UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Kreis, B. K., Hermann, A., Pachur, T., & Groß, J. (2025). Hindsight Bias Through Knowledge Updating: A Conceptual Replication of Groß et al. (2023). *Collabra: Psychology*, *11*(1). https://doi.org/10.1525/collabra.147499

Cognitive Psychology

Hindsight Bias Through Knowledge Updating: A Conceptual Replication of Groß et al. (2023)

Barbara K. Kreis[1][a], Antje Hermann[1], Thorsten Pachur[2,3], Julia Groß[1]

[1] Department of Psychology, University of Mannheim, Mannheim, Germany, [2] School of Management, Technical University of Munich, Munich, Germany, [3] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

Keywords: hindsight bias, knowledge updating, replication, real-world estimation

https://doi.org/10.1525/collabra.147499

Collabra: Psychology

Vol. 11, Issue 1, 2025

Hindsight bias is the phenomenon that after learning facts about previously judged objects people tend to recall their previous judgments of the objects as closer to the facts than they actually were. Groß, Kreis, Blank, and Pachur (2023) found that hindsight bias emerged not only when people learned actual values for previously judged objects, but also when they learned the values of other objects in the same domain. Moreover, hindsight bias co-occurred with improved estimates for new objects. These findings challenge the traditional view that hindsight bias reflects a cognitive error and instead suggest that it results from adaptive knowledge updating. Groß et al. (2023) obtained their findings in the domain of country populations, a domain where people tend to be unfamiliar with the content and the numerical range (up to several million); this lack of familiarity may affect the link between knowledge updating and hindsight bias. In a high-powered conceptual replication (N = 300), we tested whether the findings generalize to the sugar content of food items—a domain where people are more familiar with both content and numerical range (up to 50 grams). Participants provided original judgments for items, learned numerical information, then recalled their original judgments, and lastly provided judgments for a new set of items. Our results replicate the key results of Groß et al. (2023), showing a close link between hindsight bias and knowledge updating in a more familiar domain. We discuss implications for theories of hindsight bias and propose directions for future research.

Introduction

Have you ever wondered about the amount of sugar in a jam doughnut? Say you guess 15 grams, then look it up and discover that the true figure is 29 grams. Later, when recalling your initial estimate, you might confidently think, “I guessed 25 grams.” What you have experienced is *hindsight bias* (Blank et al., 2007; Fischhoff, 1975, 2025; Hawkins & Hastie, 1990; Roese & Vohs, 2012). Hindsight bias occurs when people estimate a quantity or predict an event’s outcome, learn the actual value or result, and then recall their initial judgment as being closer to the truth than it actually was. This phenomenon has been demonstrated across various contexts and materials, including real-world knowledge (e.g., answers to almanac questions, numerical estimates; Erdfelder & Buchner, 1998; Groß et al., 2023) and outcomes (e.g., medical cases, elections; Arkes et al., 1988; Blank et al., 2003). Hindsight bias is traditionally viewed as an illusion or cognitive error that results from the limitations of the human mind (e.g., Pohl, 2022). One account posits that it is due to people’s use of the *anchoring-and-adjustment* heuristic (Tversky & Kahneman, 1974), the assumption being that people use the new information (i.e., the actual value or outcome) as an anchor when reconstructing their initial judgment, but fail to adjust sufficiently away from it (Hawkins & Hastie, 1990; Tversky & Kahneman, 1974).

More recently, an alternative view has been proposed, according to which hindsight bias might represent a by-product of an adaptive learning process—knowledge updating (Groß et al., 2023; Hoffrage et al., 2000; Nestler et al., 2012). From this perspective, people update the knowledge base they used to generate their initial judgment based on the new information. When asked to recall their initial judgment, they then rejudge the object based on this updated knowledge, leading to a systematic discrepancy from the initial judgment (Hoffrage et al., 2000). From this *knowledge-updating-and-rejudgment* perspective, hindsight bias is a consequence of adaptive information integration, rather than reflecting a cognitive error.

a Correspondence concerning this article should be addressed to Barbara K. Kreis at barbara.kreis@uni-mannheim.de

A set of studies by Groß et al. (2023) has supported the knowledge-updating account of hindsight bias. The authors derived a set of predictions (detailed below) and tested them in the context of numerical estimation, with country populations (ranging in the millions) as the knowledge domain. Their results demonstrated for the first time that hindsight bias arises not only when people learn the actual values of objects (as commonly shown), but also when they learn the values of other objects in the same domain. That is, hindsight bias emerged when participants were presented with any type of numerical information that allowed for knowledge updating. These results challenge the traditional theoretical assumptions on the psychological underpinnings of hindsight bias.

In light of their novelty and theoretical relevance, it is important to test whether Groß et al.'s (2023) findings can be replicated and extended beyond the domain of country populations. This extension is critical because people have limited familiarity with both the objects of this domain (Brown & Siegler, 1993; LaVoie et al., 2002) and the numerical range it covers (in the millions; Landy et al., 2013; Resnick et al., 2017), which may affect both hindsight bias and knowledge updating processes. Why? Greater familiarity with objects is known to facilitate memory formation and recall (Bellana et al., 2021) and could thus reduce hindsight bias (e.g., Christensen-Szalanski & Willham, 1991; Hertwig et al., 2003). Additionally, participants' calibration to a given numerical range, which is typically poorer in less familiar ranges, is linked to errors in their numerical judgments (e.g., Boyce-Jacino et al., 2022; Landy et al., 2013; Louie et al., 2015; Resnick et al., 2017); it might thus also influence how effectively participants improve their judgments through knowledge updating. However, it is unclear whether hindsight bias and knowledge updating are equally influenced by familiarity and whether the degree of familiarity impacts the observed association between them. It is thus important to explore whether the link established by Groß et al. (2023) generalizes to a domain with which people are more familiar and that involves a smaller and thus more familiar numerical range than with country populations. In this article, we address this issue by conducting a preregistered (https://osf.io/2pr76) conceptual replication of Groß et al. (2023) in which we expand the study to a different domain, the sugar content of food items, while retaining the general structure, design, and analytic approach of the original study.

The Knowledge-Updating Account of Hindsight Bias

To examine the knowledge-updating account of hindsight bias, Groß et al. (2023) connected hindsight bias research to research on *seeding effects* in numerical real-world estimation (Brown & Siegler, 1993). This research has shown that presenting actual values for a set of objects, so-called *seed facts* (e.g., that Spain has a population of 48 million inhabitants), leads to an improvement of people's estimation accuracy not only for subsequent judgments of the seeded objects (e.g., Spain), but also for nonseeded objects from the same domain (e.g., France or Canada; Bröder et al., 2023; Brown & Siegler, 1993; Groß et al., 2024; LaVoie et al., 2002). We refer to this effect as *transfer learning*. It is assumed that the seed facts lead to an update of the underlying metric knowledge about the domain (i.e., knowledge of the range, central tendency, or distribution of country populations; Brown & Siegler, 1993), thus improving estimation accuracy.

Groß et al. (2023) proposed that the actual values of objects that lead to hindsight bias should also function as seed facts, prompting an update of the underlying metric knowledge and thereby improving subsequent judgments. Incorporating insights from seeding research, the authors derived three predictions that should hold if hindsight bias emerges, at least partly, due to knowledge updating. First, presenting actual values for a set of objects should not only elicit hindsight bias for those objects, but also prompt an update of the underlying metric knowledge for the whole domain, leading to improved estimates for both the original objects and other objects from the domain (i.e., transfer learning). Second, if hindsight bias is a consequence of knowledge updating, then any numerical information that people can use to update their domain knowledge should induce hindsight bias. Therefore, presenting values for other objects from the same domain should not only elicit transfer learning but also trigger hindsight bias for the previously estimated objects (hindsight bias via domain information). Third, when the values presented do not provide information relevant to updating the knowledge base (e.g., because they refer to a different domain), no hindsight bias should be triggered (no hindsight bias via irrelevant information).

