AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 11, 3167–3188
https://doi.org/10.1037/xge0001452

Knowledge Updating in Real-World Estimation: Connecting Hindsight Bias and Seeding Effects

Julia Groß[1], Barbara K. Kreis[1], Hartmut Blank[2], and Thorsten Pachur[3, 4]
[1] Department of Psychology, University of Mannheim
[2] Department of Psychology, University of Portsmouth
[3] School of Management, Technical University of Munich
[4] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

When people estimate the quantities of objects (e.g., country populations), are then presented with the objects' actual quantities, and subsequently asked to remember their initial estimates, responses are often distorted towards the actual quantities. This *hindsight bias*—traditionally considered to reflect a cognitive error—has more recently been proposed to result from adaptive knowledge updating. But how to conceptualize such knowledge-updating processes and their potentially beneficial consequences? Here, we provide a framework that conceptualizes knowledge updating in the context of hindsight bias in real-world estimation by connecting it with research on *seeding effects*—improvements in people's estimation accuracy after exposure to numerical facts. This integrative perspective highlights a previously neglected facet of knowledge updating, namely, recalibration of metric domain knowledge, which can be expected to lead to transfer learning and thus improve estimation for objects from a domain more generally. We develop an experimental paradigm to investigate the association of hindsight bias with improved estimation accuracy. In Experiment 1, we demonstrate that the classical approach to induce hindsight bias indeed produces transfer learning. In Experiment 2, we provide evidence for the novel prediction that hindsight bias can be triggered via transfer learning; this establishes a direct link from knowledge updating to hindsight bias. Our work integrates two prominent but previously unconnected research programs on the effects of knowledge updating in real-world estimation and supports the notion that hindsight bias is driven by adaptive learning processes.

Public Significance Statement

When people try to recall their previous judgment on some issue (e.g., "What is the population of Sweden?") and had in the meantime learned the true value ("Sweden's population is about 10.4 million"), they seem to misremember their previous judgment to be closer to the true value than it had actually been. This *hindsight bias* has traditionally been interpreted as reflecting a deficiency of the mind. Here we demonstrate in the context of real-world quantitative estimation that hindsight bias likely represents a side effect of adaptive learning processes. Specifically, we show that hindsight bias emerges because people use the acquired true numerical values of objects to recalibrate the metric underlying their estimates of objects in a domain. Our results suggest that hindsight bias, rather than indicating a mental flaw, is the product of people's smart integration of new information into their world knowledge.

Keywords: hindsight bias, judgment, real-world estimation, seeding effects, transfer of knowledge

Supplemental materials: https://doi.org/10.1037/xge0001452.supp

This article was published Online First August 3, 2023.
Julia Groß https://orcid.org/0000-0002-1555-1070
This work was funded by Grant GR 4649/4-1 (PA 1925/2-1) from the German Research Foundation (DFG) and supported by the University of Mannheim's Graduate School of Economic and Social Sciences. The data and analysis code are available via the Open Science Framework (https://osf.io/va2jf/). The authors thank Ute J. Bayen for financial support (online pilot study) and discussions at earlier stages of the project, Michael Ohlinger for technical support and programming of the experiments, Jann Wäscher for data collection (Experiment 1), Andreas Brandmaier for discussions, Ulrich Hoffrage for comments on an earlier version of the manuscript, and Susannah Goss for editing the manuscript.

Julia Groß served as lead for data curation, formal analysis, project administration, resources, visualization, and writing–original draft. Barbara K. Kreis contributed equally to data curation, and served in a supporting role for formal analysis, investigation, methodology, and writing–review and editing. Hartmut Blank served in a supporting role for investigation, resources, and writing–review and editing. Thorsten Pachur served as lead for supervision, contributed equally to resources, and served in a supporting role for writing–original draft, formal analysis, and visualization. Julia Groß, Hartmut Blank, and Thorsten Pachur contributed equally to conceptualization and methodology. Julia Groß and Thorsten Pachur contributed equally to investigation, writing–review and editing, and funding acquisition. The authors have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Julia Groß, Department of Psychology, University of Mannheim, A5 – C, Room 202, D-68159 Mannheim, Germany. Email: julia.gross@uni-mannheim.de

Being able to accurately estimate quantities—the time it will take to finish a task, the distance from home to a target location, or the amount of sugar contained in a food item—helps individuals navigate their environment (e.g., Nieder, 2020). Although intuitive estimates of real-world quantities can be quite good (e.g., Griffiths & Tenenbaum, 2006), they are rarely perfectly accurate. When people who have given an estimate (e.g., "It takes 160 min to travel from Paris to Bordeaux by train") are then provided with the actual value (e.g., 130 min), two things can usually be observed.

First, when asked to recall their original estimates, people's responses tend to be closer to the actual value than their original estimates were. This well-established phenomenon is known as *hindsight bias* (e.g., Christensen-Szalanski & Willham, 1991; Groß & Pachur, 2019; Hawkins & Hastie, 1990), and was initially introduced by Fischhoff (1975). It has been shown for estimates of real-world quantities (e.g., lengths of rivers, heights of towers; see Bayen et al., 2006; Bernstein et al., 2011; Erdfelder & Buchner, 1998; Groß & Bayen, 2015; Pohl & Hell, 1996), but also for two-alternative-choice tasks, judgments of event outcomes (e.g., historical and medical), and other tasks and materials (see Pohl, 2007, for an overview). Traditionally, hindsight bias has been viewed as a manifestation of the limitations of human information processing and as an impediment to accurate judgment (e.g., Arkes, 1981). It is interpreted as indicating that people are inevitably anchored on information in the environment (e.g., numerical facts) and have difficulty inhibiting this information when trying to assess their initial state of knowledge (Hawkins & Hastie, 1990; Tversky & Kahneman, 1974). From this perspective, hindsight bias is a negative consequence of exposure to information (justifying the term "bias").

Second, after having been provided with the actual value of an object, people's subsequent estimates for both this object and other objects from the same domain (e.g., the time it takes to travel from Paris to Marseille by train) tend to be more accurate than before exposure to the information. This suggests that people are able to integrate new information into existing knowledge and thus to improve their judgment performance. For the context of real-world quantitative estimation, such as country populations, city-to-city distances, latitudes and longitudes, university tuition fees, or food calories, these improvements have become known as effects of *seeding the knowledge base* (or *seeding effects*; Brown, 2002; Brown & Siegler, 1993; Friedman & Brown, 2000; Groß et al., in press; Lawson & Bhagat, 2002; Wohldmann, 2015). Thus, exposure to numerical facts about objects in a domain also has a positive consequence: knowledge about the domain is updated, and subsequent judgments are more accurate.

Hindsight bias and seeding effects have been studied in separate research traditions. In this article, we connect the two lines of research to map out and examine a perspective on hindsight bias that deviates from the common view that it reflects a cognitive error; specifically, we consider the possibility that hindsight bias results from processes of adaptive knowledge updating (Hawkins & Hastie, 1990; Hoffrage et al., 2000). Despite initial evidence in support of this idea (Hoffrage et al., 2000; Nestler et al., 2012), the nature of the knowledge updating potentially underlying hindsight bias and its consequences on judgment performance are not yet well understood. To date, research on a possible link between hindsight bias and knowledge updating has mainly focused on local effects of learning—that is, the updating of cue knowledge for a specific set of objects. Less attention has been paid to the potentially more global downstream benefits of knowledge updating—for instance, that it can also facilitate more accurate judgments for other objects from the same domain.

We use concepts, measures, and insights from the literature on seeding effects to distinguish ways in which knowledge updating can occur in the context of real-world quantitative estimation. For instance, learning can occur both for the objects on which information has been provided (which we call *seed objects*), but also for other objects in the same domain (which we call *nonseed objects*) due to recalibration of people's metric representation of the domain. Extrapolating from the literature on seeding effects, we predict that hindsight bias should be accompanied by *transfer of knowledge*, that is, improvement for nonseed objects. This possibility has not yet been considered. Furthermore, our integrative perspective yields the novel prediction that hindsight bias for an object could be triggered by any manipulation that leads to transfer learning—even when no direct information about the object is provided. We test these predictions within an integrative experimental paradigm that combines elements from hindsight-bias and seeding experiments.

In the following, we first introduce the phenomenon of hindsight bias and the idea that it might reflect processes of knowledge updating. We then explain how research on seeding effects can be used to conceptualize, distinguish, and measure different effects of knowledge updating in the context of real-world estimation. Exploiting the parallels of the experimental paradigms used to study hindsight bias and seeding effects, we propose an integrative paradigm that allows us to examine the link between the two phenomena. In Experiment 1, we demonstrate that providing numerical facts—as done in the typical hindsight paradigm—produces both hindsight bias and transfer learning, consistent with the idea that hindsight bias reflects adaptive processes of knowledge updating. In Experiment 2, we demonstrate a direct link from transfer learning to hindsight bias by providing evidence for the novel prediction that hindsight bias for a set of objects can be triggered solely by providing relevant knowledge about the knowledge domain, not about the objects directly.

Hindsight Bias and Knowledge Updating

Hindsight bias is a multifaceted phenomenon (manifested in inevitability and foreseeability impressions as well as distorted memory; for overviews, see Blank et al., 2007, 2008; Roese & Vohs, 2012). In tasks that require participants to estimate real-world quantities such as historical dates, heights of buildings, lengths of rivers, or the population of cities (e.g., Bayen et al., 2006; Bernstein et al., 2011; Erdfelder & Buchner, 1998; Groß & Bayen, 2015), hindsight bias has often been investigated in the so-called memory paradigm (Pohl, 2007), where it manifests as distorted memory. In this paradigm, participants are first presented with questions whose quantitative answers they are unlikely to know exactly, but can estimate (e.g., "How long is the Amazon river [in km]?"). The responses (e.g., "4,000 km") are referred to as *original judgments* (OJ). In a subsequent phase, participants are told the actual values (6,992 km) for some objects (experimental items) but not for others (control items); they are then asked to report their initial estimates for all items, referred to as *recall of original judgments* (ROJ). The characteristic finding is that ROJs (e.g., "4,600 km") are closer to the objects' actual values (i.e., more accurate) than the OJs were. Typically, the shift is relatively pronounced for experimental items. Some shift can also occur for control items, but if so, it is typically much smaller.[1]

Why does an ROJ shift toward the actual value? One idea is that hindsight bias is due to the *anchoring-and-adjustment* heuristic (Tversky & Kahneman, 1974). If people cannot directly recall their OJ, they might use the actual value as a starting point (i.e., as an anchor) for reconstructing the OJ, but then adjust insufficiently from the anchor (e.g., Hawkins & Hastie, 1990; Pohl, 1998; Pohl et al., 2003; Wilson et al., 2021). Consequently, the ROJ is closer to the actual values than the OJ was. Hindsight bias might thus arise from the use of a judgment heuristic; from this perspective, the distortion that it generates can be viewed as a cognitive error.

Hoffrage et al. (2000) proposed an alternative view on hindsight bias, arguing that it could reflect adaptive knowledge updating. Specifically, someone given the actual value of an object might update the mental model of the knowledge domain that was used to construct the OJ. Following Gigerenzer et al. (1991), we refer to a mental model as the collection of knowledge that might be relevant to construct a judgment, including probabilistic cues, knowledge about the correlation of the cues and the criterion (i.e., cue weights), knowledge about the criterion values of objects in the domain, as well as more generic information about the domain, such as the distribution of criterion values (e.g., the central tendency, variability, or range). Let us illustrate the proposal made by Hoffrage et al. (2000) by assuming that someone is asked to estimate the population of the Italian city of Verona (i.e., the criterion). To come up with an estimate, the person considers cues that are predictive of the criterion (e.g., whether the city is home to a university, whether it has an airport, or whether it has a soccer team in the Serie A; cf. Lee et al., 2017) as well as more general domain knowledge (e.g., the typical population of Italian cities). After giving an estimate, the person learns the actual population of Verona (257,353 inhabitants) and updates their mental model (see Hoffrage et al., 2000)—for instance, by imputing missing cue values (e.g., adding that Verona does have a soccer team in the Serie A) or by modifying domain knowledge (e.g., typical population of Italian cities) and existing cue values (e.g., that Verona does have a university) to cohere with the new information. When the person is asked for an ROJ and they cannot directly recall their OJ, they repeat the initial judgment process, but now based on an updated mental model. The ROJ will then be closer to the actual value than the OJ had been, reflecting the updated knowledge. Hindsight bias could thus arise as a result of *updating and rejudgment*.[2]

Two empirical studies have tested the notion that hindsight bias might be due to knowledge updating. Hoffrage et al. (2000) had participants learn cue information for a set of food items (e.g., the amount of saturated fat or protein in a pie or a cake). In a paired-comparison task, the participants were then asked to choose which of two objects had a higher value on a criterion (e.g., "Which has more cholesterol, cake or pie?") and to indicate their confidence in that choice. In a subsequent phase of the experiment, they were told the criterion (i.e., cholesterol) values for each object and asked to recall for each paired comparison their original choice as well as their original confidence in that choice. In a final task, participants were asked to recall the cue information for the objects. The retrospectively reported choices indicated a hindsight bias, with the recalled choices corresponding more closely to the correct choice (i.e., the one implied by the objects' criterion values) than the original choices did. In addition, people's retrospectively reported confidence in their original choices was more aligned with the correct choice than their original confidence had been (i.e., confidence increased for originally correct choices and decreased for originally incorrect choices). Finally, when Hoffrage et al. (2000) derived choices for the paired comparisons based on the cue values that participants recalled after being presented with the criterion values, they found that these simulated choices were more often correct than participants' original choices had been—an indication that the participants had updated their cue knowledge consistent with the criterion values. Thus, the study by Hoffrage et al. (2000) provides evidence that hindsight bias and knowledge updating co-occurred.

