randomly assigned to one of the three conditions (see above). On the next page, participants read that we were interested in their perception and evaluation of an event, and were asked to carefully read the presented material.

Participants in the foresight and hindsight condition received general information about the Shushenskaya hydroelectric power station in Siberia. Participants in the hindsight condition additionally received outcome information. Specifically, they were informed of the accident that took place on August 17, 2009, in which 75 people died (“On August 17, 2009, there was an accident with 75 deaths at the Sayano–Shushenskaya hydroelectric power station. It was caused by the flooding of the engine house after several pipes broke due to high water pressure.”). Participants in the hindsight plus article condition read the  $t_3$  article version about the hydroelectric power station, which we had used in Study 1. This article version contained detailed information about the accident on August 17, 2009 and potential causal antecedents. In Study 1, our indicators of hindsight bias had revealed that the article was highly suggestive of the event—in hindsight, but not in foresight.

Next, participants in all experimental conditions were asked for their personal likelihood estimates of four alternative events including the original accident (in percent,

summing up to 100% for all four events). Participants in the hindsight conditions were urged to ignore their outcome knowledge when answering this question. The alternative events were phrased to be mutually exclusive. Afterwards, participants answered six items rating their personal impressions of inevitability (see Blank et al., 2008 for item wordings; 1 = disagree, 5 = agree) and another three items tapping their personal impression of foreseeability on a 5-point scale (1 = disagree, 5 = agree). For the foreseeability scale, Cronbach's  $\alpha$  was 0.67. The six items of the inevitability scale, in contrast, proved to be inconsistent (Cronbach's  $\alpha = 0.40$ ). We, therefore, ran a factor analysis and aggregated the three items with the highest loadings on the first factor (explaining 25% of the variance; Cronbach's  $\alpha = 0.63$ ).

Subsequently, participants were asked whether they had heard of the event before, whether they knew about its causes and whether they had been familiar with the original Wikipedia article before participating in our study. After providing information about their general trust in Wikipedia, Wikipedia engagement, and basic demographic information, participants were debriefed and informed about the fact that the article they had read was old.

## Results and discussion

### Hindsight bias

Given that we expected a linear increase in participants' perceptions of likelihood, inevitability, and foreseeability (foresight < hindsight < hindsight plus article), we computed linear contrasts for all three dependent variables with foresight condition coded as  $c = -1$ , hindsight condition coded as  $c = 0$ , and hindsight plus Wikipedia article condition coded as  $c = 1$  (see Table 3 for descriptives). All three contrasts yielded significant linear increases for each of the three dependent variables,  $F_{\text{likelihood}}(1, 132) = 5.08$ ,  $p = .03$ ,  $\eta_p^2 = 0.03$ ,  $F_{\text{foreseeability}}(1, 132) = 4.53$ ,  $p = .04$ ,  $\eta_p^2 = 0.03$ ,  $F_{\text{inevitability}}(1, 132) = 35.83$ ,  $p < .001$ ,  $\eta_p^2 = 0.21$ . In addition, we compared the hindsight condition and the hindsight plus Wikipedia article condition directly to inspect the additional effect of reading more closely. There was a significant difference with regard to perceived inevitability,  $t(80) = 2.85$ ,  $p < .01$ ,  $d = 0.62$ , and a marginally significant difference into the predicted direction with regard to perceived likelihood,  $t(80) = 1.73$ ,  $p = .09$ ,  $d = 0.38$ . The difference with regard to perceived foreseeability was likewise into the predicted direction, however, fell short of significance,  $t(80) = 1.44$ ,  $p = .15$ . Taken together, we found the perceived likelihood, inevitability, and foreseeability for the Shushenskaya disaster to increase with outcome knowledge—the classic hindsight bias—but also and in addition to the classic

**Table 3** Mean hindsight evaluations (SDs) of Study 3 as a function of outcome knowledge and article reading

	Foresight	Hindsight	Hindsight + article
Likelihood <sup>a</sup>	17.47 (14.23)	18.47 (18.85)	25.82 (19.60)
Inevitability	2.36 (0.70)	2.82 (0.63)	3.27 (0.83)
Foreseeability	1.96 (0.83)	2.08 (0.71)	2.32 (0.85)

<sup>a</sup> Since the 11-point likelihood scale reflected percentages (0–100%), we present likelihood estimates as percentages

hindsight bias, we found perceptions of likelihood and inevitability to increase with the perception of the  $t_3$  Wikipedia article version.