To investigate hindsight bias and transfer learning simultaneously and test whether they might co-occur as consequences of knowledge updating, Groß et al. (2023) developed an *Integrated Hindsight-Bias-and-Seeding Paradigm*. In this paradigm, participants first judge a set of objects, are then presented with numerical information, are subsequently asked to recall their initial judgments (as a measure of hindsight bias), and finally provide estimates for other objects from the same domain (as a measure of transfer learning). Using this paradigm, Groß et al. (2023, Exp. 2) obtained support for all three predictions. The knowledge domain used in Groß et al. (2023), country populations, is commonly used in seeding research. We investigated the extent to which their results replicate in a domain where people can be expected to be more familiar with both the content and the numerical range.

The Present Study

Here we conduct a conceptual replication of Groß et al. (2023, Exp. 2), using the sugar content of food items as the knowledge domain. It is plausible that people are relatively familiar with this domain, as they are often exposed to information on sugar content—for example, via nutrition labels or recipes. Relative familiarity with the numerical range of the objects can also be assumed (typically 0 to 50 grams of sugar per serving or purchase size).

As in the original study, we used the Integrated Hindsight-Bias-and-Seeding Paradigm (Groß et al., 2023). Participants first completed an estimation task for a set of objects (original judgments, OJs). In an information phase, they were then presented with the actual values for those same objects (feedback group), the values for other objects in the same domain (domain-information group), or the actual values for the original objects but relabeled as belonging to another domain (irrelevant-information group). In the control group, no actual values were presented. Instead, participants read a text unrelated to the topic of the study. Participants were then asked to recall their original estimates (recall of original judgments, ROJs) and to provide estimates for a new set of objects (new judgments, NJs). If the ROJs are closer to the actual values than the OJs, then this would suggest the presence of hindsight effects; if the NJs are closer to the actual values than the OJs, then this would suggest the presence of transfer learning effects.

We tested the three predictions that Groß et al. (2023, Exp. 2) derived from the knowledge-updating account of hindsight bias. Our hypotheses were as follows.[1] First, we expected that presenting the actual sugar content values for a set of previously estimated food items would both trigger hindsight effects for those food items and improve estimation accuracy for new food items (hindsight and transfer learning effects in the feedback group). Second, we expected that not only presenting actual sugar content values for a set of previously estimated food items but also presenting values for other items from the same domain would also both trigger hindsight effects for the previously estimated items and improve estimation accuracy for new food items (hindsight and transfer learning effects in the domain-information group). Third, we expected that presenting the actual sugar content values relabeled as referring to an unrelated domain—here, longitudes of European cities—would not lead to either hindsight effects or improved estimation accuracy for new food items (no hindsight or transfer learning effects in the irrelevant-information group). Finally, we did not expect to observe hindsight or transfer learning effects in the control group, where no actual values were presented (no hindsight or transfer learning effects in the control group).

Method

The experiment was preregistered (see https://osf.io/2pr76). Deviations from the preregistration are reported in an overview document on OSF (see https://osf.io/hr5bv/).

Participants

The experiment was conducted online. Participants were recruited through Prolific (www.prolific.com) with the same general eligibility criteria as in the original study by Groß et al. (2023, Exp. 2): Participation was limited to native German speakers aged 18 to 45 years. Additionally, we required that the study was conducted on a laptop, desktop computer, or tablet, rather than a smartphone. This was to ensure that the items, which were accompanied by food images, were properly displayed. We additionally excluded individuals who reported type 1 or type 2 diabetes or an eating disorder (e.g., anorexia nervosa, bulimia nervosa, binge-eating disorder). The median completion time of the experiment was 18.09 minutes, and participants received a fixed compensation of £4.80.

Overall, $N = 340$ participants completed the experiment. They were, on average, 27.9 years old, with 142 identifying as women, 193 as men, and 5 as diverse. In the original study ($N = 295$), participants were, on average, 27.8 years, with 116 women, 178 men, and 1 other. In the present study, most participants reported a university degree as their highest level of education (157), followed by a certificate of higher secondary education (122), vocational training (24), a certificate of lower secondary education (24), and a PhD (13). The majority of participants were currently working (177) or studying (university or college; 126); 13 participants were not working but looking for work; 4 were not working and not looking for work; 14 were apprentices; 6 were in high school. In the original sample, 130 participants were working, 122 studying (university or college), 29 looking for work, 7 in high school, 6 apprentices, and 1 was retired. The present sample was thus similar to that of Groß et al. (2023, Exp. 2) in terms of these demographic characteristics.

The target sample size for the present study was determined by conducting an a priori frequentist simulation-based power analysis with the `mixedpower` package in R (Kumle et al., 2021). The code for this analysis is available via https://osf.io/hr5bv/. We based the simulation on the results of the effect of interest of Experiment 2 in Groß et al. (2023): the hindsight effect in the domain-information group. To account for uncertainty of the expected effect size, we simulated power for the smaller boundary (in absolute terms) of the $CI_{90\%}$ of the effect. We ran 5,000 simulations and defined a critical t value of 2, reflecting an α level of 5%. The simulations showed that a power of 80% would be achieved with 80 participants, resulting in an overall target sample size of 320 for four groups. To account for potential exclusions, we collected data from 20 additional participants, resulting in a total sample size of 340.

Materials

Participants estimated the sugar content (in grams) of common food items. In contrast to the original study, where items were presented as text only (e.g., “How many people

1 Note that in the preregistration, hypotheses were formulated from the perspective of the effects (hindsight effects and transfer learning effects), whereas here they are presented from the perspective of the experimental groups, which we considered more intuitive for understanding hindsight effects and transfer learning effects as co-occurring consequences of knowledge updating. This represents a difference in framing rather than a substantive deviation from the preregistered hypotheses.

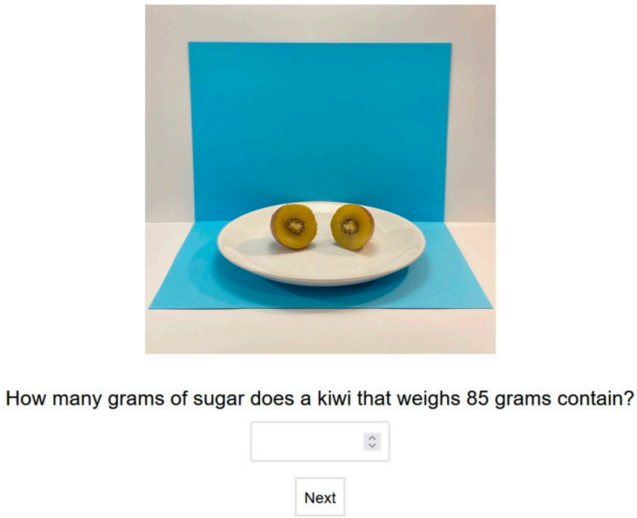

Figure 1. Sample Item from the Original Judgment Task

Note. Translated from German.

live in Peru?"), we included an image and specified the weight of each food item (e.g., "How many grams of sugar does a kiwi that weighs 85 grams contain?") to clarify the size of the food item and increase ecological validity (see Figure 1 for a sample item).

As in the original study by Groß et al. (2023), we used 96 items. These were taken from eight food subcategories (grains and grain products; vegetables; fruit; milk and dairy products; fish, meat, and sausage; oils and fats; drinks; and sweets and snacks), with each subcategory containing 12 items.[2] For most items, participants were asked to estimate the sugar content of a typical serving size, such as a banana; for others, the sugar content of a typical purchase size, such as a can of coconut milk. The food items varied in sugar content ($M = 7.6$ grams, $SD = 7.7$, $min = 0$, $max = 33$) and degree of processing (based on the NOVA classification system; Monteiro et al., 2019): unprocessed or minimally processed foods (e.g., a banana or whole milk), processed culinary ingredients (e.g., a tablespoon of butter or a tablespoon of honey), processed foods (e.g., a serving of spaghetti or a can of sweetcorn), and ultra-processed foods (e.g., a bowl of sausage salad or a glass of energy drink). We photographed all items and edited the pictures to blur any visible labels, nutritional information, and numbers. All food images are available at https://osf.io/hr5bv/. A list of all food items, their subcategories, size (serving or purchase size in grams or milliliters), and sugar content (in grams) can be found in Appendix A1. From these 96 items, we created three sets of 32 items each. Each set contained four items from each of the eight food subcategories (see Appendix A2 for details). The sets were designed to be comparable in terms of degree of processing and statistical characteristics of the sugar content (mean, median, range). Thus, assignment of items to item sets followed similar principles as in the original study.