In a second study testing the knowledge-updating account of hindsight bias, Nestler et al. (2012) showed participants pictures of (nonacquainted) target persons and asked them to rate the targets' personality on several dimensions (e.g., extraversion, conscientiousness, and openness to experience). Subsequently, participants received feedback, namely, information about how the target persons had rated themselves on the same dimensions. Finally, participants were asked to recall their initial ratings. These retrospective judgments showed a shift toward the feedback—that is, a hindsight effect. In addition, regression analyses (based on the lens model framework; Brunswik, 1956) suggested that participants relied on more valid cues in their retrospective judgments than in their initial judgments (e.g., trimness of hairstyle to predict conscientiousness; Study 1) and that cue knowledge was updated in line with the feedback, as indicated by participants' judgments of the cues before and after presentation of feedback (Study 2). These findings indicate that the feedback prompted the updating of relevant cue knowledge and cue weights (e.g., how strongly a person's hairstyle is predictive of their personality). The sizes of individual hindsight effects correlated with individual improvements in judgments and utilization of more valid cues. Thus, hindsight bias and knowledge updating also co-occurred in the study by Nestler et al. (2012).

The investigations by Hoffrage et al. (2000) and Nestler et al. (2012) showed that presenting the actual values of objects leads to changes in the (cue) knowledge underlying people's judgments. However, it remains unclear to what extent this updating of knowledge improves people's judgments more generally. In principle, it is conceivable that feedback about the actual value of an object leads to an update of (cue) knowledge for that specific object only; judgments for other objects from the judgment domain could be unaffected. Alternatively, the update could change the underlying knowledge representation more broadly, such that judgments for other objects from the domain also become more accurate. Such transfer of knowledge could be tested by asking people to provide

[1] Previous work has defined hindsight bias as the difference in shift between experimental items and control items (e.g., Pohl, 2007). Based on the assumption that shifts for control items are due to repeated testing or regression toward the mean of the OJ (Pohl, 2007), the difference should indeed provide a purer measure of bias. In the research presented here, we consider effects of knowledge updating as a possible explanation for systematic shifts in judgments. As we will outline, these effects may occur not only for experimental items but also for control items via transfer of knowledge. We, therefore, examine the shifts for experimental items and control items separately, and use the term "hindsight effect" for both types of items.

[2] Here, we focus on possible cognitive mechanisms for hindsight bias in real-world quantitative estimation; for other domains and tasks, such as retrospective judgments of event outcomes, motivational factors also play a role (e.g., Blank et al., 2008; Hawkins & Hastie, 1990). We return to a broader perspective on the hindsight-bias phenomenon in the "General Discussion" section.

estimates for other objects from the judgment domain. The notion that hindsight bias is due to adaptive knowledge updating implies that it should be possible to demonstrate beneficial downstream effects of the information that "biases" hindsight judgments.

To test this possibility and to better understand the nature and consequences of knowledge updating in real-world quantitative estimation—one of the most frequently studied contexts of hindsight bias (e.g., Bayen et al., 2006; Bernstein et al., 2011; Coolin et al., 2016; Erdfelder & Buchner, 1998; Groß & Bayen, 2015; Pohl et al., 2018, 2010; Pohl & Hell, 1996)—we connect the investigation of hindsight bias with research on seeding effects (Brown & Siegler, 1993).[3] Our integrative framework allows for a more differentiated conceptualization of the processes of knowledge updating and its downstream consequences. It links two previously unconnected lines of research on quantitative real-world estimation and makes it possible to measure hindsight bias and knowledge updating concurrently.

Knowledge Updating in Real-World Quantitative Estimation: Seeding Effects

Brown and Siegler (1993) proposed a framework for studying the processes and knowledge structures underlying people's estimation of real-world quantities and how intuitive estimates can be improved. When people are provided with representative numerical facts (*seed facts*) about objects in a knowledge domain (e.g., populations of countries), there is an improvement in their judgments of both those seed objects and other (nonseed) objects in the domain. In a typical seeding experiment, participants first provide estimates for a set of objects (Brown & Siegler, 2001; LaVoie et al., 2002). They are then provided with the numerical facts for either a subset of these objects or a different but representative set of objects. These seed facts are supposed to help the participants to recalibrate their judgments about objects of the domain. Finally, participants are asked to provide estimates either for seed objects, for which the actual value had been provided, or for nonseed objects, for which the actual value had not been provided. The (re)estimates for the seed objects allow for a measurement of *direct learning*, the estimates for the nonseed objects allow for a measurement of *transfer learning*. To illustrate, Brown and Siegler (1996) asked participants to estimate 99 country populations. Next, they were presented with the actual populations of a subsample of 24 countries (which served as seed facts), and then were asked to estimate again all 99 country populations. A robust finding in such seeding experiments is that participants' postseeding estimates are more accurate than their preseeding estimates, both for seed objects (reflecting direct learning) and for nonseed objects (reflecting transfer learning; Brown & Siegler, 1993, 1996; LaVoie et al., 2002). Providing people with seed facts thus seems to lead to an updating of the knowledge structure underlying people's estimates.

To better understand the seeding effects, Brown and Siegler (1993) distinguished between two aspects of numerical knowledge: metric knowledge and mapping knowledge. *Metric knowledge* refers to knowledge of the statistical properties of the distribution of objects in a domain (i.e., central tendency, variability, and shape) and allows people to give estimates within a plausible range. People can have metric knowledge (e.g., about the typical range or the average of country populations) even in the absence of knowledge about individual objects from the domain (e.g., the population of France being 67 million). Metric knowledge is usually measured in terms of the order of magnitude error (OME; e.g., Brown, 2002; Brown & Siegler, 1996, 2001):

$$\mathrm{OME}_i = \left|\log_{10}\left(\frac{\mathrm{estimate}_i}{\mathrm{actual}_i}\right)\right|. \tag{1}$$

The OME quantifies the (log-transformed) discrepancy between the actual and the estimated value of an object i and converts the discrepancy into an order of magnitude. The rationale for using the OME is as follows. Real-world distributions (e.g., country populations) are often highly skewed (e.g., Bak, 2013). Both the actual values and the estimates for such distributions can differ by orders of magnitude, and individual estimation errors can thus have an undue impact when measuring alignment of quantities with more conventional indices (e.g., Pearson correlation and mean deviation). A log-based measure such as the OME helps to minimize the distorting effects of outliers in such a distribution. By log-transforming the discrepancy, over- and underestimates of the same order of magnitude are weighted equally (Brown & Siegler, 2001).[4]

Mapping knowledge, by contrast, refers to information that allows people to determine the relative magnitudes of a given set of objects from the domain (e.g., that Poland has a smaller population than France, which in turn has a smaller population than Germany). This facet of knowledge is independent of knowledge about the absolute magnitudes of the objects (e.g., how much smaller the population of France is than that of Germany). The accuracy of mapping knowledge is usually measured as the rank-order correlation between estimated and actual values (e.g., Brown & Siegler, 1993, 1996).

Several studies have demonstrated that providing seed facts results in improvements in metric knowledge (e.g., Brown & Siegler, 1996; LaVoie et al., 2002). These improvements do not seem to be superficial or short-lived (e.g., due to anchoring processes), but instead reflect a genuine *recalibration* of people's metric knowledge: People seem to revise the metric assumptions underlying their responses (e.g., range, frequent exemplars, and central tendency). Brown and Siegler (1996) showed that people's postseeding estimates for nonseed objects remained improved for up to 4 months after seeding (possibly even longer), supporting the idea that seeding effects are driven by such recalibration. Anchoring effects can be ruled out because specific seed facts are likely forgotten after such a long time; anchoring would require that the seed facts are retrievable from memory (Brown & Siegler, 2001; Friedman & Brown, 2000). In another study, Brown and Siegler (2001) directly contrasted predictions from the anchoring and recalibration hypotheses, using a preselected set of objects for which the two processes would have opposite consequences for postseeding estimates (larger postseeding estimates would speak for anchoring; smaller postseeding estimates for recalibration). The authors found clear evidence for recalibration.

[3] In a workshop presentation, Brown and Lee (2005) also linked seeding effects and hindsight bias. To our knowledge, this work was not devised to test theories of hindsight bias or to locate the two phenomena within a broader framework of adaptive knowledge updating; instead, it demonstrated hindsight "side effects" of seeding.

[4] An OME of 1 means that the actual value has been misestimated by one order of magnitude (e.g., an estimate of 10 million or 100,000 when the actual value is 1 million); an OME of 2 means that the actual value has been misestimated by two orders of magnitude (e.g., an estimate of 100 million or 10,000 when the actual value is 1 million).

Whereas seeding affects metric knowledge both for seed and nonseed objects, mapping knowledge, in contrast, has been shown to improve primarily for seed objects (Brown & Siegler, 1996, 2001; LaVoie et al., 2002). In addition, improvements in mapping knowledge are rather short-lived and hinge on the presence or retrievability from memory of the specific seed facts (LaVoie et al., 2002).

Taken together, research on seeding effects offers a framework for conceptualizing processes of knowledge updating in real-world quantitative estimation. Providing seed facts—representative numerical facts for a given domain—leads to long-term adjustments in people's metric representation of the domain: The accuracy of subsequent judgments is improved not only for the seed objects, indicating direct learning, but also for nonseed objects, indicating transfer learning. It does not, however, lead to changes in mapping accuracy for nonseed objects. Research on seeding effects thus points to a way to measuring the different types of learning effects—direct learning and transfer learning—that could result from exposure to numerical facts in the context of hindsight bias. Next, we describe an integrative experimental paradigm that combines elements from hindsight-bias and seeding experiments, making it possible to test whether the occurrence of hindsight bias is associated with the occurrence of direct learning and transfer learning.

An Integrative Framework for Investigating the Link Between Hindsight Bias and Knowledge Updating in Real-World Estimation

As described above, it has been proposed that hindsight bias results from processes of adaptive knowledge updating (Hoffrage et al., 2000). It is currently unclear, however, whether hindsight bias in the estimation of real-world quantities reflects a local, object-specific effect (e.g., due to direct learning) or, in addition, a change in the overall metric representation, which should be manifested in effects of transfer learning. Based on insights from research on seeding effects, the actual value of an object provided prior to a hindsight judgment could function as a seed fact. If so, hindsight bias should co-occur with both direct learning and transfer learning. In other words, the information leading to hindsight bias would confer beneficial effects of learning and estimation improvement for both seed and nonseed objects. In the classical hindsight-bias paradigm, any potential benefits in terms of improved estimation accuracy for seed and nonseed objects remain invisible because changes in estimation accuracy are not measured. Furthermore, if hindsight bias results from processes of adaptive knowledge updating, a manipulation that produces transfer learning should produce hindsight bias in an object even if no specific information about that object is given.

To investigate these possibilities, we exploit the conceptual and procedural similarities between the hindsight-bias and seeding paradigms as well as their potential to complement each other. As shown in Figure 1A and B, participants in both paradigms first perform an estimation task (OJ task); they are then provided with actual values (seed facts) for some or all of those objects. In the hindsight-bias paradigm (Figure 1A) participants are asked to report their original answers (ROJ task), whereas in the seeding paradigm (Figure 1B) they are given another estimation task; this task involves either the same objects (OJ_{again} task), allowing for a measurement of direct learning for seed objects and a measurement of transfer learning for nonseed objects, or it involves new objects (OJ_{new} task), allowing for a measurement of transfer learning.

The Integrated Hindsight-Bias-and-Seeding Paradigm

To study hindsight bias and effects of seeding simultaneously and quantify both their sizes, we propose the *integrated hindsight-bias-and-seeding* (iHBS) paradigm. As shown in Figure 1C, participants in the iHBS paradigm first perform an estimation task (OJ task). They are then provided with the actual values (seed facts) for some or all of the previously estimated items, and are asked to report their original judgments (ROJ task). Finally, they perform a new estimation task, either for the same items (OJ_{again}) or for new items (OJ_{new}).

For the hindsight-bias paradigm, it has been shown that providing actual values leads to biased ROJs. Biased reconstruction can occur for both seed objects (experimental items), and—due to transfer of knowledge—nonseed objects (control items); we, therefore, use the term "hindsight effect" for both types of items (see also Footnote 1). Mapped onto the iHBS paradigm, there is a hindsight effect if the estimation error (i.e., OME) is smaller for the ROJ than for the OJ.[5]

For the seeding paradigm, it has been shown that providing seed facts leads to changes in metric knowledge. Mapped onto the iHBS paradigm, there is knowledge updating in the form of direct learning if the OME is smaller for the OJ_{again} than for the OJ for seed objects. There is knowledge updating in the form of transfer learning if the OME is smaller for the OJ_{again} than for the OJ for nonseed objects, and if it is smaller for the OJ_{new} than for the OJ.

Research Questions and Hypotheses

Our experiments addressed two main sets of research questions and hypotheses.

Does Hindsight Bias Co-Occur With Transfer Learning?

First, presenting an object's actual value to participants before the ROJ, which typically triggers hindsight bias, should prompt an update of knowledge for this (seed) object, observable as an improvement in OJ_{again} relative to OJ—that is, a direct-learning effect (Hypothesis 1a). Second and importantly, if hindsight bias also reflects a recalibration of metric knowledge (i.e., more general knowledge updating), presenting an object's actual value should also lead to a transfer-learning effect (Hypothesis 1b). Therefore, we expect an improvement of judgments of new objects from the same domain (all items in the OJ_{new} task) relative to the judgments in the OJ task. In addition, we expect an improvements of judgments of previously encountered objects for which actual values had not been provided (i.e., the control items in the OJ_{again} and ROJ tasks) relative to the judgments in the OJ task.