## General discussion

Web 2.0 enables laypersons to collaborate with others at an unprecedented scale. Such mass collaboration has numerous benefits as documented in a number of studies (see Cress et al., 2016). Moreover, mass collaboration comes along with a certain context such as rules and norms that guide collaboration. In case of Wikipedia, several norms aim at the prevention of personal evaluations to ensure the construction of recognized knowledge, which is the ultimate goal of any encyclopedia. It is for this reason that we chose to examine Wikipedia articles with regard to hindsight bias: although prior research has documented the robustness and pervasiveness of hindsight bias, it has solely been investigated in terms of individuals' subjective perceptions and evaluations—that is, in reception—but never in production in a context like Wikipedia. Another reason was Wikipedia's popularity. If Wikipedia articles were biased, this would likely shape the views of millions.

### Hindsight bias in the production of Wikipedia articles

With regard to Wikipedia articles, we found evidence for hindsight bias only in one particular category of events, namely disasters, but not in other event categories. We will first discuss the absence of hindsight bias in the majority of articles and then turn to the disasters category.

The absence of hindsight bias in the majority of articles is of great interest and importance as it stands in stark contrast to prior research with individuals. Although great caution is warranted when interpreting null effects, it is noteworthy that research with individuals has documented hindsight bias as a robust and pervasive error (e.g., Guilbault et al., 2004; Pohl et al., 2002). Therefore, it seems unlikely that we simply might have accidentally selected events for which no hindsight bias had occurred. For instance, with regard to elections, our finding lacking



evidence of hindsight bias in Wikipedia articles contrasts a substantial body of research documenting hindsight bias in individuals (e.g., Blank, Fischer, & Erdfelder, 2003; Blank & Nestler, 2006; Fischer & Budescu, 1995; Leary, 1982; Powell, 1988).

A potential explanation for our results is that Wikipedia's authors may have—personally—succumbed to hindsight bias, but that their hindsight bias did not enter the article. One reason for this could be Wikipedia's request for verifiable and reliable information. Consequently, if individuals' hindsight bias was based on information that is not verifiable and from a reliable source, it should not be inserted into a Wikipedia article or be deleted by others if someone included it nevertheless (e.g., Oeberst et al., 2014a, b). Hence, if individual's biases are based on personal opinions or speculations about why the event has happened one would expect a contrast in the results of research with individuals and research with Wikipedia articles. In the case of the marriage of Prince William and Kate Middleton, for instance, people might be personally convinced that they foresaw the marriage all along, but as long as they do not have verifiable information from reliable sources that support this conviction, it will not be included into the article. Wikipedia's rules might thus possibly heighten the threshold for biases to enter and effectively lead to fewer instances of hindsight bias—even if they may not entirely preclude it as we saw for the disasters category.