In the experiment, one set of items was presented to all groups in both the OJ and ROJ task. In the information phase, actual values for the same set were presented to participants in the feedback group; the same values, but relabeled as longitudes of European cities (which have a similar numerical range to the sugar content values), to participants in the irrelevant-information group. Participants in the domain-information group were presented with values for a second set of food items. Finally, all groups were presented with a third set of food times in the NJ task.

To assess participants' familiarity with the sugar content of food items in terms of engagement, we asked the following question: "At the beginning of this study, you answered several questions about the sugar content of food items. How *frequently*, prior to this study, have you engaged with this topic (e.g., at school, at work, or in your leisure time)?" (translated from German). Participants indicated their answer on a 7-point scale, ranging from *very rarely* to *very frequently*.

Design and Procedure

The design and procedure of the present experiment closely followed that of Groß et al. (2023, Exp. 2). The experiment was programmed with lab.js (Henninger et al., 2022) and hosted via Jatos (Lange et al., 2015). It consisted of four phases: Participants (1) completed the OJ task, (2) were presented with actual values of different types depending on the experimental group to which they were randomly assigned, (3) completed the ROJ task, and (4) completed the NJ task. This resulted in a three (task: OJ, ROJ, NJ) by four (group: control, feedback, domain information, irrelevant information) mixed design.

After providing informed consent, participants entered the first phase, the OJ task, where they estimated the sugar content of 32 food items (e.g., "How many grams of sugar does a kiwi that weighs 85 grams contain?"). The items were presented sequentially in a randomized order for each participant. Participants took a median of 6.4 seconds per estimate (which was faster than in the original study: 9.4 seconds). In the second phase, the information phase, participants were presented with different types of values depending on their assigned group. Those in the *feedback* group were shown the actual values of the 32 food items from the OJ task (e.g., "A kiwi that weighs 85 grams contains 8 grams of sugar."). Those in the *domain-information* group were shown the values of a different set of 32 food items (e.g., "A plum that weighs 75 grams contains 7 grams of sugar."). Those in the *irrelevant-information* group saw the same values as participants in the feedback group (i.e., the actual sugar content values) but relabeled as the lon-

2 The definition of the eight food subcategories was based on the categorization guidelines by the German Society for Nutrition (Deutsche Gesellschaft für Ernährung e. V., 2024) and the Federal Center for Nutrition (Bundeszentrum für Ernährung, 2023).

gitudes of 32 well-known European cities, along with the flag of the respective country (e.g., “Basel, a major city in Switzerland, is located at longitude 8.”). This domain was selected to match the mean and range of the sugar content values across the three sets (see Appendix A2 for details). In all three groups, each item was presented for 8 seconds, resulting in an overall presentation duration of 256 seconds. In the *control* group, participants were presented with a text that was unrelated to the topic of the study (filler task) for an overall duration of 256 seconds. The presentation time was extended from 5 seconds per item in the original study to 8 seconds per item here to account for the more complex combination of quantity information and an accompanying image.

In the third phase, the ROJ task, participants were asked to recall their original judgments in the OJ task for all 32 items, presented in the same randomized order (e.g. “How many grams of sugar does a kiwi that weighs 85 grams contain? What was your ORIGINAL answer?”). In the fourth and final phase, the NJ task, participants estimated the sugar content of a new set of 32 food items (“How many grams of sugar does a tangerine that weighs 90 grams contain?”). The assignment of sets was counterbalanced, with each set being presented with comparable frequency across participants and experimental groups. After completing all four phases, participants were asked whether they had cheated or clicked through the experiment and whether they had any problems with the display of images, before indicating their prior engagement with the domain. Finally, they had the opportunity to enter any additional comments into an open response field.

Data Diagnostics

We preregistered several data assessment steps comparable to those applied in Groß et al. (2023, Exp. 2) to ensure data quality by excluding both extreme outliers and data indicating non-compliance. First, Prolific automatically excluded participants who did not finish the experiment within the specified 87-minute time limit. Second, we excluded all participants whose meta-data indicated that contrary to the eligibility criteria they had used a smartphone (six participants). Third, we excluded participants who indicated non-compliance, such as cheating (one participant) or just clicking through the experiment (one participant). Fourth, we excluded participants who reported technical problems or disruptions (two participants who reloaded the experiment, one participant who reported a disturbance that led to a long break during the experiment, and three participants who reported problems with the display of several food images). Fifth, we excluded one participant for providing estimates that exceeded the weight of the food item for 10 or more items, as such sugar content is not possible. Sixth, we planned to exclude cases in which participants provided estimates in under 1000 ms for 10 or more items, as such short estimation processes are implausible; however, there were no such cases. Seventh, we excluded participants whose responses indicated extreme outliers. To identify such outliers, we relied on robust statistics (median and interquartile range rather than mean and standard deviation), as these measures are more robust in the context of skewed distributions that naturally occur in real-world data. Specifically, we excluded 21 participants whose estimates showed a median absolute deviation from the actual values (see Eq. 1 below) that exceeded the threefold interquartile range in at least one of the three experimental tasks (three in the control group, five in the feedback group, eight in the domain-information group, and five in the irrelevant-information group). Eighth, we excluded individual estimates that exceeded the weight of the food item (30 responses, 0.09% of all responses) and responses given in less than 1000 ms (one response, 0.003% of all responses). The final sample consisted of $N = 300$ participants. Sample size was comparable across experimental groups, with $n = 76$ in the control group, $n = 73$ in the feedback group, $n = 74$ in the domain-information group, and $n = 77$ in the irrelevant-information group.

Analytic Approach

To quantify estimation accuracy in each task—OJ, ROJ, and NJ—we calculated $|\Delta|$, the absolute deviation of the estimate from the actual value, for each item i and for each participant j:

$$|\Delta_{ij}| = |estimate_{ij} - actual_i| \tag{1}$$

Smaller values of $|\Delta|$ indicate a higher estimation accuracy. In Groß et al.'s (2023) original study, estimation accuracy was quantified with a logarithmic deviation measure, the order of magnitude error (OME). The use of this measure was not indicated in the present study for two reasons. First, the OME is indicated for highly skewed distributions spanning several orders of magnitude. This applied to the country population domain used by Groß et al. (2023), but does not apply to sugar content. Second, the frequent occurrence of zero values in sugar content estimates (8.9% of all estimates) and actual values (19.9% of all actual values) makes logarithmic measures problematic, as the logarithm of zero is undefined. In all other respects, the analytic approach in the present study was identical to that in the original study.

We applied Bayesian linear mixed-effects models to compare estimation accuracy across judgment tasks and groups, defining $|\Delta|$ as the criterion variable. As fixed effects, we specified the task, information group, and their interaction. Furthermore, we included random intercepts and random slopes for participants and items to capture by-person and by-item variability in estimation accuracy. For parameter estimation we used the `brms` package (Bürkner, 2017, 2018), which calls `STAN` for MCMC sampling (Stan Development Team, 2019). We report prior specification and sensitivity analyses in Appendix B. The general conclusions were robust across two alternative specifications of the priors.

To test our hypotheses, we compared a full model including the fixed-effect predictor of interest, M_1, with a baseline model M_0, that did not include that predictor but did include all random effects present in the full model. The model comparisons are described in detail in “Results”. We compared the models using the `bayes_factor` function in brms, which computes Bayes Factors (BF) based on bridge sampling (e.g., Gronau et al., 2017). The BF_{10} indicates the evidence for the alternative hypothesis relative to the null hypothesis, when comparing the full model M_1 to the baseline model M_0.[3] The data and the analysis code are available at https://osf.io/hr5bv/. To decide whether the present study replicated the findings of Groß et al. (2023, Exp. 2), we compared the Bayes Factors (BFs) for each effect with those reported in the original study. An effect was considered replicated if the BFs in both studies indicated evidence in the same direction and provided at least moderate evidence (i.e., BF > 3 in favor of the effect or BF $< 1/3$ against an effect). An effect was considered not replicated if the BFs indicated conflicting directions of evidence or if the evidence was weak (i.e., BF $> 1/3$ and < 3).