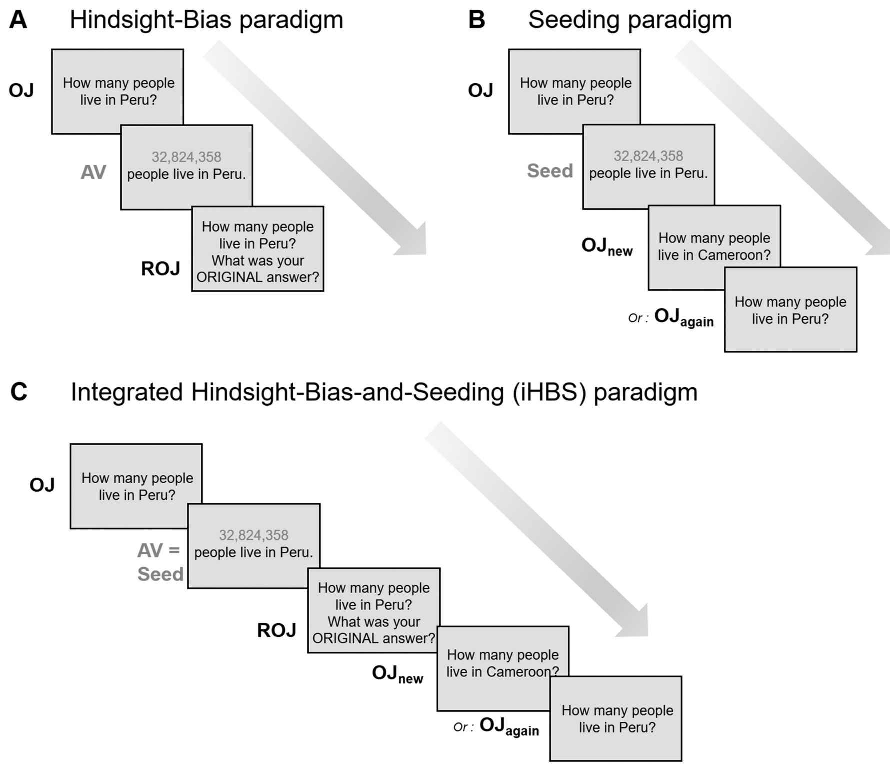
Can Hindsight Bias Can Be Triggered via Transfer of Knowledge?

The demonstration of a co-occurrence of hindsight bias and transfer learning would support the knowledge-updating account of hindsight

[5] A prominent way to quantify hindsight was proposed by Rüdiger Pohl (e.g., Pohl, 2007; Pohl & Hell, 1996), according to which hindsight bias is quantified as $\Delta z_i = |z(\text{estimate}_i) - z(\text{actual}_i)|$. This approach is based on z-transformed values of the estimated and actual quantities in order to be able to compare hindsight bias for items from domains that are located on different scales (e.g., heights of buildings in meters vs. lengths of rivers in kilometers). As we use a single domain (with all items on the same scale) with a highly skewed distribution, the OME is a better suited measure for our present investigation.

Figure 1

The Elements of (A) the Hindsight-Bias Paradigm, (B) the Seeding Paradigm, and (C) the Integrated Hindsight-Bias-and-Seeding Paradigm



Note. OJ = original judgment; AV = actual value; ROJ = recall of original judgment; Seed = seed fact; OJ_{again} = repeated original judgment; OJ_{new} = new original judgment.

bias, but it would not necessarily imply that knowledge updating is a direct cause of hindsight bias. In principle, both phenomena could occur independently of each other, even if they are both produced by the presentation of the actual value. If, however, hindsight bias is in fact due to processes of knowledge updating in the form of a recalibration of underlying metric knowledge, a novel prediction of our framework is that it should be possible to trigger a hindsight effect merely by providing relevant domain knowledge (e.g., actual values for other objects from the domain; Hypothesis 2a). By the same token, providing numerical information that cannot be used to update domain knowledge (e.g., because it refers to a different domain) should *not* trigger a hindsight bias (Hypothesis 2b). Conversely, if such information also leads to hindsight bias, this would suggest that the changes in people's responses in the ROJ task are due to anchoring processes rather than due to knowledge updating.

The Present Research

We conducted two experiments using the iHBS paradigm to test the hypotheses and research questions described in the previous section. In Experiment 1, we tested the co-occurrence of hindsight bias and the two types of learning (Hypotheses 1a and 1b). In Experiment 2, we tested the novel prediction that hindsight bias can be triggered via transfer of knowledge by providing relevant domain knowledge rather than direct knowledge about seed objects; we also tested for the possible operation of anchoring processes (Hypotheses 2a and 2b).

Both experiments involved the same knowledge domain (country populations). We developed seven sets of items—four used in Experiment 1 and three used in Experiment 2—from data collected in a separate pilot study conducted online; in this study, 100 participants estimated the populations of 96 countries. Across sets, the mean population was comparable, as were people's metric knowledge and mapping knowledge, and the distribution of countries from different continents was similar. Details can be found in Appendix Tables A1, A2, and A3. The studies were not preregistered.

Experiment 1: Does Hindsight Bias Co-Occur With Transfer Learning?

Experiment 1 aimed at testing whether presenting the actual values of objects leads simultaneously to hindsight bias and transfer learning. We used a full version of the iHBS paradigm that (a) included a within-participants manipulation of feedback (i.e., experimental items, control items) and (b) allowed for a measurement of hindsight effects, direct learning, and transfer learning. As a straightforward consequence of presenting the actual values for objects, we predicted, first, a hindsight effect when participants were asked for an ROJ; this hindsight effect should be particularly pronounced for experimental items, but it might also emerge to some extent for control items due to transfer of knowledge. Second, we expected that presenting the actual values for objects would lead to an updating for these objects, observable as direct-learning effects when participants were asked to provide estimates for the same objects again (i.e., experimental items in the OJ_{again} task; Hypothesis 1a); in addition and importantly, updating should also lead to improved estimation accuracy when participants were asked to provide estimates for new objects (OJ_{new}) or for previously encountered objects for which no actual values had been provided (control items in the OJ_{again} task), signaling transfer-learning effects (Hypothesis 1b) (note that both control items in the ROJ and OJ_{again} tasks as well as all items in the OJ_{new} task are nonseed objects).

Previous research on hindsight bias has employed different setups to present actual values. To reflect this, we considered two presentation formats: first, the actual values were presented *during* the ROJ task ("concurrent feedback"; e.g., Bayen et al., 2006; Erdfelder & Buchner, 1998; Groß & Bayen, 2015); second, they were presented *prior* to the ROJ task ("preceding feedback"; e.g., Erdfelder & Buchner, 1998). Based on the previous findings, we expected to observe hindsight effects for both presentation formats. In addition, we expected that hindsight effects might be more pronounced with concurrent feedback than with preceding feedback, where the actual values are accessible only by memory retrieval (Erdfelder & Buchner, 1998, Experiment 2). If so, our experiment allows us to explore whether corresponding differences would also emerge for learning effects (i.e., direct and transfer learning).

Method

Participants

A total of $N = 322$ participants took part in a laboratory experiment at the Max Planck Institute for Human Development (Center for Adaptive Rationality). Of these, 81 were assigned to an experimental condition whose manipulation did not lead to learning effects. For transparency, we describe this manipulation and the results in Appendix B. Here, we focus on the 241 participants assigned to the other conditions. Previous studies (Brown & Siegler, 1996, 2001; LaVoie et al., 2002) showed reliable seeding effects on OME with approximately 30 participants per condition. As our analytic approach (see below) involves the estimation of various by-participant random effects, we approximately doubled our planned participant sample size. We verified the ability of our experimental setup to reveal both hindsight and seeding effects in a pilot study (not reported in the article). Post-hoc power analyses showed that for the smallest observed effects in Experiments 1 and 2 power was 0.83 or higher. All participants were native speakers of German aged 18–45 years ($M = 26.1$ years, $SD = 4.7$, 146 women, 94 men, 1 other), and were recruited via the participant database maintained by the Center for Adaptive Rationality. Participants took about 45–50 min to complete all tasks and received a fixed fee of 12 € as compensation. The study protocol was approved by the Ethics Committee of the Max Planck Institute for Human Development.

Materials

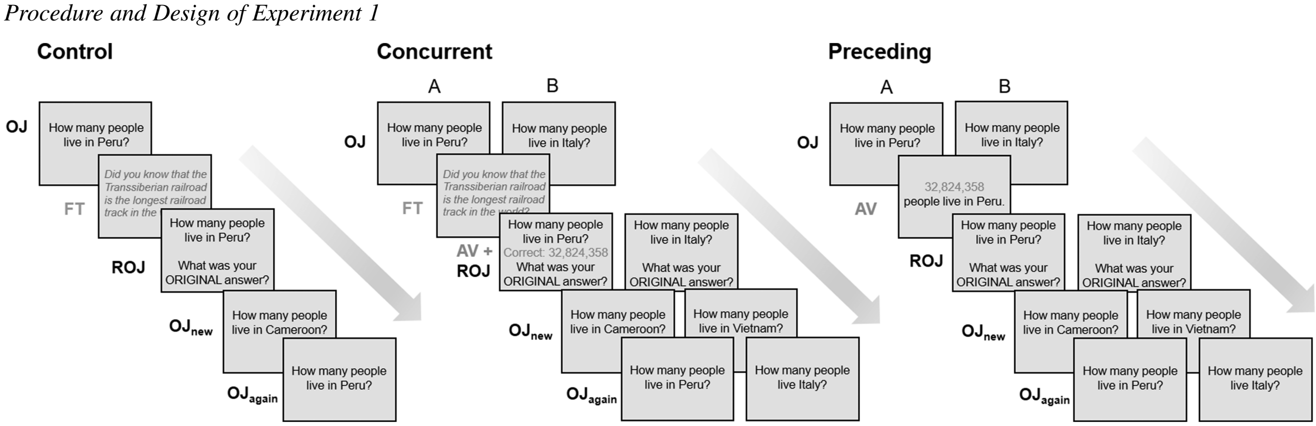
We created four sets of items, each consisting of 23 countries, as listed in Appendix Table A1. Two of the four sets were presented in the OJ, ROJ, and OJ_{again} tasks—one as the experimental items (i.e., for which actual values were provided) and the other as the control items. The remaining two sets were presented in the OJ_{new} task. Assignment of sets to tasks was counterbalanced, such that each combination appeared with comparable frequency across participants and experimental conditions.

Procedure and Design

The experimental procedure is shown in Figure 2. Participants were randomly assigned to be presented with the actual values of objects either prior to the ROJ task (*preceding-feedback* condition), during the ROJ task (*concurrent-feedback* condition), or not at all (*no-feedback control* condition).

In the first phase (OJ task), all participants estimated the populations of 46 countries. Items were presented sequentially in an order that was randomized by participant. The median time a participant took to estimate a country population was 8.3 s (range: 4.2–17.4 s). In the second phase, participants in the concurrent-feedback and control conditions read a short text that was unrelated to the topic of the study (filler task), whereas participants in the preceding-feedback condition were presented with the actual values for half of the previously estimated items. That is, 23 countries served as

Figure 2
Procedure and Design of Experiment 1



Note. Control = no-feedback control condition; Concurrent = concurrent-feedback condition; Preceding = preceding-feedback condition; OJ = original judgment; FT = filler task; ROJ = recall of original judgment; OJ_{new} = new original judgment; OJ_{again} = repeated original judgment; AV = actual value. For the concurrent-feedback and preceding-feedback conditions, Streams A and B show examples for experimental and control items, respectively.

experimental items (right panel, Figure 2, Stream A), the remaining 23 countries were control items (Stream B). The actual values were presented for 4 s each, followed by a blank screen for 500 ms. In the third phase (ROJ task), all participants were asked to recall their 46 initial OJs in the same randomized order as in the OJ task. Additionally, participants in the concurrent-feedback condition were presented with the actual values for half of the previously estimated items. That is, 23 countries served as experimental and control items, respectively (Figure 2, middle panel). In the fourth phase (OJ_{new} task), all participants estimated the populations for two new sets of items (i.e., 46 countries). In the fifth and final phase (OJ_{again} task), all participants again estimated all 46 country populations from the OJ task (for half of which participants in the experimental conditions had been presented with the actual values) in the same randomized order. The design was thus a four (phase: OJ, ROJ, OJ_{new}, OJ_{again}) by three (feedback condition: control, concurrent, preceding) mixed design.

Data Diagnostics

We performed several checks to identify data points unlikely to reflect compliance with task instructions. Some participants entered numbers below 1,000. These extremely implausible estimates of country populations may reflect typing errors, lapses in attention, or failures to comply with task instructions. In the first step, we therefore excluded all data from participants for whom we could not ensure sufficient compliance with the task instructions, namely, participants for whom more than 15% of responses were smaller than 1,000 in at least two of the four experimental phases (28 participants; 11.6%). In the second step, we excluded remaining single responses smaller than 1,000, which we assumed to be due to typing errors or lapses in attention (1.77% of all data points). We further excluded all responses equal to or larger than 8,000,000,000, which is roughly the world population (0.34% of all data points). Finally, we excluded five participants (one in the concurrent-feedback and four in the preceding-feedback conditions) whose median estimation accuracy (in terms of OME) fell outside of three times the interquartile range in at least one of the four experimental phases. This left a total of 208 participants in the analyses ($n = 71$ control, $n = 70$ concurrent, $n = 67$ preceding). We checked the robustness of the results with regard to participant exclusions by also performing the analyses leaving the 28 participants with >25% responses <1,000 in the analysis (but excluding all single responses <1,000). This led to the same result pattern. The data and analysis code are available at https://osf.io/va2jf/.