Only with regard to disasters we found that later articles suggested to a greater extent that the disaster was more likely, more inevitable, and more foreseeable compared to earlier (foresight) articles. Interestingly, disasters have hardly ever been used in hindsight bias research (see Verplanken & Pieters, 1988, for an exception), presumably, because they pose methodological challenges for hindsight researchers (e.g., Hawkins & Hastie, 1990). With the benefit of hindsight, it is plausible why it was particularly the category of disasters that exhibited a hindsight bias: usually, disasters are not initially expected and thus surprising. They are also negative and mostly consequential by causing death, injuries, or damage, thereby eliciting a particular need to explain how it could happen which, in turn, fosters hindsight distortions—at least if one can come up with an explanation (Ash, 2009; Guilbault et al., 2004; Musch, 2003; Pezzo, 2003; Schkade & Kilbourne, 1991). From this perspective, it becomes clear how an article may be highly suggestive of the event (in hindsight), even though it does not contain any explicit phrases expressing hindsight bias (e.g., “It was clear that...”). If the article contains an explanation which suggests that all antecedents spoke for the occurrence of this event (as post hoc explanations lack an appreciation of event-inconsistent antecedents), the article itself becomes highly suggestive of the event. Thus, one might expect a pronounced hindsight bias

in the case of disasters, which might, in turn, transfer to a “collective” hindsight bias in Wikipedia articles, as Wikipedia's norms may heighten the threshold for hindsight bias but not prevent it all along. Recall, that we have argued that hindsight bias is not per se in conflict with Wikipedia's guidelines and the results of our lab study (Study 2) corroborate this notion. As long as verifiable information from reliable sources is available and presented neutrally, it may go unnoticed that it is biased by the fact that it focuses on event-consistent information and underweights or ignores event-inconsistent information, which is characteristic for hindsight bias (Carli, 1999; Nestler et al., 2008). In sum then, we suggest that Wikipedia's norms contribute to an unbiased presentation of events (see also Postmes et al., 2001), but may not prevent any bias to occur. Although this interpretation has to be tested in future research, it raises an exciting novel possibility to reduce or prevent hindsight bias.

Despite the fact that we found evidence for hindsight bias in only one out of many different event categories, the relevance of our results should not be underestimated for at least two reasons: First, disasters and calamities usually attract a particularly broad audience (e.g., Keegan et al., 2011). The Fukushima article, for instance, was retrieved more than 100,000 times in May 2011 alone (the time frame into which our  $t_3$  article version falls: <http://www.stats.grok.se>; this number includes the traffic to the article regarding the nuclear power plant as well as the newly created article “Nuclear disaster of Fukushima-Daiichi” to which the elaborations regarding the disaster were migrated). In other words, even if only certain Wikipedia articles might be biased by hindsight, our results indicate that these could likely reach a great number of people.

A second aspect that speaks to the same argument is that highly negative events such as disasters are closely linked to questions of responsibility and guilt (e.g., Harley, 2007; Rachlinski, 1998). Particularly if a damage or harm seems to be foreseeable in hindsight and thus as preventable accusations of negligence come to mind and are frequently voiced. This may not only result in broadly shared—but biased—attributions of guilt but even in juridical affirmations of negligence that are biased by hindsight (e.g., Hastie, Schkade, & Payne, 1999; LaBine & LaBine, 1996; Smith & Greene, 2005).

### Elicitation and increases of hindsight bias through reading Wikipedia articles

In consideration of the fact that we did find some articles to contain traces of hindsight bias, we further examined their impact on readers. The present findings show that reading biased Wikipedia articles elicits a hindsight bias in readers who are unfamiliar with the event. In addition, and beyond



prior research, we also found that hindsight bias further increased readers' already existent hindsight bias when they knew already about the event. The latter result is interesting as we provided a cause for the event outcome in all hindsight conditions. Informing vs. not informing about a cause for the accident (Yopchick & Kim, 2012) hence cannot explain these findings. However, we believe that causal modeling could be otherwise involved: Wikipedia articles could (1) add new knowledge (i.e., causes) to participants' existing causal model, (2) reactivate participants' existing causal model, or (3) provide a coherent presentation of the causal information, which participants had previously lacked. The first aspect might explain the results of our Study 3 as the Wikipedia article had contained more information than what had been presented to participants in the classic hindsight condition. The second and third explanations may be particularly relevant in real-world settings: When time between learning the outcome and reading about it has passed, it is possible that reading the article reactivates the causal model which, in turn, contributes to participants' evaluations of the event. Moreover, we usually learn of real-world events by receiving information that is distributed over time (e.g., news-ticker) and/or sources (e.g., news reports). If an article then provides readers with a single coherent presentation of the entire event, it may foster comprehension and a coherent causal model (Pennington & Hastie, 1988; see also McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996).