We performed additional analyses investigating another facet of domain knowledge—namely, the accuracy of the ranking of objects (mapping knowledge; Brown & Siegler, 1993)—as detailed in Appendix C.

Treatment of Perfectly Accurate Judgments

For all analyses, we excluded all OJs and corresponding ROJs where the OJ was an exact match for the actual value (7.4%), as in these cases neither a hindsight effect nor a transfer learning effect can occur (as NJ items are new items, there were no corresponding NJs to be excluded).[4] In Groß et al. (2023, Exp. 2), there were no cases of perfectly accurate judgments.

Treatment of Perfect OJ Reproductions

As in Groß et al. (2023, Exp. 2), we also identified perfect OJ reproductions, that is, cases where the ROJ matched the OJ exactly (e.g., due to correct recollection of the OJ or successful guessing). In such cases, hindsight effects are zero, and the higher the percentage of such cases, the smaller the overall size of hindsight effects will be. Any difference in hindsight effects between groups might thus be due to a genuine difference in hindsight effects, to different percentages of perfect OJ reproductions, or to both. As in previous investigations (Dehn & Erdfelder, 1998; Erdfelder et al., 2007; Erdfelder & Buchner, 1998), the percentages of perfect OJ reproductions in our study differed significantly between groups ($X^2(3) = 34.04$, $p < .001$), being lower in the feedback (35.23%) and the domain-information groups (33.08%) than in the irrelevant-information (39.67%) and control groups (39.67%). Following previous studies (e.g., Groß & Pachur, 2019; Pohl, 2007), we therefore excluded all perfectly matching OJ/ROJ pairs from our analyses.[5] In Groß et al. (2023, Exp. 2), the percentage of such cases was similar across groups (and they were therefore not excluded); additional analyses showed that the conclusions were unaffected by whether perfect OJ reproductions were excluded or not.

Results

Participants' self-reported engagement with the sugar content of food items (ranging from 1 = *very rarely* to 7 = *very frequently*) was, on average, 3.44 ($SD = 1.45$).[6] Participants' estimation accuracy in each of the four groups and three judgment tasks is plotted in Figure 2.

Transfer Learning Effects

We first tested whether presenting actual values led to transfer learning effects. This would be the case if $|\Delta|$ (i.e., the deviation of the estimate from the actual value) was smaller for the NJs than for the OJs. To test for transfer learning effects, we compared a model including the fixed-effect predictor task (OJ versus NJ) to a baseline model that did not include the predictor. As expected, in both experimental groups that saw domain-relevant values, there were transfer learning effects (feedback group: $b = -6.59$, $CI_{95\%} = [-8.63, -4.55]$, $BF_{10} > 10,000$; domain-information group: $b = -3.71$, $CI_{95\%} = [-5.09, -2.32]$, $BF_{10} > 10,000$). An interaction analysis indicated that the transfer learning effect was larger in the feedback group than in the domain-information group ($b = -3.11$, $CI_{95\%} = [-5.06, -1.16]$, $BF_{10} = 11$). There was no transfer learning effect in the irrelevant-information group ($b = 0.52$, $CI_{95\%} = [-0.45, 1.49]$, $BF_{10} = 0.08$) or the control group ($b = -0.92$, $CI_{95\%} = [-2.00, 0.15]$, $BF_{10} = 0.19$), with an interaction analysis indicating no difference between these two groups ($b = 1.35$, $CI_{95\%} = [-0.06, 2.77]$, $BF_{10} = 0.37$).

Hindsight Effects

Next we tested whether presenting actual values triggered a hindsight effect. This would be the case if $|\Delta|$ was smaller for the ROJs than for the OJs. To test for hindsight effects, we compared a model including the fixed-effect predictor task (OJ versus ROJ) to a baseline model that did not include the predictor. As expected, there was a

3 A BF_{10} below 1/10 is generally taken to indicate strong evidence, a BF_{10} between 1/10 and 1/3 moderate evidence, and a BF_{10} between 1/3 and 1 weak evidence for M_0. A BF_{10} larger than 10 is generally taken to indicate strong evidence, a BF_{10} between 3 and 10 moderate evidence, and a BF_{10} between 1 and 3 weak evidence for M_1 (e.g., Doorn et al., 2023; Jeffreys, 1998).

4 Robustness analyses showed that the patterns of results remained the same, whether these cases were excluded or not.

5 Robustness analyses showed that the transfer learning results were unaffected by whether those cases were excluded or not.

6 We tested whether self-reported engagement with the domain was associated with initial estimation accuracy (i.e., estimation accuracy in the OJ task), and the results were ambiguous. The credible interval of the regression weight did not include zero ($b = -0.51$, $CI_{95\%} = [-0.95, -0.07]$), indicating that more engagement was associated with higher initial estimation accuracy; however, the BF indicated moderate evidence against an effect ($BF_{10} = 0.24$). In addition, self-reported engagement was not associated with either hindsight effects or transfer learning effects in any of the experimental groups (all credible intervals included zero and all BF_{10} were below 0.32).

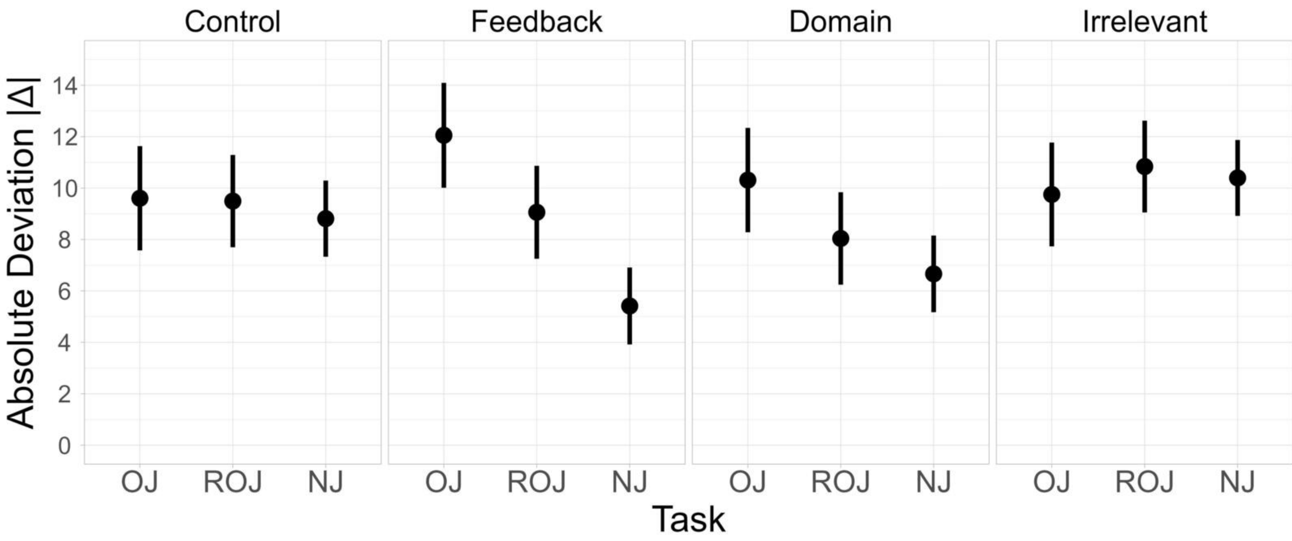

Figure 2. Estimation Accuracy

Note. Shown are the conditional predictions based on the mixed-effects model (estimated means and 95% credible intervals). OJ = Original Judgment, ROJ = Recall of Original Judgment, NJ = New Judgment. Control = Control group, Feedback = Feedback group, Domain = Domaininformation group, Irrelevant = Irrelevant-information group.