Analytic Approach

For each of the tasks—OJ, ROJ, OJ_{new}, and OJ_{again}—we calculated the deviation of the responses from the actual values in terms of the OME (Equation 1). To assess how the OMEs for ROJ, OJ_{new}, and OJ_{again} responses compared to those of the OJ responses, we applied a Bayesian linear mixed-effects model with treatment contrast coding and the OJ phase as the reference category.[6] We included phase, feedback condition, and their interaction as fixed effects. The estimated regression weights can be used to assess how the OME of the responses in the ROJ, OJ_{new}, or OJ_{again} tasks in the different experimental conditions differ from those in the OJ task in the control condition (which serves as a reference). We further estimated random intercepts for participants (capturing participant variability in OME in the OJ phase), random intercepts for items (capturing item variability in OME in the OJ phase), and random slopes for participants (capturing variability across participants in the effect of phase on OME). The random slopes for items showed negligible variability and were thus not included in any of the models. For parameter estimation, we used the `brms` package (Bürkner, 2017, 2018), which calls `STAN` for MCMC sampling (Stan Development Team, 2019). Prior specification is described in Appendix C. Hypothesis tests were conducted by comparing a model that included the fixed-effect predictor of interest (full model, M_1) to a model that did not (baseline model, M_0). The baseline models included all random effects that were specified in the full model; that is, in the baseline model it was assumed that the random effects of the respective predictor vary across participants and items but are 0 on average. For hypothesis testing, we used the `bayes_factor` function in `brms`, which computes Bayes factors (BFs) based on bridge sampling. The BF quantifies the evidence for the alternative hypothesis relative to the evidence for the null hypothesis (BF_{10}), or vice versa (BF_{01}), by comparing M_1 to M_0 (or vice versa).[7]

Treatment of Cases in Which the ROJ Equals the OJ

Hindsight bias is zero for cases in which people's ROJs perfectly reproduce the respective OJs (which likely reflects successful recollection of the OJ). The overall size of hindsight bias is thus smaller the higher the percentage of such cases. If this percentage differs between different conditions, an observed difference in hindsight bias can be due to a genuine difference in hindsight bias, different percentages of cases in which the ROJ equals the OJ, or both. Therefore, whenever this percentage differs across conditions, it is important to separate cases in which the ROJ equals the OJ from cases where the ROJ deviates from the OJ in the analysis (e.g., Erdfelder & Buchner, 1998; Groß & Pachur, 2019; Pohl, 2007). In our experiment, however, we observed a comparable percentage of OJ = ROJ cases across experimental conditions (see Appendix Table D1 for details). The analyses we report in the main text therefore include all of the data (i.e., OJ = ROJ and OJ $\neq$ ROJ). In the online supplemental materials, we also report the results for the OJ $\neq$ ROJ cases. The two analytic approaches led to the same conclusions.

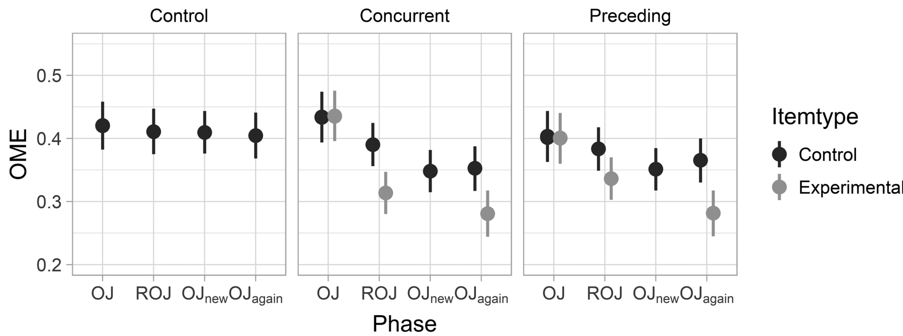
Results

Participants' accuracy (in terms of OME) in the three conditions is shown in Figure 3. We first tested whether the presentation of the actual values resulted in a hindsight effect. To recap, a hindsight effect is indicated by the OME being smaller for the ROJ than for the OJ. To

[6] Although the OMEs were better described by a log-normal distribution, we used a Gaussian distribution as the likelihood function because the parameters estimated from this model can be directly interpreted as mean OME (intercept) and changes in mean OME (slope). For both the log-normal and Gaussian likelihood functions, the Markov chain Monte Carlo (MCMC) diagnostics (e.g., traceplots and number of effective samples) and results were comparable.

[7] A common interpretation scheme is that a BF_{10} below 1/10 indicates strong, a BF_{10} between 1/10 and 1/3 indicates moderate, and a BF_{10} between 1/3 and 1 indicates weak evidence for M_0, whereas a BF_{10} larger than 10 indicates strong, a BF_{10} between 3 and 10 indicates moderate, and a BF_{10} between 1 and 3 indicates weak evidence for M_1 (e.g., Jeffreys, 1961; Lee & Wagenmakers, 2013; van Doorn et al., 2023).

Figure 3
Judgment Accuracy in Experiment 1



Note. The conditional predictions based on the mixed-effects model (estimated means and 95% credible intervals). OME = order of magnitude error; OJ = original judgment; ROJ = recall of original judgment; OJ_{new} = new original judgment; OJ_{again} = repeated original judgment; Control = no-feedback control condition; Preceding = preceding-feedback condition; Concurrent = concurrent-feedback condition.

test for a hindsight effect, we compared a model including the fixed-effect predictor phase (OJ vs. ROJ) to a baseline model without that predictor. As expected, there was strong evidence for a hindsight effect in the two experimental conditions (concurrent-feedback: $BF_{10} > 10{,}000$, preceding-feedback: $BF_{10} = 33.9$), but not in the control condition ($BF_{10} = 0.01$). To test whether the size of the hindsight effects differed between experimental and control items in the experimental conditions, we compared a model that included the Phase × Itemtype interaction to a baseline model with phase as the only predictor. In the concurrent-feedback condition, the hindsight effect was larger for experimental than for control items ($BF_{10} > 10{,}000$). The evidence for a hindsight effect was strong for experimental items ($BF_{10} > 10{,}000$) and moderate for control items ($BF_{10} = 3.6$). In the preceding-feedback condition, there was weak evidence that the hindsight effect was larger for experimental than for control items ($BF_{10} = 1.9$). There was strong evidence for a hindsight effect in experimental items ($BF_{10} > 10{,}000$) and against a hindsight effect in control items ($BF_{10} = 0.04$).

In the second step, we examined to what extent there were direct-learning effects, the most immediate learning consequence of exposure to the actual values (Hypothesis 1a). To recap, a direct-learning effect is indicated by the OME being smaller for the OJ_{again} than for the OJ in experimental items (i.e., seed objects). As expected, there was strong evidence for direct learning in both the concurrent-feedback and the preceding-feedback condition (both $BF_{10} > 10{,}000$), but no evidence for direct learning in the control condition ($BF_{10} = 0.04$). The presentation of the actual values thus resulted in more accurate judgments when the same items were estimated again.

Does the Presentation of Actual Values Before the ROJ Lead to Transfer Learning?

In the third step, we tested whether exposure to the actual values led to transfer-learning effects (Hypothesis 1b). To recap, a transfer-learning effect is indicated by the OME being smaller for the OJ_{new} than for the OJ. We found strong evidence for transfer learning in the concurrent-feedback condition ($BF_{10} > 10{,}000$) and moderate evidence for transfer learning in the preceding-feedback condition ($BF_{10} = 7.6$). The size of transfer-learning effects did not differ between the two experimental conditions ($BF_{10} < 0.01$). By contrast, there was no transfer-learning effect in the control condition ($BF_{10} = 0.01$).

A transfer-learning effect is also indicated by the OME being smaller for the OJ_{again} than for the OJ for control items (for which no actual values were provided). There was strong evidence for transfer learning in the OJ_{again} task in the concurrent-feedback ($BF_{10} > 10{,}000$) but only weak evidence for transfer learning in the preceding-feedback condition ($BF_{10} = 2.1$).

In addition to testing for changes in metric accuracy—measured in terms of the OME—we tested whether the presentation of actual values also led to changes in mapping accuracy between OJ and ROJ, OJ and OJ_{new}, as well as OJ and OJ_{again}, measured as the rank-order correlation between estimated and actual values. To recap, mapping knowledge should only improve for items for which actual values were presented (i.e., ROJ and OJ_{again} for experimental items); no improvements are expected for control items. Our results were in line with these expectations. Details are presented in Appendix Tables E1 and E2.

Discussion

It has been proposed that hindsight bias might be the by-product of an adaptive process of knowledge updating in response to learning about facts. Connecting this proposal to the literature on seeding effects, we predicted that the actual value presented before the ROJ functions as a seed fact and should therefore lead to a transfer of knowledge. Experiment 1 provided support for this prediction: we observed improvement of judgments not only for objects whose actual values were presented (i.e., OJ_{again} for experimental items) but also for objects whose actual values were not presented (i.e., OJ_{again} for control items and OJ_{new}). This pattern of results suggests that the presentation of numerical facts led to a recalibration of the underlying domain knowledge. In line with the idea that this type of knowledge updating contributes to hindsight bias, there was also a (small) hindsight effect for control items (however, only for concurrent feedback and not for preceding feedback). Presenting numerical facts thus causes both distortion in the retrospective assessment of one's previous judgments and improvement in judgment accuracy more generally.

Hindsight and transfer-learning effects emerged irrespective of whether the actual values had been presented concurrently with or preceding the ROJ. The co-occurrence of the two phenomena thus seems to be robust across different ways in which hindsight bias had been induced in previous research. Both hindsight and transfer learning were more pronounced with concurrent feedback than with preceding feedback. This commonality further bolsters the idea that hindsight bias and learning are closely linked. A possible explanation for the effects being stronger with concurrent than with preceding feedback is that the former facilitates the linking of the actual value to existing knowledge because its presence during the rejudgment process promotes the integration of prior knowledge with new information (Erdfelder & Buchner, 1998, Experiment 2).

Experiment 1 showed that hindsight effects are associated with transfer learning. However, this does not necessarily imply that hindsight bias is a direct consequence of knowledge updating. It is conceivable that both hindsight and transfer-learning effects occur independently as a consequence of the presentation of the actual values. In Experiment 2, we tested for a direct link between hindsight effects and transfer learning.

Experiment 2: Can Hindsight Bias Be Triggered by Transfer of Knowledge?

In Experiment 2, we took a different approach to testing the proposal that hindsight bias reflects knowledge updating. Recall that we proposed that the relevant process of knowledge updating consists of a recalibration of metric knowledge, brought about by the presentation of actual values of the judgment domain. For such recalibration, it is irrelevant whether the actual values refer to the objects for which an OJ was given, or whether they refer to other objects from the same domain. Therefore, if hindsight bias reflects processes of knowledge updating, it should be possible to trigger a hindsight effect for a set of objects merely by providing the actual values for a set of other objects from the domain ("domain information"). In other words, hindsight bias should be triggered via transfer of knowledge (Hypothesis 2a).

Furthermore, if hindsight bias reflects processes of knowledge updating, there should be no hindsight effect if the values provided cannot be used to update knowledge of the judgment domain (Hypothesis 2b). To test this prediction, participants were exposed to the actual values for previously estimated objects but were told that the numbers represented actual values for objects from a different domain ("irrelevant knowledge"). This type of information should *not* lead to an updating of metric knowledge of the judgment domain; however, it might influence judgments (and produce a hindsight effect) via anchoring processes, if anchoring contributes to hindsight bias in the current context.

Method

Participants

The experiment was conducted online. We recruited a total of $N = 295$ participants via Prolific (https://www.prolific.co/). All participants were native speakers of German aged 18–45 years ($M = 27.8$ years, $SD = 6.5$, 116 women, 178 men, 1 other). They took about 30–40 min to complete the experiment and received a fixed fee of £5.60. The study protocol was approved by the Ethics Committee of the University of Mannheim.

Materials

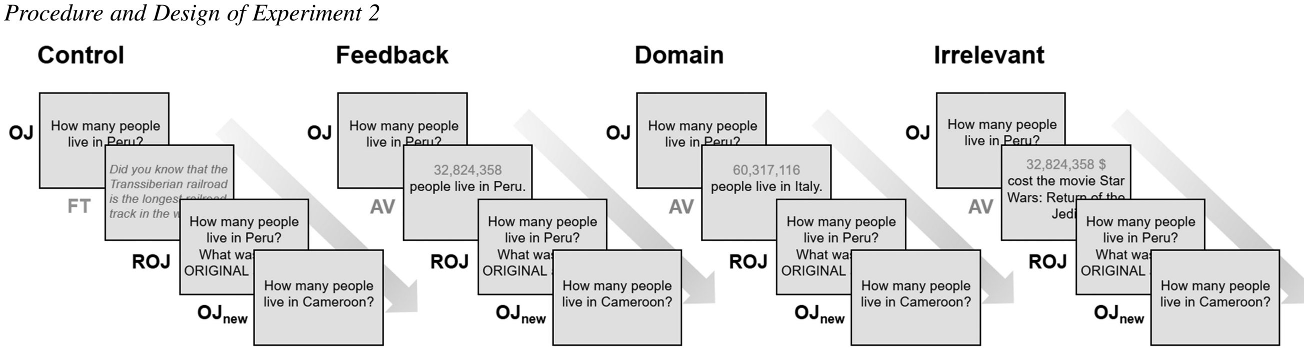
We created three sets of items, each consisting of 32 countries (see Appendix Table A2 for details). Assignment of sets to tasks was counterbalanced, such that each set was presented in each task (OJ/ROJ, (Actual) Values, OJ_{new}) with comparable frequency across participants and experimental conditions.