### Limitations and future prospects

To examine the occurrence of hindsight bias in Wikipedia articles, a foresight article version of that article was necessary. Hence, the present results are limited to events for which articles existed before the event took place (e.g., the power plant article before the Fukushima disaster). Therefore, we cannot exclude selection effects. Presumably, the existence of an article about a topic in Wikipedia likely depends on the relevance of the topic. Thus, it remains unclear whether our results are generalizable to rather irrelevant topics. Recall, however, that we had also included unfamiliar events, but popularity of the event did not affect the results. In addition, we have included events from various different event categories, which clearly extends prior research, which typically focused on one event or a particular category of events (e.g., elections).

Concerning future prospects, we believe that it might be interesting to investigate whether hindsight bias is present in other collaborative products, as well. We have argued that biases will enter collaborative products if they are widely shared and when there are no guidelines effectively preventing their occurrence. This reasoning implies that

our findings should not be limited to Wikipedia articles, but the bias is likely to be present in other media as well, as long as the production rules will not prevent it. Especially for disasters, one often finds post hoc articles claiming that a disaster was inevitable and foreseeable (e.g. The Telegraph, 2011; The Express Tribune, 2011), whereas foresight articles warning about the upcoming disaster (which should be possible if it was, indeed, foreseeable) are missing. Second, our reasoning could well extend to other biases (e.g., descriptions favoring the own group in inter-group conflicts, see Oeberst, Cress, Back, & Nestler, 2016). Again, we believe that this is an interesting endeavor for future research.

### Implications for the production of collective knowledge

With regard to the accuracy of Wikipedia, it has repeatedly been documented that Wikipedia articles are comparatively accurate (e.g., Giles, 2005). In addition, research showed that there is a positive relationship between the number of authors and the quality of an article (Kittur & Kraut, 2008). Interestingly, however, we obtained evidence for bias particularly in the category of articles that usually attracts an extraordinary high number of authors (e.g., Keegan et al., 2011; Oeberst et al., 2014a, b). More authors, hence, do not automatically lead to less biased articles. Probably, it is not the number of authors that is decisive, but rather the heterogeneity of the authors involved (Schulz-Hardt, Frey, Lüthgens, & Moscovici, 2000; Schulz-Hardt, Jochims, & Frey, 2002). In addition, while a larger number of authors increases the likelihood of a heterogeneous perspective (e.g., in a controversy), it may be of no effect in cases of widely shared and potentially large biases such as the hindsight bias in the context of disasters.

### Conclusion

To conclude, the present studies extend prior research on hindsight bias in individuals to a collective level and point to interesting differences between varying types of events as well as the potential power of guidelines, such as present in Wikipedia, on the prevention of hindsight bias. Moreover, our findings indicate that biased Wikipedia articles may, again, nourish hindsight bias in individuals who read those articles. As this paper has hopefully shown, the use of a highly ecological setting led to empirical and theoretical advances and identified several questions for future research, which, in the long run, may foster a more elaborated understanding of biases in the real world.