hindsight effect in the feedback group ($b = -2.88$, $CI_{95\%}$ = $[-4.19, -1.54]$, $BF_{10} = 268$). Importantly, there was also a hindsight effect in the domain-information group ($b = -2.25$, $CI_{95\%}$ = $[-3.32, -1.16]$, $BF_{10} = 80$). An interaction analysis showed that the size of the hindsight effect did not differ between these two groups ($b = -0.70$, $CI_{95\%}$ = $[-2.11, 0.71]$, $BF_{10} = 0.07$). Critically, there was no hindsight effect in the irrelevant-information group ($b = 0.87$, $CI_{95\%}$ = $[-0.06, 1.81]$, $BF_{10} = 0.23$) or the control group ($b = 1.10$, $CI_{95\%}$ = $[-0.15, 2.35]$, $BF_{10} = 0.25$), with an interaction analysis indicating no difference between the control and the irrelevant-information group ($b = -0.26$, $CI_{95\%}$ = $[-1.15, 0.64]$, $BF_{10} = 0.05$). In other words, when the actual values were presented to participants but relabeled as referring to a different domain (as in the irrelevant-information group), this did not impact the ROJs. This analysis was based on $|\Delta|$, the absolute deviation of the estimate from the actual value of each item. It is possible that the relabeled values may have affected ROJs in other ways—for instance, via the central tendency of the presented values, or only the most recently presented values. Such results could point to the operation of anchoring-and-adjustment processes. However, additional analyses showed no indication for such alternative influences of the relabeled values (for details, see Appendix D, and Groß et al. (2023)).

Comparison of Replication and Original Study Results

All key findings reported in this Results section replicated those reported in Groß et al. (2023, Exp. 2), that is, their corresponding Bayes Factors each indicated the same presence or absence of an effect. A detailed comparison of the Bayes Factors for each effect in the current study and those reported by Groß et al. (2023, Exp. 2) is provided in Appendix E. There were only two exceptions where the original results and ours diverged, both regarding the interaction analyses of the transfer learning effects. One, the transfer learning effect in the present study was larger in the feedback group than in the domain-information group, whereas in Groß et al. (2023, Exp. 2) there was no difference in the size of this effect between the two groups. Two, in the present study, the evidence against a difference in the size of this transfer learning effect between the irrelevant-information group and the control group was only weak, while the evidence against this effect was strong in Groß et al. (2023, Exp. 2). Importantly, this was likely due to the fact that the irrelevant-information group showed a slightly lower estimation accuracy in the OJ than in the NJ phase, as can be seen in Figure 2.

In sum, the results replicate the key novel finding of Groß et al. (2023, Exp. 2). Hindsight effects were triggered for objects even when participants were presented with the values of other objects in the same domain, not only when they were presented with the actual values of the original objects. That is, hindsight effects occurred when knowledge updating was enabled through the provision of relevant numerical information. Thus, hindsight bias also seems to be triggered by knowledge updating in a domain where people are relatively familiar with both the content and the underlying numerical range. Our results also replicate the finding in Groß et al. (2023) that a hindsight effect is not triggered when the numerical information presented is relabeled as referring to a different domain, thus hindering knowledge updating. Together, these findings support a clear link between hindsight bias and knowledge updating.

Discussion

Hindsight bias has long been seen as resulting from limitations of the human mind (Hawkins & Hastie, 1990; Tversky & Kahneman, 1974). More recently, it has been proposed that hindsight bias might in fact be a by-product of an adaptive learning process—knowledge updating (Groß et al., 2023; Hoffrage et al., 2000; Nestler et al., 2012). Investigating hindsight bias in the context of country population estimates, Groß et al. (2023) provided evidence that hindsight bias for objects can be elicited not only by actual values for those specific objects, but also by values for other objects in the same domain. This finding suggests that hindsight judgments reflected re-judgments of the objects based on an updated metric representation of the knowledge domain.

In the present study, we tested whether these findings can be replicated in the domain of sugar content of food items, with which participants are likely to be more familiar. There are several indicators for such higher familiarity. First, the initial accuracy of the ranking of objects, quantified as the rank-order correlation between estimates and actual values in the OJ task, was higher for sugar content (ρ = 0.61) than for country populations (ρ = 0.52; Groß et al., 2023).[7] Second, perfectly accurate judgments—that is, cases where the OJ matched the actual value exactly—were more frequent for sugar content than for country populations (7.4% vs. 0%; Groß et al., 2023), OJs were given faster (6.4 vs. 9.4 seconds per response; Groß et al., 2023), and participants reported more prior engagement with sugar content than did a comparable sample for country populations ($M = 3.44$ vs. $M = 2.58$; Kreis et al., 2024).[8]

As hypothesized, we observed indications for a close link between hindsight bias and knowledge updating, even in this more familiar judgment domain. Most importantly, we replicated the key novel finding by Groß et al. (2023) that hindsight effects for previously estimated objects can be triggered even without presenting the actual values of those objects—it can suffice to present the actual values of other objects in the same domain. Further, presenting actual values for previously estimated objects but relabeled as referring to an unrelated domain did not lead to either hindsight effects or transfer learning effects. Hindsight bias thus consistently emerged when participants were presented with relevant numerical information that allowed for knowledge updating, but not when that numerical information was relabeled, thus hindering knowledge updating.

It should be noted that our finding that a robust link between hindsight bias and knowledge updating emerges even in a domain with which people are more familiar does not imply that the amount of prior knowledge is irrelevant for hindsight bias and knowledge updating. Both are likely to be affected in tandem by whether prior knowledge is high or low. The greater the amount of prior knowledge, the less knowledge updating there is likely to be and hence the smaller hindsight bias. To examine this link for the present study, we compared the size of transfer learning and hindsight effects across the eight food subcategories (see section "Materials"), and examined whether the sizes of the two effects covaried with the amount of prior knowledge of the subcategories, reflected in participants' initial estimation accuracy (i.e., accuracy of the OJs). As Figure 3 shows, the sizes of transfer learning and hindsight effects seem to be closely linked and both depend on the amount of prior knowledge in a given subcategory. Food subcategories with higher initial estimation accuracy showed smaller transfer learning and hindsight effects, while those with lower initial estimation accuracy showed greater transfer learning and hindsight effects. For example, in the feedback group, for the Oils and fats subcategory, where initial estimation accuracy was rather high ($|\Delta|$ = 7.08), transfer learning and hindsight effects were relatively small, at $|\Delta|$ = 4.20 and 1.93, respectively; whereas for the Drinks subcategory, where initial estimation accuracy was rather low ($|\Delta|$ = 20.26), transfer learning and hindsight effects were relatively large, at $|\Delta|$ = 9.46 and 5.08, respectively.[9]

Implications

The results of our conceptual replication provide further evidence that hindsight bias, while traditionally viewed as cognitive error, might actually reflect knowledge updating and rejudgment (Groß et al., 2023; Hoffrage et al., 2000; Nestler et al., 2012). The findings indicate that hindsight bias is triggered whenever people learn numerical information about a domain that they can use to update their underlying knowledge—whether actual values for previously estimated items or values for other items from the same domain.

This finding highlights the pervasiveness of hindsight bias and may help explain why it has proven so difficult to eliminate in the context of numerical estimation. For example, hindsight bias has been shown to persist when participants are fully informed about the bias and then retested (Pohl & Hell, 1996). Even discrediting the actual values prior to the ROJ task (Erdfelder & Buchner, 1998) or presenting them as another person's estimates (Pohl, 1998) seems to only partially reduce hindsight bias. Yet this robustness of hindsight bias is to be expected if people integrate any numerical information that can be used to inform their underlying metric knowledge about a domain rather automatically (Fischhoff, 1975; Hawkins & Hastie, 1990). To the extent that people usually rejudge an object when trying to reconstruct their initial judgment for the object, hindsight bias thus represents an inevitable by-product of knowledge updating.