Procedure and Design

The experimental procedure is shown in Figure 4. Participants were randomly assigned to one of four experimental conditions as described below. The experiment consisted of four phases. In the first phase (OJ task), all participants estimated the population of a set of 32 countries. Items were presented in an order randomized by participant. The median time a participant took to estimate a country population was 9.4 s (range: 3.7–32.8 s).

In the second phase, participants were provided with information depending on the experimental condition. In the *feedback* condition, they were presented with actual values for the same set of 32 countries that they had estimated in the OJ task (declared as such; this procedure is analogous to that for the preceding-feedback condition in Experiment 1). In the *domain-information* condition, participants

Figure 4
Procedure and Design of Experiment 2



Note. Control = no-feedback control condition; Feedback = feedback condition; Domain = domain-information condition; Irrelevant = irrelevant-information condition; OJ = original judgment; FT = filler task; ROJ = recall of original judgment; OJ_{new} = new original judgment; AV = actual value.

were presented with the actual values for a different set of 32 countries. In the *irrelevant-information* condition, they were presented with the same numerical information as in the feedback condition (i.e., actual values for the previously estimated objects), but the numbers were labeled as budgets (in U.S.$) for 32 popular movies (e.g., "Star Wars: Return of the Jedi"). We chose this domain because the budgets approximately matched the mean and range of country populations across the three sets (see Appendix Table A2 for details). In all three conditions, each number was shown for 5 s. In the *control* condition, participants instead read a short text that was unrelated to the topic of the study (filler task).

In the third phase of the experiment (ROJ task), participants were asked to recall their 32 initial OJs in the same randomized order as in the OJ task. In the fourth and final phase (OJ_{new} task), participants estimated the populations for a new set of 32 countries. The design was thus a three (phase: OJ, ROJ, OJ_{new}) by four (condition: control, feedback, domain, irrelevant) mixed design. Following the experimental procedure, all participants completed a number-mapping task (results are reported in a separate manuscript).

Data Diagnostics and Analytic Approach

In the first step, we excluded the data of participants who reported technical problems (one participant), insufficient compliance with task instructions (one participant), or having been disturbed during participation (one participant). We then checked for data points that were unlikely to reflect compliance with task instructions (contrary to Experiment 1, in Experiment 2 participants were not allowed to enter numbers smaller than 1,000 as responses). As in Experiment 1, we excluded all responses equal to or larger than 8,000,000,000 (0.09% of all data points). Finally, we excluded an additional seven participants (three in the control condition, three in the irrelevant-information condition, and one in the feedback condition) whose median estimation accuracy (in terms of OME) fell outside of three times the interquartile range in at least one of the four experimental phases. This left a total of 285 participants for the analyses ($n = 69$ control, $n = 74$ feedback, $n = 72$ domain information, $n = 70$ irrelevant information). The data and analysis code are available at https://osf.io/va2jf/. We applied the same analytic approach as in Experiment 1.

Results

Participants' accuracy (in terms of OME) in the four conditions is shown in Figure 5. In a first step, we examined to what extent there were transfer-learning effects. Replicating Experiment 1, there was evidence for transfer learning in the feedback condition ($BF_{10} > 10{,}000$). There was also evidence for transfer learning in the domain-information condition ($BF_{10} > 10{,}000$); the size of the effect did not differ from that in the feedback condition ($BF_{10} < 0.01$). There was evidence for no transfer learning both in the control condition ($BF_{10} = 0.01$) and, importantly, in the irrelevant-information condition ($BF_{10} = 0.07$). The latter result indicates that the transfer-learning effects were not due to anchoring on the actual values, but likely reflect genuine knowledge updating.[8]

Can Hindsight Effects Be Triggered via Transfer of Knowledge?

After establishing that the presentation of actual values led to transfer learning, we next tested whether transfer learning produced hindsight effects. Most crucial for this question is the domain-information condition, in which no actual values for previously estimated objects were presented; as a consequence, hindsight effects could arise only via transfer learning (Hypothesis 2a). We indeed found evidence of a hindsight effect in the domain-information condition ($BF_{10} = 88$), and the size of the effect did not differ from that in the feedback condition ($BF_{10} < 0.01$). These results demonstrate, to our knowledge for the first time, that hindsight bias for objects can arise not only when feedback is given on these objects (as in the standard hindsight-bias paradigm), but also when information is provided on other objects from the same domain.

Importantly, there were no hindsight effects in the irrelevant-information condition ($BF_{10} < 0.01$), where the same values were provided as in the feedback condition, but labeled as referring to a different domain (Hypothesis 2b). This shows that hindsight bias is not triggered by the mere presentation of the values (e.g., through anchoring processes). There was no evidence of hindsight effects in the control condition either ($BF_{10} = 0.01$); the irrelevant-information condition did not differ from the control condition in this respect ($BF_{10} < 0.01$).

As in Experiment 1, we also tested for changes in mapping accuracy between OJ and ROJ as well as between OJ and OJ_{new}. As expected, mapping accuracy improved only for items for which the actual values were presented (i.e., for ROJ in the feedback condition). Details can be found in Appendix Tables E3 and E4.

Discussion

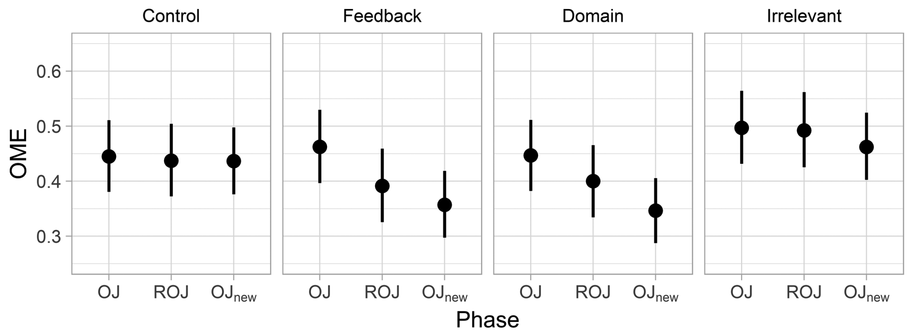
Replicating Experiment 1, Experiment 2 showed that presenting actual values for objects causes hindsight effects for those objects as well as transfer-learning effects for other objects. Transfer learning occurred both when the actual values of previously estimated objects were presented ("feedback") and when the actual values of other objects from the same domain were presented ("domain information"). Experiment 2 thus showed that presenting information about any objects in a domain can improve judgment accuracy in that domain (see also LaVoie et al., 2002).

Furthermore, we established for the first time that hindsight effects can be triggered not only by providing actual values for previously estimated objects, but also by providing actual values for other objects in the domain. This could mean, for instance, that when a person estimates the population of Peru, then learns about the population of Italy, and is then asked to recall their prior estimate of Peru, they might, due to the update of underlying knowledge, report as their prior estimate a value that is more accurate than their actual prior estimate, even without having learned anything about Peru directly. Going beyond Experiment 1, Experiment 2 thus points to a direct link between transfer learning and hindsight bias: Whereas in Experiment 1 the two could have co-occurred independently, this possibility can be ruled out in the domain-information condition in Experiment 2, where no actual values for previously estimated objects were presented.

[8] This conclusion on possible anchoring effects is based on the assumption that the potential anchors—the full set of values presented as movie budgets—would recalibrate participants' responses and thereby lead to a general improvement in estimation accuracy. In the online supplemental materials, we present analyses based on alternative assumptions about which specific values might have served as anchors. For example, participants could have anchored on what they perceived as a central tendency of the presented information, or on the central tendency of the most recent pieces of information presented. Even with these alternative assumptions, there was no evidence that participants' responses were affected by exposure to the values.

Figure 5

Judgment Accuracy in Experiment 2



Note. The conditional predictions based on the mixed-effects model (estimated means and 95% credible intervals). OME = order of magnitude error; OJ = original judgment; ROJ = recall of original judgment; OJ_{new} = new original judgment; Control = no-feedback control condition; Feedback = feedback condition; Domain = domain-information condition; Irrelevant = irrelevant-information condition.

Additionally, Experiment 2 provided evidence that the hindsight and learning effects observed are unlikely to be due to anchoring processes: When the same numbers that produced transfer learning (when presented as the actual values for the country populations) were presented as being numbers of an unrelated domain (movie budgets), neither transfer-learning nor hindsight effects emerged. This, in turn, suggests that participants in the feedback and domain-information conditions integrated the domain-relevant information into their knowledge structures. While some studies have shown that irrelevant numbers can induce anchoring effects (Englich et al., 2006; Reitsma-van Rooijen, 2006), we did not find them to trigger hindsight bias in our study.

General Discussion

Hindsight bias has traditionally been regarded as a cognitive illusion arising from deficits of the human mind. More recently, it has been proposed that hindsight bias might instead reflect the operation of an adaptive process: knowledge updating (Hoffrage et al., 2000). There has been little conceptual work, however, on the nature and potentially beneficial downstream effects of such knowledge-updating processes. By connecting research on hindsight bias with research on seeding effects (e.g., Brown & Siegler, 1993), we developed a framework that allows the effects of two different types of knowledge updating to be differentiated and quantified, namely, the local updating of information for specific items—as manifested in direct learning—and global updating in the form of recalibration of metric domain knowledge—as manifested in transfer learning. Our framework thus highlights a previously neglected possible benefit of the "biasing" exposure to numerical facts, namely, a general improvement of estimation accuracy in the domain. In our experiments, we separated transfer learning from direct learning and tested the link between hindsight bias and transfer learning.

Experiment 1 demonstrated that the presentation of numerical facts leads simultaneously to hindsight bias and transfer learning. Experiment 2 provided evidence for the novel prediction that also the provision of domain information can trigger hindsight bias.[9] Transfer-learning and hindsight effects were not observed when the same numeric information was presented but labeled as referring to a different domain. This suggests that the effects were due to genuine knowledge updating rather than to simple anchoring effects. Overall, our findings provide support for the notion that hindsight bias results (in part) from knowledge updating. In the following, we discuss various implications of our framework and findings.

Theoretical Implications

In their research on the knowledge updating account of hindsight bias, Hoffrage et al. (2000) and Nestler et al. (2012) showed how hindsight bias can result from an adaptive adjustment of cue values and cue weights, respectively, in people's mental model of a domain. Our research supports and extends this research in several ways. First, we elaborated the consequences of knowledge updating in a further facet of the mental model of a domain, namely, the recalibration of metric knowledge. Second, whereas Hoffrage et al. (2000) and Nestler et al. (2012) documented changes in elements of the mental model, we focused on the downstream consequences of knowledge updating—that is, the effects on people's estimation ability in a given domain.[10]

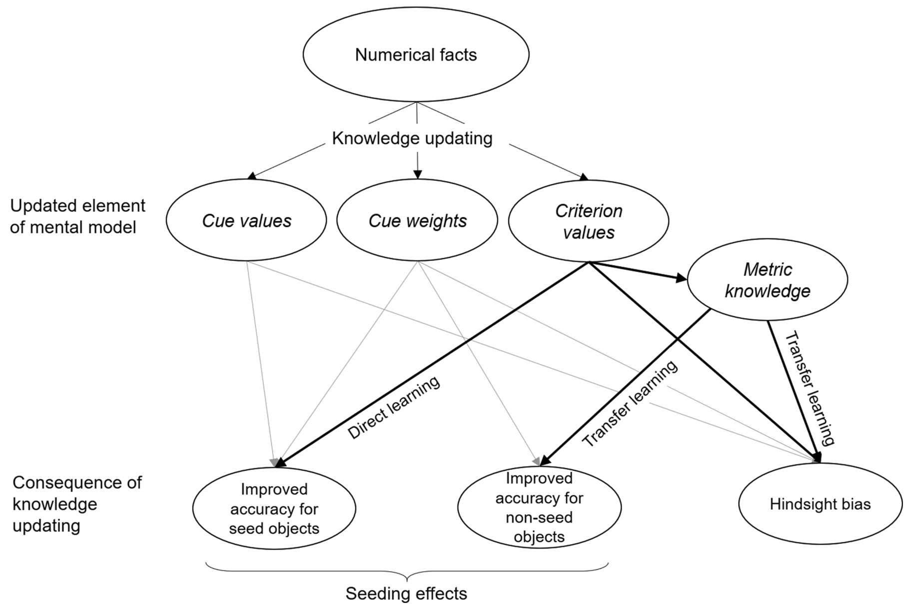
In Figure 6, we summarize and organize the insights from Hoffrage et al. (2000), Nestler et al. (2012), and the present research on different types of knowledge updating. The figure distinguishes three elements of the mental model that can be updated after learning information about objects in a domain: cue values, cue weights, and criterion values. Specifically, knowledge about the cue values of the objects can be updated, missing cue values can be imputed, cues might even be added, cue weights can be adjusted, and the criterion values of individual objects can be updated. Work on seeding effects by Brown and Siegler (1993) indicates that providing information about criterion values also leads to an updating of the metric

[9] This finding closely matches that of Brown and Lee (2005), who showed that being provided with actual values and applying the updated metric after seeding (by answering new items) is necessary to establish a transfer of knowledge and a hindsight "side effect."