### Compliance with ethical standards

**Funding** This research was funded by the German Research Foundation (Grant Numbers OE 604/1-1 and NE 1485/5-1).

**Conflict of interest** All authors declare that they have no conflicts of interest.

**Ethical standards** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

### References

- Adler, B. T., Alfaro, L., de Mola-Velasco, S. M., Rosso, P., & West, A. G. (2011). Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (pp. 277–288). Berlin: Springer.
- 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*, 916–933.
- Barron, B., Martin, C. K., Mertl, V., & Yassine, M. (2016). Citizen science: Connecting to nature through networks. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education*. Cham: Springer International Publishing.
- Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research*, *43*, 210–236.
- Bellomi, F., & Bonato, R. (2005). Network analysis for Wikipedia. In *Proceedings of the first international Wikimedia conference Wikimania*. Retrieved from <http://www.uvm.edu/~pdodds/files/papers/others/2009/bellomi2009a.pdf>. Accessed 10 Apr 2017.
- Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, *11*, 491–504.
- Blank, H., & Nestler, S. (2006). Perceiving events as both inevitable and unforeseeable in hindsight: The Leipzig candidacy for the Olympics. *British Journal of Social Psychology*, *45*, 149–160.
- Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition*, *25*, 132–146.
- Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? *Cognition*, *106*, 1408–1440.
- Bukszar, E., & Conolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, *31*, 628–641.
- Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology*, *62*, 1899–1915.
- Carli, L. L. (1999). Cognitive reconstruction, hindsight, and reactions to victims and perpetrators. *Personality and Social Psychology Bulletin*, *25*, 966–979.
- Chen, S., Shechter, D., & Chaiken, S. (1996). Getting at the truth or getting along: accuracy- versus impression-motivated heuristic and systematic processing. *Journal of Personality and Social Psychology*, *71*, 262–275.
- Choi, I., Koo, M., & Choi, J. A. (2007). Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin*, *33*, 691–705.
- Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*, 147–168.
- Cress, U., Jeong, H., & Moskaliuk, J. (2016). Mass collaboration as an emerging paradigm for education? Theories, cases, and research methods. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education*. Cham: Springer International Publishing.
- Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, *3*, 105–122.
- Fallis, D. (2008). Toward an epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*, 1662–1674.
- Ferron, M., & Massa, P. (2011). Collective memory building in Wikipedia: The case of North African uprisings. In *Proceedings of the Wikisym*. Retrieved from <http://www.slideshare.net/phauly/collective-memory-building-in-wikipedia-the-case-of-north-african-uprisings>. Accessed 10 Apr 2017.
- Fields, D. A., Kafai, Y. B., & Giang, M. T. (2016). Coding by choice: A transitional analysis of social participation patterns and programming contributions in the online Scratch community. In U. Cress, J. Moskaliuk, & H. Jeong (Eds.), *Mass collaboration and education*. Cham: Springer International Publishing. (in press).
- Fischer, I., & Budescu, D. V. (1995). Desirability and hindsight biases in predicting results of a multi-party election. In J.-P. Caverni, M. Bar-Hillel, F. H. Barron, & H. Jungermann (Eds.), *Contributions to decision making* (Vol. 1, pp. 193–211). Amsterdam: Elsevier.
- Fischhoff, B. (1975). Hindsight  $\neq$  foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299.
- Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 349–358.
- Forte, A., & Bruckman, A. (2008). Scaling consensus: Increasing decentralization in Wikipedia governance. In *Proceedings of the 41st annual Hawaii international conference on system sciences (HICSS'08)*. Washington, DC: IEEE.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*, 900–901.
- Gowers, T., & Nielsen, M. (2009). Massively collaborative mathematics. *Nature*, *461*, 879.
- 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*, 103–117.
- Hayes, J. R. (1989). Writing research: The analysis of a very complex task. In D. Klahr, Kotovsky, K. (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 209–234). Lawrence Erlbaum: Hillsdale, New Jersey.
- Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition*, *25*, 48–63.
- 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*, 597–614.
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgment of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327.
- Hecht, B., & Gergle, D. (2009). Measuring self-focus bias in community-maintained knowledge repositories. In *Proceedings of the 4th international conference on communities and technologies* (pp. 11–20). New York: ACM.
- Hecht, B., & Gergle, D. (2010). The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual

- context. In *CHI '10 Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 291–300). New York: ACM.
- Isenberg, D. J. (1986). Group polarization: a critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*, 1141–1151.
- Jennings, M. M., Lowe, D. J., & Reckers, P. M. J. (1998). Causality as an influence on hindsight bias: An empirical examination of judges' evaluation of professional audit judgment. *Journal of Accounting and Public Policy*, *17*, 143–167.
- Keegan, B., Gergle, D., & Contractor, N. (2011). Hot off the wiki: Dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. In *Proceedings of the 7th international symposium on wikis and open collaboration (WikiSym'11)* (pp. 105–113). New York: ACM Press.
- Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in Wikipedia: Quality through coordination. In *Proceedings of the 2008 ACM conference on computer supported cooperative work* (pp. 37–46). New York: ACM Press.
- Küfner, A. C. P., Nestler, S., Back, M. D., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality*, *44*, 427–435.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.
- LaBine, S. J., & LaBine, G. (1996). Determinations of negligence and the hindsight bias. *Law and Human Behavior*, *20*, 501–516.
- Leary, M. R. (1982). Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin*, *8*, 257–263.
- Louie, T. A. (2005). Hindsight bias and outcome-consistent thoughts when observing and making service provider decisions. *Organizational Behavior and Human Decision Processes*, *98*, 88–95.
- Magnus, P. D. (2009). On trusting Wikipedia. *Episteme*, *6*, 74–90.
- Mark, M. M., & Mellor, S. (1991). Effect of self-relevance of an event on hindsight bias: The foreseeability of a layoff. *Journal of Applied Psychology*, *76*, 569–577.
- Massa, P., & Scrinzi, F. (2012). Manypedia: Comparing language points of view of Wikipedia communities. In *Proceedings of the 8th international symposium on wikis and open collaboration (WikiSym'12)*, Linz, Austria. Retrieved from [http://www.gnu.org/files/papers/manypedia\\_006\\_wikisym\\_2012\\_camera\\_ready.pdf](http://www.gnu.org/files/papers/manypedia_006_wikisym_2012_camera_ready.pdf). Accessed 10 Apr 2017.
- McNamara, D. S., & Kintsch, W. (1996). Learning from one text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.
- Merz, M., & Döring, N. (2010). Aktive Beteiligung an Wikipedia aus sozial-kognitiver Perspektive. [Active participation in Wikipedia from a social-cognitive perspective]. Retrieved from <http://purl.org/merz/20100926>. Accessed 27 Nov 2012.
- Musch, J. (2003). Personality differences in hindsight bias. *Memory*, *11*, 473–489.
- Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1399–1413.
- 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*, 1043–1054.
- 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*, 1539–1544.
- 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. *Psychologische Rundschau*.
- Oeberst, A., Halatchliyski, I., Kimmerle, J., & Cress, U. (2014a). Knowledge construction in Wikipedia: A systemic-constructivist analysis. *Journal of the Learning Sciences*, *23*, 149–176.
- Oeberst, A., von der Beck, I., & Nestler, S. (2014b). Reading about explanations enhances perceptions of inevitability and foreseeability. A cross-cultural study with Wikipedia articles. *Cognitive Processing*, *15*, 343–349.
- 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). Cham, Switzerland: Springer International Publishing.
- Olick, J. K. (1999). Collective memory: The two cultures. *Sociological Theory*, *7*, 333–348.
- Pennington, N., & Hastie, R. (1988). Explanation-based decision making: The effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 521–533.
- Pentzold, C. (2009). Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies*, *2*, 255–272.
- Pezzo, M. V. (2003). Surprise, defense, or making sense: What removes the hindsight bias? *Memory*, *11*, 421–441.
- Pohl, R. F. (2017). *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory*. London: Routledge.
- Pohl, R. F., Bender, M., & Lachmann, G. (2002). Hindsight bias around the world. *Experimental Psychology*, *49*, 270–282.
- 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). London: 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*, 49–58.
- Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality and Social Psychology*, *80*, 918–930.
- Potthast, M., Stein, B., & Gerling, R. (2008). Automatic Vandalism Detection in Wikipedia. In C. Macdonald et al (Eds.), *Advances in information retrieval. 30th European conference on IR research (ECIR 08), Lecture notes in computer science* (vol. 4956, pp. 663–668). Berlin: Springer.
- Powell, J. L. (1988). A test of the knew-it-all-along effect in the 1984 presidential statewide elections. *Journal of Applied Social Psychology*, *18*, 760–773.
- Rachlinski, J. J. (1998). A positive psychological theory of judging in hindsight. *University of Chicago Law Review*, *65*, 571–625.
- Robinson, M. D., Cassidy, D. M., Boyd, R. L., & Fetterman, A. K. (2015). The politics of time: conservatives differentially reference the past and liberals differentially reference the future. *Journal of Applied Social Psychology*, *45*, 391–399.
- Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: the lay assessment of subclinical depression from written self-descriptions. *Journal of Personality*, *78*, 575–597.
- Roese, N. J., & Olson, J. M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology*, *32*, 197–227.
- Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives in Psychological Science*, *7*, 411–426.