7 For further information on the analyses and results of the rank-order correlations, see Appendix C.

8 A direct comparison between the domains in terms of initial estimation accuracy is not possible because it needed to be quantified differently: For country populations, a logarithmic accuracy measure was needed to take into account the skewness and large range of values; the same did not apply to sugar content, where logarithmic measures were unsuitable due to the high prevalence of zero values (as discussed in "Analytic Approach").

9 Reported are the conditional predictions of the means based on the mixed-effects model.

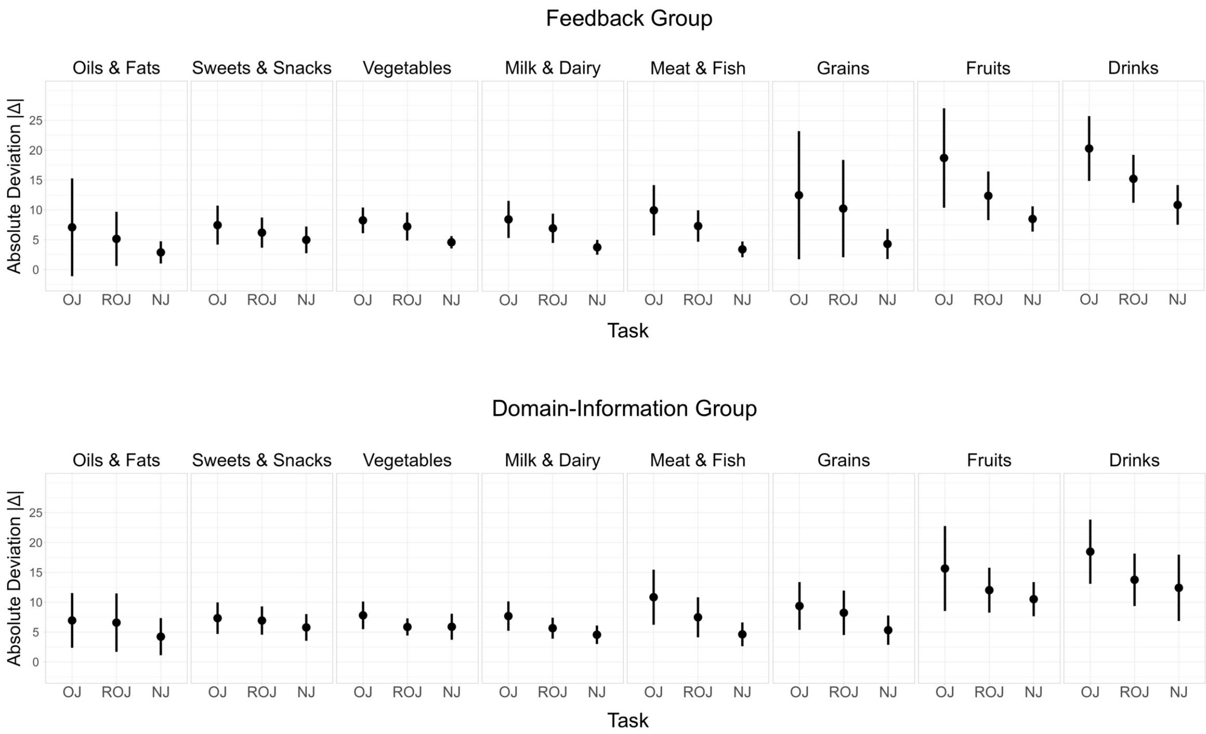

Figure 3. Comparison of Estimation Accuracy by Food Subcategory in the Feedback and Domain-Information Groups

Note. Shown are the conditional predictions based on the mixed-effects models by group and food subcategory (estimated means and 95% credible intervals). Food subcategories are sorted ascending by initial estimation accuracy (estimated mean |Δ| in the OJ task) in the feedback group. OJ = Original Judgment, ROJ = Recall of Original Judgment, NJ = New Judgment.

However, the results of the present study indicate that processes other than knowledge updating might also contribute to hindsight bias. In both the feedback and the domain-information groups, hindsight effects were smaller than transfer learning effects, with the ROJs being further away from the actual values than the NJs (see also Groß et al., 2023, Exp. 2). If both types of responses were exclusively (re-)judgments based on updated knowledge, that should not be the case. However, note that in the ROJ task, participants are asked to recall their original judgments, whereas in the NJ task, they are asked to provide estimates for new objects. Thus, participants seem to engage in additional processes in the ROJ task (e.g., adjusting the ROJ using episodic memory traces of the OJ to account for inaccuracies in initial estimation).

Outlook

While our findings add to the increasing evidence for a knowledge-updating account of hindsight bias, future research could further test this account by addressing additional aspects. For example, another approach to investigating the link between knowledge updating and hindsight bias could be to manipulate the time interval between the information phase and the ROJ task. Research on seeding effects suggests that the updating of metric knowledge is maintained over extended periods, with transfer learning effects remaining stable for up to four months regardless of whether the specific seed facts can still be recalled (Brown & Siegler, 1996; LaVoie et al., 2002). Thus, to the extent that hindsight bias is driven by processes of knowledge updating, it should remain stable over similar time frames. If hindsight bias diminishes over time, this would suggest that additional processes—such as anchoring and adjustment, which relies more on the recall of specific values—may be involved.

In the present study, we concluded that hindsight bias in the context of real-world estimation is primarily driven by knowledge updating, and we found no evidence for the influence of anchoring-and-adjustment processes—which would have manifested as a hindsight effect in the irrelevant-information group. It is conceivable, however, that the contribution of anchoring-and-adjustment processes depends on the order of presentation of actual values and ROJ task. In the present study, all actual values were presented first, and only then did participants provide ROJs. If actual values and ROJs were presented and provided alternately, anchoring processes might play a larger role; this is because the relabeled values could be more directly linked to subsequent ROJs, potentially increasing their influence (Hawkins & Hastie, 1990; Tversky & Kahneman, 1974).

Another promising avenue for future research is to investigate the relationship between hindsight bias and other judgment distortions arising from informational input, such as the misinformation effect (Loftus et al., 1978). This perspective raises important questions about how specifically subsequent information shapes retrospective judgments. For instance, misinformation research has long debated whether new input overwrites earlier representations or whether both coexist, with new information overshadowing the original at recall (Ayers & Reder, 1998). Addressing such possibilities in the context of seeding effects and hindsight bias may not only further clarify the mechanisms underlying these phenomena but also situate them more firmly within the broader landscape of memory and judgment distortions, thereby opening new directions for integrative research.

Conclusion

The nature and underlying mechanisms of hindsight bias have been debated for decades (Blank et al., 2007; Fischhoff, 1975; Hawkins & Hastie, 1990; Hoffrage et al., 2000; Roese & Vohs, 2012). The present study, together with that of Groß et al. (2023), contributes to the increasing evidence that hindsight bias reflects knowledge updating processes rather than a cognitive error (see also Hoffrage et al., 2000; Nestler et al., 2012). With this replication, which demonstrates the robustness and generalizability of the link between hindsight bias and knowledge updating across domains, we contribute to establishing a solid empirical foundation that can inform the continued development of theory and future research on hindsight bias (Eronen & Bringmann, 2021).

CRediT authorship contribution statement

Barbara K. Kreis served as lead for investigation, data curation, formal analysis, visualization, writing-original draft and validation and in a supporting role for resources and software. Antje Hermann served as a lead for resources and software and in a supporting role for investigation, data curation and validation. Thorsten Pachur served in a supporting role for conceptualization and supervision. Julia Groß served in a supporting role for validation, resources and writing-original draft. Barbara K. Kreis, Antje Hermann, Thorsten Pachur and Julia Groß served equally in writing-review and editing. Barbara K. Kreis, Antje Hermann and Julia Groß served equally in conceptualization and methodology. Barbara K. Kreis and Julia Groß served equally in project administration and supervision. Thorsten Pachur and Julia Groß served equally in funding acquisition.

Acknowledgements

The authors thank Susannah Goss for editing the manuscript.