[10] Hoffrage et al. (2000) showed by means of computer simulations that judgment accuracy *could* improve with updated cue values; they did not test, however, whether this potential was actually realized in people's judgments.

Figure 6

Forms of Knowledge Updating and Their Consequences for Subsequent Estimation Performance (Seeding Effects) and Retrospective Estimation Assessment (Hindsight Bias) in Real-World Estimation



Note. Bold arrows indicate the processes targeted in our experiments; grey arrows indicate additional possibilities for processes leading to knowledge updating and hindsight bias.

knowledge about a domain (consisting of knowledge about the central tendency, range, and distribution of objects in the domain).

In addition to providing a differentiated conceptualization of knowledge updating, Figure 6 specifies the various consequences that can follow from updating cue values, cue weights, and criterion values. While the updating of all three elements can lead to improved estimation accuracy for seed objects (i.e., objects for which actual values were provided), the updating of cue weights and metric knowledge can also lead to improved estimation accuracy for nonseed objects in the domain. In addition to these benefits of knowledge updating on estimation accuracy (i.e., seeding effects), updating can also have negative consequences in the form of hindsight bias. Hindsight bias can result from the updating of cue values, cue weights, and criterion values (as discussed by Hoffrage et al., 2000; Nestler et al., 2012, and in this article). Our work considered and found support for an additional possibility, namely, that hindsight bias can occur via the updating of metric knowledge—that is, via transfer learning (rightmost arrow in Figure 6). Furthermore, if general processes of knowledge updating lead to hindsight bias, it might be useful to consider the role of updating processes in other phenomena that reflect distorted judgment due to the presentation of information—such as the misinformation effect or the illusory truth effect (Hasher et al., 1977; Loftus et al., 1978).

Our work points to intriguing issues for future research on quantitative estimation and knowledge updating. For instance, what are the learning mechanisms by which numerical facts are integrated into a mental model, leading to metric recalibration? And how can the observed benefits of updated metric knowledge (Brown & Siegler, 1993, 2001) be accommodated within theoretical models of estimation such as exemplar-based and rule-based strategies (Juslin et al., 2003; Juslin & Persson, 2002; Trippas & Pachur, 2019; von Helversen & Rieskamp, 2008, 2009)? In general, it is important to formally model the mechanisms underlying real-world estimation (e.g., Brown, 2002; Schweickart & Brown, 2014).

Hindsight Bias and Knowledge Updating Beyond Real-World Estimation

In this article, we focused on hindsight bias in the context of real-world quantitative estimation, where metric knowledge plays an important role for judgment performance and can be updated as a result of exposure to numerical facts. However, hindsight bias has also been observed in domains and tasks where metric knowledge plays no central role, for instance, when judging event probabilities (e.g., Fischhoff, 1975), or identifying visual stimuli (e.g., Giroux et al., 2022). Does this mean that knowledge updating can be expected to be less relevant in these other contexts?

Consider the study by Goethals and Reckman (1973) on the consistency of attitudes. Participants were asked to indicate their level of agreement with a then controversial political measure of desegregation. They were then presented with arguments that ran counter to their opinion before being asked to recall as accurately as possible their initial level of agreement with the measure. These latter assessments indicated hindsight bias, such that participants' reconstruction of their initial level of agreement was biased toward the newly presented arguments. This was the case irrespective of whether participants were initially for or against the measure. Participants' distorted hindsight assessments can be interpreted as resulting from an updating of arguments in the mental model that formed the basis of their attitude (e.g., integrating new arguments in the mental model or modifying the weights given to arguments). That is, the notion that knowledge in a mental model is updated and contributes to hindsight bias might also apply beyond real-world quantitative estimation.

Furthermore, knowledge updating might also contribute to other manifestations of hindsight bias. In the research presented in the present article, we considered the primary approach used to study hindsight bias: the memory paradigm. Here, participants are asked to recall or reconstruct their previous judgment. But hindsight bias can also become manifest in other ways (Blank et al., 2008, see also footnote 2). First, after learning about the outcome of an event (e.g., a football match), people may indicate that they "knew it all along" or were able to predict the outcome, as indicated by increased *foreseeability impressions* (relative to a preoutcome rating or a no-outcome control condition). Second, after learning the outcome of an event, people may indicate that it was "bound to happen," as indicated by increased probability ratings or *inevitability impressions* (relative to a preoutcome rating or no-outcome control condition). In Nestler et al. (2010), people showed increased impressions of inevitability after being told the outcomes of applied science problems (e.g., a television set will continue to work when immersed in salad oil); however, this occurred only when the outcome feedback was accompanied by an explanation of the underlying factors, enabling causal learning (e.g., salad oil does not conduct electricity; see also Nestler et al., 2008). One way to interpret this result is that hindsight bias emerged because the explanation allowed participants to update their causal mental model of the domain.

Although updating of knowledge in a mental model may thus contribute to hindsight bias more broadly, there may also be boundary conditions for the link between hindsight bias and knowledge updating. We examined real-world estimation in a domain where people's knowledge is relatively limited, but where there is an actual (true) value at the time of judgment. Both aspects are typical of many real-world domains (e.g., knowledge of geography or food calories; Brown & Siegler, 1993; Friedman & Brown, 2000; Wohldmann, 2015). However, in some domains people's knowledge is more accurate, and in some domains there is irreducible uncertainty. For instance, in forecasting tasks (e.g., predicting stock prices or temperatures), performance is to a large extent limited by how predictable an outcome is, and even decision makers having rich knowledge of a domain might show only limited predictive accuracy. In such situations, the relative contribution of the processes of knowledge updating summarized in Figure 6 (e.g., updating cue values and metric knowledge) to hindsight bias may be more limited; instead, other factors may be more important—for example, the need for self-enhancement, coping, or other motivational influences (e.g., Biais & Weber, 2009; Blank et al., 2008; Groß et al., 2017; Renner, 2003).

Note that if the relative contribution of knowledge updating (and other processes) to hindsight bias depends on the amount of prior knowledge of a domain, this might have interesting implications for the study of developmental differences in hindsight bias. Compared with young adults, children and older adults are more prone to hindsight bias (e.g., Bernstein et al., 2011; Groß & Pachur, 2019; Pohl et al., 2018). Given that children and older adults differ in the amount of knowledge they have (e.g., Horn et al., 2016; Pachur et al., 2009), knowledge updating might play different roles in the two age groups: Knowledge-updating processes might drive hindsight bias in children, who tend not to have much prior domain knowledge, whereas other factors (e.g., age-related difficulties with the inhibition of new information; Coolin et al., 2016; Groß & Bayen, 2015) might drive hindsight bias in older adults, who tend to have much richer domain knowledge. The effect of prior knowledge on hindsight bias has also been discussed by Hertwig et al. (2003) and Christensen-Szalanski and Willham (1991).

On the Adaptivity of Hindsight Bias

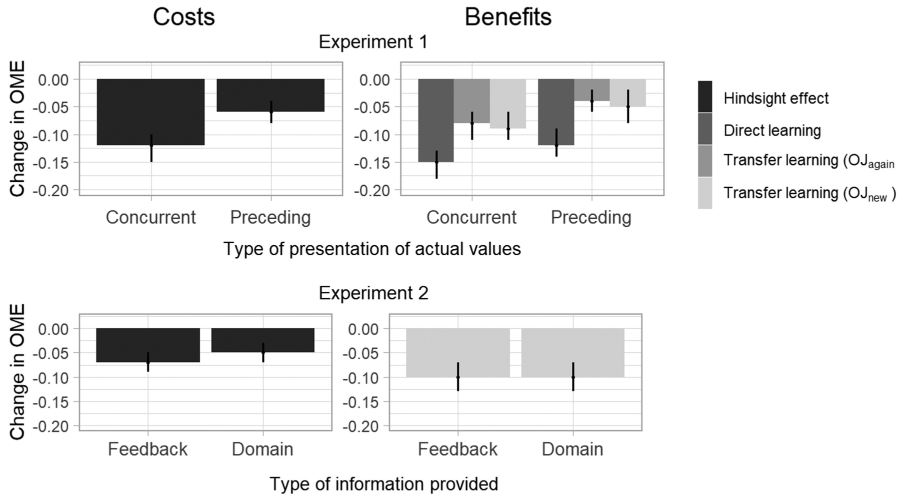
In their proposal that hindsight bias might result from knowledge updating, Hoffrage et al. (2000) suggested that "the disadvantage of hindsight bias is a relatively cheap price to pay for making better inferences" (p. 579, see also Hawkins & Hastie, 1990, p. 323). But how large is the benefit, relative to the price? If hindsight bias is a by-product of knowledge updating, it might be argued that the costs should not exceed the benefits, neither for direct learning nor transfer learning. To evaluate this claim, it would be necessary to gauge the costs of exposure to feedback (i.e., bias in hindsight judgments) against the benefits (i.e., improvement in estimation accuracy). The iHBS paradigm enables a measurement of both hindsight bias and effects of learning within the same judgment domain, and to compare the size of the effects directly by comparing the respective regression weights. Figure 7 shows the changes in OME resulting from the presentation of numerical facts in both Experiments 1 and 2. As can be seen, the beneficial effects in terms of improved estimation ability (direct learning and transfer learning) seem to be of a similar size as the bias in hindsight judgment. Statistical analyses indicated that overall, the benefits were of similar size as the costs, or even outweighed the costs.[11]

It should be noted that comparing the sizes of hindsight and learning effects shown in Figure 7 implies that both effects are of equal importance. However, for the domain of country populations, reconstructing a prior judgment is often much less important than being able to provide accurate responses in future judgments. In such situations, hindsight bias might indeed be a "cheap price to pay." Yet there might well be situations where accurately reconstructing a prior judgment is key. Consider, for example, a medical malpractice lawsuit. A jury has to evaluate whether a radiologist should have been able to detect a lung tumor on a chest X-ray at time t_1 (example taken from Berlin, 2000). The tumor became clearly visible only in a second X-ray at t_2—knowledge that the radiologist did not possess at t_1, but is now available to the jury and to expert witnesses. In order to make a fair verdict, the jury and experts have to reconstruct, in hindsight, the radiologist's state of knowledge at t_1. If the jury and experts overestimate this original state of knowledge due to hindsight bias, the costs can be high, as they might assign undue responsibility or unjust awards (Berlin, 2000)—even if they acquire from the case new knowledge that might be helpful in the future (for reviews on hindsight bias in the legal domain, see Giroux et al., 2016; Harley, 2007).

[11] To statistically evaluate the relative size of the effects, we compared a model that included as fixed effect the three-level predictor phase (OJ vs. ROJ vs. OJ_{again}/OJ_{new}) to a model that included a modified two-level predictor phase that assumed hindsight and learning effects to be of equal size (OJ vs. ROJ = OJ_{again}/OJ_{new}). In Experiment 1 (upper panel), the average direct-learning effect exceeded the average hindsight effect in the preceding-feedback condition ($BF_{10} = 125.7$), in the concurrent-feedback condition the size of the two effects did not differ ($BF_{10} = 0.2$). The sizes of the transfer-learning effects (for both previously estimated and new items) and hindsight effects did not differ in either experimental condition ($BF_{10} < 1.0$). In Experiment 2 (lower panel), the transfer-learning effect was larger than the hindsight effect in the domain-information condition (with weak evidence, $BF_{10} = 2.5$), but there was no evidence for a difference between the effects in the feedback condition ($BF_{10} = 0.9$).

Figure 7

Costs and Benefits of Being Presented With Numerical Facts (i.e., the Actual Values of Objects) in Experiments 1 and 2



Note. Estimated mixed-effects model regression weights (i.e., change in OME compared to the OJ phase) along with their 95% credible intervals. Costs = hindsight effect; Benefits = direct learning and transfer learning; Concurrent = concurrent-feedback condition; Preceding = preceding-feedback condition; Feedback = feedback condition; Domain = domain-information condition. The procedure of the feedback condition of Experiment 2 is analogue to that of the preceding-feedback condition of Experiment 1.

Furthermore, the relative size of hindsight bias and improved accuracy might also change dynamically across time. In our experiments, we compared hindsight bias and learning effects at a single point in time. In future research it would be interesting to gauge their relative sizes after, say, a day, a week, or a month. While research on seeding effects has shown that metric knowledge updating is stable over time (Brown & Siegler, 1996), research on hindsight bias tends to find an increase with longer delays between original judgment and outcome information (e.g., Blank et al., 2003; Hell et al., 1988). To comprehensively evaluate the balance between benefits and costs on more extended time horizons, it thus appears to be essential to consider not only the relative size of the effects but also their relative importance, as the accuracy of one's previous judgment should indeed matter less and less over the course of time in most applied context (except, e.g., in legal contexts).

Theory Integration

Our research underlines the merits of connecting research traditions that are usually pursued in separation. Here, we highlight the commonalities between paradigms that investigate hindsight bias and seeding effects in real-world estimation (for theory integration between hindsight bias in confidence judgments and the reiteration effect, see Hertwig et al., 1997). Our integrative perspective fosters theoretical enrichment in several ways. Specifically, it helps to refine the understanding of hindsight bias and to conceptually differentiate processes of knowledge updating. Rather than being driven (primarily) by anchoring processes, hindsight bias in real-world quantitative estimation seems to be due to rejudgment based on recalibrated metric knowledge, indicating that it is the by-product of an adaptive process (e.g., Hawkins & Hastie, 1990; Hoffrage et al., 2000; Nestler et al., 2012). Connecting hindsight bias and seeding effects highlights that the two may share a common cognitive basis and brings to mind the double-edged nature of some psychological processes. For example, high performance in some aspects of cognition (e.g., memory) might be associated with low performance in another aspect (e.g., abstraction and generalization; Bjork, 2011; Hills & Hertwig, 2011). In the present context, the same process that improves estimation—learning—distorts retrospection. Finally, and even more importantly, our theory integration led to a novel prediction: that hindsight bias can follow from transfer learning.