- Royal, C., & Kapila, D. (2009). What's on Wikipedia, and what's not...?: Assessing completeness of information. *Social Science Computer Review*, 27, 138–148.
- Schkade, D. A., & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes*, 49, 105–123.
- Schultheiss, O. C. (2013). Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis. *Frontiers in Psychology*, 4, 748.
- Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology*, 78, 655–669.
- Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, 88, 563–586.
- Smith, A. C., & Greene, E. (2005). Conduct and its consequences: Attempts at debiasing jury judgments. *Law and Human Behavior*, 29, 505–526.
- Stahlberg, D., Eller, F., Maass, A., & Frey, D. (1995). We knew it all along: Hindsight bias in groups. *Organizational Behavior and Human Decision Processes*, 63, 46–58.
- Sterba, S. K. (2017). Partially nested designs in psychotherapy trials: A review of modeling developments. *Psychotherapy Research*. doi:10.1080/10503307.2015.1114688.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54.
- Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, 11 (pp. 289–338). New York: Academic Press.
- The Express Tribune (2011). IAEA warned Japan over nuclear quake risk: Wikileaks. Published on March 17, 2011. Retrieved from <http://www.telegraph.co.uk/news/worldnews/wikileaks/8384059/Japan-earthquake-Japan-warned-over-nuclear-plants-WikiLeaks-cables-show.html>. Accessed 10 Apr 2017.
- The Telegraph (2011). Japan earthquake: Japan warned over nuclear plants, WikiLeaks cables show. Published on March 15, 2011. Retrieved from <http://tribune.com.pk/story/133824/iaea-warned-japan-over-nuclear-quake-risk/>. Accessed 10 Apr 2017.
- Verplanken, B., & Pieters, R. G. M. (1988). Individual differences in the reverse of hindsight bias: I never thought something like Chernobyl would happen. Did I? *Journal of Behavioral Decision Making*, 1, 131–147.
- Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on human factors in computing systems*. New York: ACM Press.
- Viégas, F. B., Wattenberg, M., Kriss, J., & van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. In *Proceedings of the 40th annual Hawaii international conference on system sciences (HICSS'07)*. Washington, DC: IEEE.
- Wilson, A. M., & Likens, G. E. (2015). Content volatility of scientific topics in Wikipedia: A cautionary tale. *PloS One*, 10(8), e0134454.
- Yama, H., Manktelow, K. I., Mercier, H., van der Henst, J. B., Soo Do, K., Kawasaki, Y., et al. (2010). A cross-cultural study of hindsight bias and conditional probabilistic reasoning. *Thinking & Reasoning*, 16, 346–371.
- Yopchik, J. E., & Kim, N. S. (2012). Hindsight bias and causal reasoning: A minimalist approach. *Cognitive Processing*, 13, 63–72.