Funding

This work was funded by Grant GR-4649/4-1 (PA 1925/2-1) from the German Research Foundation (DFG) which was awarded to Julia Groß and Thorsten Pachur. Barbara K. Kreis was further supported by the University of Mannheim's Graduate School of Economic and Social Sciences.

Data accessibility statement

Data and analysis code and research materials are available via the Open Science Framework https://osf.io/hr5bv/.

Ethics statement

The study protocol was approved by the Ethics Committee of the University of Mannheim (EK Mannheim 58/2020).

Competing interests

The authors have no conflicts of interest to disclose.

Editors: Sebastiaan Mathôt (Senior Editor)

Submitted: June 27, 2025 PST. Accepted: September 25, 2025 PST. Published: December 05, 2025 PST.

References

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, *73*(2), 305–307. https://doi.org/10.1037/0021-9010.73.2.305

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*(1), 1–21. https://doi.org/10.3758/BF03209454

Bellana, B., Mansour, R., Ladyka-Wojcik, N., Grady, C., & Moscovitch, M. (2021). The influence of prior knowledge on the formation of detailed and durable memories. *Journal of Memory and Language*, *121*, 104264. https://doi.org/10.1016/j.jml.2021.104264

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

Boyce-Jacino, C., Peters, E., Galvani, A. P., & Chapman, G. B. (2022). Large numbers cause magnitude neglect: The case of government expenditures. *Proceedings of the National Academy of Sciences*, *119*(28), e2203037119. https://doi.org/10.1073/pnas.2203037119

Bröder, A., Dülz, E., Heidecke, D., Wehler, A., & Weimann, F. (2023). Improving carbon footprint estimates of food items with a simple seeding procedure. *Applied Cognitive Psychology*, *37*(3), 651–659. https://doi.org/10.1002/acp.4060

Brown, N. R. (2002). Real-world estimation: Estimation modes and seeding effects. In *Psychology of learning and motivation* (Vol. 41, pp. 321–359). Academic Press. https://doi.org/10.1016/S0079-7421(02)80011-1

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

Bundeszentrum für Ernährung. (2023). *Die Ernährungspyramide [The food pyramid]*. Bundeszentrum für Ernährung. https://www.bzfe.de/ernaehrung/die-ernaehrungspyramide/die-ernaehrungspyramide-eine-fuer-alle/

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

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

Christensen-Szalanski, J. 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

Dehn, D. M., & Erdfelder, E. (1998). What kind of bias is hindsight bias? *Psychological Research*, *61*(2), 135–146. https://doi.org/10.1007/s004260050020

Deutsche Gesellschaft für Ernährung e. V. (2024). *DGE-Ernährungskreis [DGE Nutrition Circle]*. DGE. http://www.dge.de/gesunde-ernaehrung/gut-essen-und-trinken/dge-ernaehrungskreis/

Doorn, J. van, 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

Erdfelder, E., Brandt, M., & Bröder, A. (2007). Recollection biases in hindsight judgments. *Social Cognition*, *25*(1), 114–131. https://doi.org/10.1521/soco.2007.25.1.114

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

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, *16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B. (2025). Fifty years of hindsight bias research-Reflection on Fischhoff (1975). *Journal of Experimental Psychology. Human Perception and Performance*, *51*(2), 143–150. https://doi.org/10.1037/xhp0001232

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, *81*, 80–97. https://doi.org/10.1016/j.jmp.2017.09.005

Groß, J., Kreis, B. K., Blank, H., & Pachur, T. (2023). Knowledge updating in real-world estimation: Connecting hindsight bias and seeding effects. *Journal of Experimental Psychology: General*, *152*(11), 3167–3188. https://doi.org/10.1037/xge0001452

Groß, J., Loose, A. M., & Kreis, B. K. (2024). A simple intervention can improve estimates of sugar content. *Journal of Applied Research in Memory and Cognition*, *13*(2), 282–291. https://doi.org/10.1037/mac0000122

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

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

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). Lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

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

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

Jeffreys, H. (1998). *The theory of probability*. Oxford University Press.

Kreis, B. K., Groß, J., & Pachur, T. (2024). Real-world estimation taps into basic numeric abilities. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02575-4

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science*, *37*(5), 775–799. https://doi.org/10.1111/cogs.12028

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

LaVoie, N. N., Bourne, L. E., Jr., & 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

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

Louie, K., Glimcher, P. W., & Webb, R. (2015). Adaptive neural coding: From biological to behavioral decision-making. *Current Opinion in Behavioral Sciences*, *5*, 91–99. https://doi.org/10.1016/j.cobeha.2015.08.008

Monteiro, C. A., Cannon, G., Lawrence, M., Costa Louzada, M. L., & Pereira Machado, P. (2019). *Ultra-processed foods, diet quality, and health using the NOVA classification system*. Food and Agriculture Organization of the United Nations. https://openknowledge.fao.org/server/api/core/bitstreams/5277b379-0acb-4d97-a6a3-602774104629/content

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

Pohl, R. F. (1998). The effects of feedback source and plausibility of hindsight bias. *European Journal of Cognitive Psychology*, *10*(2), 191–212. 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. (2022). *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (3rd ed.). Routledge.

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, *67*(1), 49–58. https://doi.org/10.1006/obhd.1996.0064

Resnick, I., Newcombe, N. S., & Shipley, T. F. (2017). Dealing with big numbers: Representation and understanding of magnitudes outside of human experience. *Cognitive Science*, *41*(4), 1020–1041. https://doi.org/10.1111/cogs.12388

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

Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasishth, S. (2023). Workflow techniques for the robust use of Bayes factors. *Psychological Methods*, *28*(6), 1404–1426. https://doi.org/10.1037/met0000472.supp

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

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

Appendices

Appendix A. Materials

Table A1. Food Items Used in the Experiment

Subcategory Food item	Serving/Purchase size (g or ml)	Sugar content (in g)	Item set
Grains and grain products			
1 soft pretzel	90	7	A
1 potato	110	1	A
1 pre-packaged pizza dough	400	4	A
1 rusk	10	1	A
1 croissant	70	4	B
1 slice of toast	25	1	B
1 serving of cooked spaghetti	125	4	B
1 wrap	65	4	B
1 slice of farmhouse bread	55	1	C
1 sweet potato	320	13	C
1 serving of cooked rice	125	0	C
1 bread roll	60	1	C
Vegetables			
1 can of corn	140	11	A
1 jar of artichokes	165	1	A
1 kohlrabi	265	10	A
1 onion	90	6	A
1 jar of bell peppers	165	9	B
1 jar of white asparagus	115	2	B
1 cucumber	420	8	B
1 carrot	60	4	B
1 can of kidney beans	265	9	C
1 jar of pickled gherkins	185	8	C
1 bell pepper	150	8	C
1 zucchini	225	4	C
Fruit			
1 can of pineapple slices	275	33	A
1 kiwi	85	8	A
1 orange	170	13	A
1 serving of dried apple slices	30	17	A
1 pear	205	18	B
1 can of peach halves	250	28	B
1 tangerine	90	8	B
1 serving of raisins	30	20	B
1 banana	110	17	C
5 dates	40	22	C
1 can of apricot halves	240	18	C
1 plum	75	7	C
Milk and dairy products			
1 cup of vegan coconut yogurt	160	2	A
1 tablespoon of cream cheese	20	1	A
1 glass of full fat milk	200	9	A
1 mozzarella	125	1	A
1 cup of low-fat quark cheese	250	9	B
1 glass of buttermilk	200	8	B