Constraints on Generality

Our results should generalize to adults aged 18–45. Similar results can be expected for older adults, children, and adolescents; however, the relative contribution of knowledge updating to hindsight effects might differ in these age groups (see the "General Discussion" section). We obtained our findings with a single knowledge domain, country populations; however, we expect our findings to generalize to other real-world domains for which people have limited metric knowledge (e.g., distances, lengths, areas, calories, sugar content, etc.). We obtained our findings in a single experimental session and with an explicit instruction to memorize the actual values presented. Based on previous research (e.g., Brown & Siegler, 1996; Erdfelder & Buchner, 1998; Groß & Bayen, 2015; Hell et al., 1988), we expect hindsight and learning effects to occur with longer retention intervals (hours, days, and weeks), and with incidental learning of the actual values. We have no reason to believe that the results depend on other characteristics of the participants, materials, and context.

Conclusion

Historically, it has been argued that hindsight bias—reflecting a lack of awareness of one's limited prior knowledge—restricts people's ability to learn (Fischhoff, 1975). Our research shows that, on the contrary, people learn and adaptively adjust their mental model of a domain when presented with factual information, even if their assessment of their prior knowledge is distorted. Hindsight bias is the result of mental changes that confer considerable benefits for cognitive functioning in terms of the ability to accurately estimate quantities in the environment. Biased retrospection and beneficial learning thus co-occur, suggesting that they are two sides of the same coin.

References

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology*, *49*(3), 323–330. https://doi.org/10.1037/0022-006X.49.3.323

Bak, P. (2013). *How nature works: The science of self-organized criticality*. Springer.

Bayen, U. J., Erdfelder, E., Bearden, J. N., & Lozito, J. P. (2006). The interplay of memory and judgment processes in effects of aging on hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1003–1018. https://doi.org/10.1037/0278-7393.32.5.1003

Berlin, L. (2000). Malpractice issues in radiology: Hindsight bias. *American Journal of Roentgenology*, *175*(3), 597–601. https://doi.org/10.2214/ajr.175.3.1750597

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Peria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 378–391. https://doi.org/10.1037/a0021971

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

Bjork, R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting* (pp. 19–40). Psychology Press.

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., Musch, J., & Pohl, R. F. (2007). Hindsight bias: On being wise after the event. *Social Cognition*, *25*(1), 1–9. https://doi.org/10.1521/soco.2007.25.1.1

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

Brown, N. R. (2002). Real-world estimation: Estimation modes and seeding effects. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 321–359). Academic Press.

Brown, N. R., & Lee, P. J. (2005, July). *Transfer of hindsight bias and its representational implications* [Paper presentation], Third Hindsight Bias Workshop, University of Leipzig, Germany.

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review*, *100*(3), 511–534. https://doi.org/10.1037/0033-295X.100.3.511

Brown, N. R., & Siegler, R. S. (1996). Long-term benefits of seeding the knowledge base. *Psychonomic Bulletin & Review*, *3*(3), 385–388. https://doi.org/10.3758/BF03210766

Brown, N. R., & Siegler, R. S. (2001). Seeds aren't anchors. *Memory & Cognition*, *29*(3), 405–412. https://doi.org/10.3758/BF03196391

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

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

Coolin, A., Erdfelder, E., Bernstein, D. M., Thornton, A. E., & Thornton, W. L. (2016). Inhibitory control underlies individual differences in older adults' hindsight bias. *Psychology and Aging*, *31*(3), 224–238. https://doi.org/10.1037/pag0000088

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality & Social Psychology Bulletin*, *32*(2), 188–200. https://doi.org/10.1177/0146167205282152

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 387–414. https://doi.org/10.1037/0278-7393.24.2.387

Fischhoff, B. (1975). Hindsight ≠ 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

Friedman, A., & Brown, N. R. (2000). Reasoning about geography. *Journal of Experimental Psychology: General*, *129*(2), 193–219. https://doi.org/10.1037/0096-3445.129.2.193

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

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

Giroux, M. E., Hunsche, M. C., Erdfelder, E., Kumar, R., & Bernstein, D. M. (2022). Hindsight bias for emotional faces. *Emotion*, *23*(1), 261–277. https://doi.org/10.1037/emo0001068

Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology*, *9*(6), 491–501. https://doi.org/10.1016/0022-1031(73)90030-9

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

Groß, J., & Bayen, U. J. (2015). Adult age differences in hindsight bias: The role of recall ability. *Psychology and Aging*, *30*(2), 253–258. https://doi.org/10.1037/pag0000017

Groß, J., Blank, H., & Bayen, U. J. (2017). Hindsight bias in depression. *Clinical Psychological Science*, *5*(5), 771–788. https://doi.org/10.1177/2167702617712262

Groß, J., Loose, A. M., & Kreis, B. K. (in press). A simple intervention can improve estimates of sugar content. *Journal of Applied Research in Memory and Cognition*.

Groß, J., & Pachur, T. (2019). Age differences in hindsight bias: A meta-analysis. *Psychology and Aging*, *34*(2), 294–310. https://doi.org/10.1037/pag0000329

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

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

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

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, *16*(6), 533–538. https://doi.org/10.3758/BF03197054

Hertwig, R., Fanselow, C., & Hoffrage, U. (2003). Hindsight bias: How knowledge and heuristics affect our reconstruction of the past. *Memory*, *11*(4–5), 357–377. https://doi.org/10.1080/09658210244000595

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, *104*(1), 194–202. https://doi.org/10.1037/0033-295X.104.1.194

Hills, T., & Hertwig, R. (2011). Why aren't we smarter already: Evolutionary trade-offs and cognitive enhancements. *Current Directions in Psychological Science*, *20*(6), 373–377. https://doi.org/10.1177/0963721411418300

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

Horn, S. S., Ruggeri, A., & Pachur, T. (2016). The development of adaptive decision making: Recognition-based inference in children and adolescents. *Developmental Psychology*, *52*(9), 1470–1485. https://doi.org/10.1037/dev0000181

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156. https://doi.org/10.1037/0096-3445.132.1.133

Juslin, P., & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): A "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, *26*(5), 563–607. https://doi.org/10.1207/s15516709cog2605_2

LaVoie, N. N., Bourne, Jr., L. E., & Healy, A. F. (2002). Memory seeding: Representations underlying quantitative estimations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1137–1153. https://doi.org/10.1037//0278-7393.28.6.1137

Lawson, R., & Bhagat, P. S. (2002). The role of price knowledge in consumer product knowledge structures. *Psychology & Marketing*, *19*(6), 551–568. https://doi.org/10.1002/(ISSN)1520-6793

Lee, M. D., Blanco, G., & Bo, N. (2017). Testing take-the-best in new and changing environments. *Behavior Research Methods*, *49*(4), 1420–1431. https://doi.org/10.3758/s13428-016-0798-x

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(1), 19–31. https://doi.org/10.1037/0278-7393.4.1.19

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 and causal attribution. *Social Psychology*, *39*(3), 182–188. https://doi.org/10.1027/1864-9335.39.3.182

Nestler, S., Egloff, B., Küfner, 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

Nieder, A. (2020). The adaptive value of numerical competence. *Trends in Ecology & Evolution*, *35*(7), 605–617. https://doi.org/10.1016/j.tree.2020.02.009

Pachur, T., Mata, R., & Schooler, L. J. (2009). Cognitive aging and the adaptive use of recognition in decision making. *Psychology and Aging*, *24*(4), 901–915. https://doi.org/10.1037/a0017211

Pohl, R. F. (1998). The effects of feedback source and plausibility of hindsight bias. *European Journal of Cognitive Psychology*, *10*(2), 191–213. https://doi.org/10.1080/713752272

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition*, *25*(1), 14–31. https://doi.org/10.1521/soco.2007.25.1.14

Pohl, R. F., Bayen, U. J., Arnold, N. R., Auer, T. -S., & Martin, C. (2018). Age differences in processes underlying hindsight bias: A lifespan study. *Journal of Cognition and Development*, *19*(3), 278–300. https://doi.org/10.1080/15248372.2018.1476356

Pohl, R. F., Bayen, U. J., & Martin, C. (2010). A multiprocess account of hindsight bias in children. *Developmental Psychology*, *46*(5), 1268–82. https://doi.org/10.1037/a0020209

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory*, *11*(4–5), 337–356. https://doi.org/10.1080/09658210244000487

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

Reitsma-van Rooijen, M. (2006). Subliminal anchoring: The effects of subliminally presented numbers on probability estimates. *Journal of Experimental Social Psychology*, *42*(3), 380–387. https://doi.org/10.1016/j.jesp.2005.05.001

Renner, B. (2003). Hindsight bias after receiving self-relevant health risk information: A motivational perspective. *Memory*, *11*(4–5), 455–472. https://doi.org/10.1080/09658210244000531

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*(5), 411–426. https://doi.org/10.1177/1745691612454303

Schweickart, O., & Brown, N. R. (2014). Magnitude comparison extended: How lack of knowledge informs comparative judgments under uncertainty. *Journal of Experimental Psychology: General*, *143*(1), 273–294. https://doi.org/10.1037/a0031451

Stan Development Team. (2019). *Stan modeling language: Users guide and reference manual*. https://mc-stan.org.

Trippas, D., & Pachur, T. (2019). Nothing compares: Unraveling learning task effects in judgment and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2239–2266. https://doi.org/10.1037/xlm0000696

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

van Doorn, J., Aust, F., Haaf, J. M., Stefan, A. M., & Wagenmakers, E. -J. (2023). Bayes factors for mixed models. *Computational Brain & Behavior*, *6*, 1–13. https://doi.org/10.1007/s42113-021-00113-2

von Helversen, B., & Rieskamp, J. (2008). The mapping model: A cognitive theory of quantitative estimation. *Journal of Experimental Psychology: General*, *137*(1), 73–96. https://doi.org/10.1037/0096-3445.137.1.73

von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: Rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 867–889. https://doi.org/10.1037/a0015501

Wilson, S., Arora, S., Zhang, Q., & Griffiths, T. L. (2021). A rational account of anchor effects in hindsight bias. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43, 1327–1332. https://escholarship.org/uc/item/5g91n4b5.

Wohldmann, E. L. (2015). Planting a seed: Applications of cognitive principles for improving food choices. *American Journal of Psychology*, *128*(2), 209–218. https://doi.org/10.5406/amerjpsyc.128.2.0209

(*Appendices follows*)

Appendix A

Materials

Table A1
Sets of Countries (Population in Brackets) Used in Experiment 1

Set 1		Set 2		Set 3		Set 4	
Pakistan	(203,177,034)	Brazil	(211,819,321)	Russia	(143,919,453)	Indonesia	(268,501,680)
Bangladesh	(167,422,187)	Nigeria	(199,045,324)	Japan	(126,976,591)	Philippines	(107,505,862)
Mexico	(131,738,729)	Egypt	(100,488,879)	Ethiopia	(109,159,044)	Thailand	(69,256,846)
Democratic Republic of the Congo	(85,705,256)	Vietnam	(97,074,662)	Iran	(82,518,959)	United Kingdom	(66,816,286)
Tanzania	(60,229,204)	Turkey	(82,592,416)	France	(65,387,848)	Italy	(59,246,609)
Colombia	(49,705,306)	South Africa	(57,812,482)	Myanmar	(54,158,522)	South Korea	(51,273,440)
Ukraine	(43,877,093)	Argentina	(44,946,136)	Poland	(38,056,163)	Spain	(46,439,538)
Malaysia	(32,294,009)	Iraq	(40,002,380)	Canada	(37,156,729)	Uganda	(45,169,147)
North Korea	(25,683,863)	Peru	(32,790,012)	Morocco	(36,468,117)	Algeria	(42,425,837)
Cameroon	(25,074,109)	Angola	(31,404,292)	Saudi Arabia	(33,912,223)	Afghanistan	(36,883,979)
Australia	(24,970,495)	Ghana	(29,858,634)	Uzbekistan	(32,642,668)	Mozambique	(31,077,768)
Niger	(22,850,032)	Nepal	(29,822,097)	Venezuela	(32,630,416)	Madagascar	(26,704,247)
Zimbabwe	(17,154,637)	Yemen	(29,329,832)	Ivory Coast	(25,295,354)	Burkina Faso	(20,106,983)
Netherlands	(17,114,912)	Malawi	(19,510,631)	Sri Lanka	(20,992,622)	Romania	(19,519,762)
Guinea	(13,270,289)	Mali	(19,471,687)	Chile	(18,284,455)	Guatemala	(17,452,973)
Benin	(11,683,042)	Kazakhstan	(18,518,517)	Zambia	(17,938,326)	Senegal	(16,574,342)
Greece	(11,133,944)	Ecuador	(17,011,566)	Chad	(15,639,892)	South Sudan	(13,132,406)
Czech Republic	(10,629,078)	Cambodia	(16,394,043)	Cuba	(11,489,711)	Bolivia	(11,318,180)
Azerbaijan	(9,980,369)	Belgium	(11,539,843)	Burundi	(11,443,124)	Haiti	(11,193,953)
Tajikistan	(9,222,905)	Dominican Republic	(10,953,914)	Portugal	(10,269,227)	Jordan	(10,000,697)
Laos	(7,027,153)	Switzerland	(8,582,983)	Sweden	(10,026,898)	Hungary	(9,667,861)
El Salvador	(6,432,904)	Slovakia	(5,450,438)	Austria	(8,758,508)	Belarus	(9,439,781)
Ireland	(4,833,127)	Croatia	(4,149,214)	Bulgaria	(7,006,598)	Georgia	(3,908,462)

(*Appendices continue*)

Table A2

Sets of Countries (Population in Brackets) and the Movie Set (Movie Budgets in Brackets) Used in Experiment 2

Set 1		Set 2		Set 3		Movie set[a]	
Pakistan	(203,177,034)	Brazil	(211,819,321)	Indonesia	(268,501,680)	Avatar	(237,000,000)
Nigeria	(199,045,324)	Bangladesh	(167,422,187)	Russia	(143,919,453)	Interstellar	(165,000,000)
Mexico	(131,738,729)	Philippines	(107,505,862)	Japan	(126,976,591)	Harry Potter and the Sorcerer's Stone	(125,000,000)
Vietnam	(97,074,662)	Egypt	(100,488,879)	Ethiopia	(109,159,044)	Gladiator	(103,000,000)
Democratic Republic of the Congo	(85,705,256)	Turkey	(82,592,416)	Iran	(82,518,959)	Bruce Almighty	(81,000,000)
Thailand	(69,256,846)	Italy	(59,246,609)	United Kingdom	(66,816,286)	Les Misérables	(65,000,000)
Tanzania	(60,229,204)	South Africa	(57,812,482)	France	(65,387,848)	Gone Girl	(61,000,000)
South Korea	(51,273,440)	Kenya	(50,220,000)	Myanmar	(54,158,522)	Moulin Rouge	(53,000,000)
Colombia	(49,705,306)	Spain	(46,439,538)	Poland	(38,056,163)	Miss Congeniality	(45,000,000)
Uganda	(45,169,147)	Argentina	(44,946,136)	Canada	(37,156,729)	Little Women	(42,000,000)
Ukraine	(43,877,093)	Algeria	(42,425,837)	Morocco	(36,468,117)	Fifty Shades of Grey	(40,000,000)
Peru	(35,000,000)	Iraq	(40,002,380)	Saudi Arabia	(33,912,223)	Hangover	(32,790,012)
Malaysia	(32,294,009)	Afghanistan	(36,883,979)	Uzbekistan	(32,642,668)	Star Wars: Return of the Jedi	(32,500,000)
North Korea	(25,683,863)	Angola	(31,404,292)	Venezuela	(32,630,416)	La La Land	(30,000,000)
Niger	(22,850,032)	Ghana	(29,858,634)	Mozambique	(31,077,768)	Die Hard	(28,000,000)
Burkina Faso	(20,106,983)	Nepal	(29,822,097)	Ivory Coast	(25,295,354)	James Bond—Octopussy	(27,500,000)
Romania	(19,519,762)	Yemen	(29,329,832)	Sri Lanka	(20,992,622)	Bodyguard	(25,000,000)
Zimbabwe	(17,154,637)	Madagascar	(26,704,247)	Chile	(18,284,455)	12 Years a Slave	(22,000,000)
Netherlands	(17,114,912)	Cameroon	(25,074,109)	Zambia	(17,938,326)	Back to the Future	(19,000,000)
Somalia	(14,600,000)	Australia	(24,970,495)	Guatemala	(17,452,973)	Borat	(18,000,000)
Guinea	(13,270,289)	Malawi	(19,510,631)	Chad	(15,639,892)	Legally Blonde	(18,000,000)
Benin	(11,683,042)	Mali	(19,471,687)	South Sudan	(13,132,406)	Pitch Perfect	(17,000,000)
Haiti	(11,193,953)	Kazakhstan	(18,518,517)	Rwanda	(11,980,000)	Black Swan	(13,000,000)
Greece	(11,133,944)	Ecuador	(17,011,566)	Cuba	(11,489,711)	The Intouchables	(12,800,000)
Czech Republic	(10,629,078)	Senegal	(16,574,342)	Burundi	(11,443,124)	Brokeback Mountain	(14,000,000)
Azerbaijan	(9,980,369)	Cambodia	(16,394,043)	Tunesia	(11,430,000)	Scream	(15,000,000)
Tajikistan	(9,222,905)	Belgium	(11,539,843)	Portugal	(10,269,227)	Saw 3	(10,000,000)
Switzerland	(8,582,983)	Bolivia	(11,318,180)	Sweden	(10,026,898)	Midsommar	(9,000,000)
Laos	(7,027,153)	Dominican Republic	(10,953,914)	Jordan	(10,000,697)	Marriage Story	(8,600,000)
El Salvador	(6,432,904)	Hungary	(9,667,861)	Belarus	(9,439,781)	Little Miss Sunshine	(8,000,000)
Ireland	(4,833,127)	Slovakia	(5,450,438)	Austria	(8,758,508)	My Big Fat Greek Wedding	(5,000,000)
Georgia	(3,908,462)	Croatia	(4,149,214)	Bulgaria	(7,006,598)	High School Musical	(4,200,000)

[a] For the movies, not their actual budgets but a country population of a comparable number was shown. Movies were selected such that the actual budgets approximately matched the mean and range of country populations across the three sets.

Table A3

Characteristics of the Country Sets

Country set	*M*	OME	*r*
Experiment 1			
1[a]	40, 726, 861	0.62	0.48
2[a]	38, 582, 901	0.59	0.49
3[a]	40, 122, 068	0.59	0.47
4[a]	41, 150, 168	0.61	0.48
Experiment 2			
1[b]	42, 070, 764	0.63	0.46
2[b]	43, 922, 799	0.61	0.44
3[b]	43, 436, 345	0.60	0.43

Note. M = mean population; OME = mean order of magnitude error; r = mean rank-order correlation. Based on an online study with $N = 100$ participants.
[a] 23 countries. [b] 32 countries.

(*Appendices continue*)

Appendix B

Additional Experimental Manipulation

The goal of this experimental manipulation ("graph" condition) was to test whether a graph containing a few informative numbers (minimum, maximum, and mean of country populations) would also elicit learning effects. If hindsight effects are a by-product of knowledge updating, then any manipulation that leads to learning should trigger a hindsight effect. The procedure was the same as in the other conditions (Figure 2), but instead of being presented with either text (control and concurrent-feedback conditions) or actual values (preceding-feedback condition) in the second phase of the experiment, participants were presented with a graph that gradually built up to show (a) the distribution of country populations across the world, together with the largest (maximum) country population ("Largest country population: China: 1.38 billion inhabitants" shown for 3 s), (b) the distribution of populations, together with the smallest (min) country population ("Smallest country population: Pitcairn islands: 54 inhabitants" for 3 s), and (c) the distribution of populations, together with the mean country population ("Mean country population: 31 million inhabitants" for 3 s). The final graph including all of the information remained on screen for an additional 3 s. Prior to the analyses, we excluded data from 10 participants for whom we could not ensure sufficient compliance with the task instructions (see main text for details), leaving a total of 71 participants in the analysis. There was clear evidence that this manipulation neither improved the accuracy of repeated judgments, OJ versus OJ_{again} ($BF_{10} < 0.01$), nor that of new items, OJ versus OJ_{new} ($BF_{10} = 0.01$), and it also did not bias the retrospectively reported judgments, OJ versus ROJ ($BF_{10} = 0.01$). The information provided in the graph was thus not effective in eliciting learning effects, and the manipulation was not suited to examining whether hindsight effects result from this type of learning.

Appendix C

Prior Specification

As the prior for the intercept parameter, we specified a normal distribution `normal(1, 1.5)`. As priors for slope parameters, we implemented two different specifications, one that is "skeptical" with regard to an effect and places a lot of prior probability around zero, `normal(0, 2.5)`, and one that can be considered "weakly informative," `normal(-0.5, 1)`, as it has more probability mass around an effect (i.e., of negative sign, indicating that the OME decreases from OJ to the other tasks), but with considerable probability mass left around zero and still allowing for unexpected effects in the opposite direction (i.e., an increase in OME). Analyses with the weakly informative priors are reported in the main text. In the online supplemental materials, we describe differences in the results emerging when we used a skeptical versus a weakly informative prior. In brief, as expected, it did not affect parameter estimation, but it did affect the size of the Bayes Factor for a few of the comparisons involving small effects. For the standard deviations of the random effects (i.e., participants and items), we specified half-normal distributions, `normal(0, 0.1)` with values > 0. For the random effect correlations, we assume an LKJ prior distribution with the prior parameter η of 2. For the residual standard deviation, we specified a half-normal distribution, `normal(0, 0.3)` with values > 0.

(*Appendices continue*)

Appendix D

Table D1
Percentage of Cases in Which the ROJ Equals the OJ

Item type	Control	Concurrent	Preceding	
Experiment 1				
Control (%)	21.8	24.2	20.7	
Experimental (%)	—	20.8	21.8	
	Control	Feedback	Domain	Irrelevant
Experiment 2				
Control (%)	27.6	—	—	—
Experimental (%)	—	23.0	27.7	25.7

Note. Control = control condition; Concurrent = concurrent-feedback condition; Preceding = preceding-feedback condition; Feedback = feedback condition; Domain = domain-information condition; Irrelevant = irrelevant-information condition; OJ = original judgment; ROJ = recall of original judgment.

Appendix E

Mapping Accuracy

Table E1
Rank-Order Correlations Between Estimated and Actual Values in Experiment 1

	Control		Concurrent		Preceding	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*
OJ	0.49	(0.15)	0.48	(0.17)	0.52	(0.19)
ROJ	0.50	(0.17)	—	—	—	—
Control	—	—	0.50	(0.19)	0.51	(0.22)
Experimental	—	—	0.69	(0.12)	0.61	(0.17)
OJ_{new}	0.50	(0.17)	0.51	(0.16)	0.53	(0.19)
OJ_{again}	0.51	(0.16)	—	—	—	—
Control	—	—	0.52	(0.21)	0.54	(0.24)
Experimental	—	—	0.67	(0.19)	0.66	(0.22)

Note. Rank-order correlations were *r*-to-*z*-transformed for statistical analyses (see Appendix Table E2); shown are the means and standard deviations of the back-transformed correlations. Control = control condition; Concurrent = concurrent-feedback condition; Preceding = preceding-feedback condition; OJ = original judgment; ROJ = recall of original judgment; OJ_{again} = repeated original judgment; OJ_{new} = new original judgment.

Table E2
Improvements in Mapping Accuracy in Experiment 1

	Control	Concurrent	Preceding
Measurement of effects	BF_{10}	BF_{10}	BF_{10}
Direct-learning effects			
OJ versus OJ_{again} (control)	0.08	0.92	0.16
OJ versus OJ_{again} (experimental)	0.08	>10,000	>10,000
Hindsight effects			
OJ versus ROJ (control)	0.04	0.15	0.09
OJ versus ROJ (experimental)	0.04	>10,000	6.5
Transfer-learning effects			
OJ versus OJ_{new}	0.06	0.20	0.09

Note. Rank-order correlations were *r*-to-*z*-transformed for statistical analyses. Shown are BFs quantifying the evidence for the model including the fixed-effect predictor of interest relative to a baseline model without the fixed-effect predictor of interest (BF_{10}) with a skeptical prior on the slope parameter, `normal(0, 0.5)`. Control = control condition; Concurrent = concurrent-feedback condition; Preceding = preceding-feedback condition; OJ = original judgment; ROJ = recall of original judgment; OJ_{again} = repeated original judgment; OJ_{new} = new original judgment; BFs = Bayes factors.

(*Appendices continue*)

Table E3
Rank-Order Correlations Between Estimated and Actual Values in Experiment 2

	Control		Feedback		Domain		Irrelevant	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
OJ	0.56	(0.21)	0.49	(0.20)	0.52	(0.19)	0.48	(0.19)
ROJ	0.56	(0.22)	0.57	(0.19)	0.54	(0.19)	0.46	(0.20)
OJ_{new}	0.54	(0.23)	0.55	(0.21)	0.54	(0.20)	0.49	(0.20)

Note. Rank-order correlations were *r*-to-*z*-transformed for statistical analyses (see Appendix Table E4); shown are the means and standard deviations of the back-transformed correlations. Control = control condition; Feedback = feedback condition; Domain = domain-information condition; Irrelevant = irrelevant-information condition; OJ = original judgment; ROJ = recall of original judgment; OJ_{new} = new original judgment.

Table E4
Improvements in Mapping Accuracy in Experiment 2

	Control	Feedback	Domain	Irrelevant
Measurement of effects	BF_{10}	BF_{10}	BF_{10}	BF_{10}
Hindsight effects				
OJ versus ROJ	0.14	191.1	0.36	0.11
Transfer-learning effects				
OJ versus OJ_{new}	0.08	1.88	0.27	0.07

Note. Rank-order correlations were *r*-to-*z*-transformed for statistical analyses. Shown are BFs quantifying the evidence for the model including the fixed-effect predictor of interest relative to a baseline model without the fixed-effect predictor of interest (BF_{10}) with a skeptical prior on the slope parameter, `normal(0, 0.5)`. Control = control condition; Feedback = feedback condition; Domain = domain-information condition; Irrelevant = irrelevant-information condition; OJ = Original Judgment; ROJ = recall of original judgment; OJ_{new} = new original judgment; BFs = Bayes factors.

Received August 25, 2022
Revision received March 10, 2023
Accepted May 18, 2023 ■