Subcategory Food item	Serving/Purchase size (g or ml)	Sugar content (in g)	Item set
1 oven-baked cheese	180	1	B
1 slice of butter cheese	30	0	B
1 cup of plain yogurt	150	7	C
1 package of feta cheese	200	0	C
1 glass of oat milk	200	10	C
1 slice of Gouda cheese	30	0	C
Fish, meat, and sausage			
1 serving of smoked salmon	25	0	A
1 serving of sausage salad	230	6	A
1 slice of cooked ham	20	0	A
8 vegetarian chicken nuggets	180	0	A
1 can of tuna	80	0	B
1 Fleischkaese[a]	220	1	B
4 German meat patties	200	3	B
1 vegetarian sausage	50	0	B
5 fish fingers	150	1	C
1 slice of salami	10	0	C
2 vegetarian cordon bleu	200	1	C
2 Wiener sausages	100	1	C
Oils and fats			
1 avocado	150	1	A
1 can of coconut milk	400	8	A
1 tablespoon of butter	20	0	A
1 serving of almonds	30	1	A
1 cup of sour cream	200	7	B
1 can of black olives	85	0	B
1 tablespoon of mayonnaise	15	0	B
1 serving of pistachios	30	2	B
1 cup of cream	200	6	C
1 tablespoon of hummus	20	0	C
1 tablespoon of sunflower oil	5	0	C
1 serving of peanuts	30	2	C
Drinks			
1 glass of energy drink	200	22	A
1 glass of carrot juice	200	12	A
1 glass of rhubarb spritzer	200	14	A
1 glass of lemonade	200	16	A
1 glass of apple spritzer	200	13	B
1 glass of coke	200	21	B
1 glass of peach iced tea	200	10	B
1 glass of orange juice	200	18	B
1 glass of bitter lemon	200	24	C
1 glass of ginger ale	200	18	C
1 glass of tomato juice	200	6	C
1 glass of grape spritzer	200	16	C
Sweets and snacks			
1 cup of vanilla pudding	150	17	A

Subcategory Food item	Serving/Purchase size (g or ml)	Sugar content (in g)	Item set
1 tablespoon of strawberry jam	20	11	A
1 serving of coated peanuts	30	2	A
1 serving of sweet popcorn	30	8	A
1 tablespoon of honey	20	15	B
1 serving of crisps	30	1	B
1 chocolate doughnut	60	8	B
1 serving of gummy bears	30	14	B
1 cookie	20	6	C
1 tablespoon of nougat hazelnut spread	20	8	C
1 raspberry jam doughnut	95	29	C
1 serving of pretzels	30	0	C

Note. German meat loaf specialty.

Table A2. Sets of Foods (Sugar Content in Brackets) and Set of Cities (Longitude in Brackets) Used in the Experiment

Set A		Set B		Set C		City Set[a]	
8 vegetarian chicken nuggets	(0)	1 vegetarian sausage	(0)	1 slice of salami	(0)	Bordeaux	(-1)
1 serving of smoked salmon	(0)	1 can of tuna	(0)	1 slice of Gouda cheese	(0)	Alicante	(0)
1 slice of cooked ham	(0)	1 slice of butter cheese	(0)	1 package of feta cheese	(0)	London	(0)
1 tablespoon of butter	(0)	1 can of black olives	(0)	1 serving of cooked rice	(0)	Le Havre	(0)
1 jar of artichokes	(1)	1 tablespoon of mayonnaise	(0)	1 tablespoon of hummus	(0)	Cambridge	(0)
1 tablespoon of cream cheese	(1)	1 piece of Fleischkaese	(1)	1 tablespoon of sunflower oil	(0)	Rouen	(0)
1 mozzarella	(1)	1 oven-baked cheese	(1)	1 serving of pretzels	(0)	Tarragona	(1)
1 rusk	(1)	1 serving of crisps	(1)	2 vegetarian cordon bleu	(1)	Norwich	(1)
1 potato	(1)	1 slice of toast	(1)	5 fish fingers	(1)	Ibiza	(1)
1 avocado	(1)	1 jar of white asparagus	(2)	2 Wiener sausages	(1)	Toulouse	(1)
1 serving of almonds	(1)	1 serving of pistachios	(2)	1 bread roll	(1)	Andorra la Vella	(2)
1 cup of vegan coconut yogurt	(2)	4 German meat patties	(3)	1 slice of farmhouse bread	(1)	Barcelona	(2)
1 serving of coated peanuts	(2)	1 wrap	(4)	1 serving of peanuts	(2)	Lille	(3)
1 pre-packaged pizza dough	(4)	1 carrot	(4)	1 zucchini	(4)	Brussels	(4)
1 onion	(6)	1 croissant	(4)	1 cup of cream	(6)	Bergen	(5)
1 serving of sausage salad	(6)	1 serving of cooked spaghetti	(4)	1 glass of tomato juice	(6)	Marseilles	(5)
1 soft pretzel	(7)	1 cup of sour cream	(7)	1 cookie	(6)	Cologne	(7)
1 kiwi	(8)	1 tangerine	(8)	1 plum	(7)	Basel	(8)
1 can of coconut milk	(8)	1 glass of buttermilk	(8)	1 cup of plain yogurt	(7)	Turin	(8)
1 serving of sweet popcorn	(8)	1 chocolate doughnut	(8)	1 jar of pickled gherkins	(8)	Zurich	(8)
1 glass of full fat milk	(9)	1 cucumber	(8)	1 bell pepper	(8)	Genoa	(9)
1 kohlrabi	(10)	1 cup of low-fat quark cheese	(9)	1 tablespoon of nougat hazelnut spread	(8)	Milan	(9)
1 can of corn	(11)	1 jar of bell peppers	(9)	1 can of kidney beans	(9)	Hamburg	(10)
1 tablespoon of strawberry jam	(11)	1 glass of peach iced tea	(10)	1 glass of oat milk	(10)	Oslo	(11)
1 glass of carrot juice	(12)	1 glass of apple spritzer	(13)	1 sweet potato	(13)	Salzburg	(13)
1 orange	(13)	1 serving of gummy bears	(14)	1 glass of grape spritzer	(16)	Prague	(14)
1 glass of rhubarb spritzer	(14)	1 tablespoon of honey	(15)	1 banana	(17)	Zagreb	(16)
1 glass of lemonade	(16)	1 pear	(18)	1 can of apricot halves	(18)	Bratislava	(17)
1 serving of dried apple slices	(17)	1 glass of orange juice	(18)	1 glass of ginger ale	(18)	Stockholm	(18)
1 cup of vanilla pudding	(17)	1 serving of raisins	(20)	5 dates	(22)	Budapest	(19)
1 glass of energy drink	(22)	1 glass of coke	(21)	1 glass of bitter lemon	(24)	Thessaloniki	(23)
1 can of pineapple slices	(33)	1 can of peach halves	(28)	1 raspberry jam doughnut	(29)	Kiev	(31)

Note. [a]Cities were selected such that their longitudes approximately matched the mean and range of sugar contents across the three sets. The values presented to the participants in the irrelevant-information group (i.e., sugar content relabeled as longitudes) therefore only approximated the actual longitudes.

Appendix B. Prior Specification and Sensitivity Analyses

As recommended by Schad et al. (2023), we verified that our prior specifications produced plausible data by running prior predictive checks. Further, to facilitate specification of the priors, we mean-centered the criterion variable $|\Delta|$.

We specified the following prior distributions. For the intercept parameter, we specified a normal distribution `normal(0,15)`. For the slope parameters, we defined two different priors to examine how these different specifications would affect the results of the analysis. One was "skeptical" with regard to a potential effect, placing a lot of prior probability around zero, `normal(0,10)`. The other prior was "weakly informative," `normal(-5,10)`, with more probability mass around an effect with a negative sign, indicating that the deviation $|\Delta|$ decreases from the OJ task to the other tasks (ROJ and NJ), but still with probability mass around zero allowing for unanticipated effects in the opposite direction. For the standard deviations of the random effects (i.e., participants, items), we specified half-normal distributions, `normal(0,15)` with values > 0 as priors. For the correlation of the random effects, we defined a Lewandowski-Kurowicka-Joe (LKJ) prior distribution with the prior parameter η of 2. For the residual standard deviation, we specified a half-normal distribution, `normal(0,15)` with values > 0.

For comparison, we report the results for both the "weakly informative" prior (reported in the main text) and the "skeptical" prior in Table B1. As shown, while the different priors led to slightly different Bayes Factors, the general conclusions and result patterns remained the same.

Table B1. Results with Skeptical Prior [and Weakly Informative Prior] on the Slope Parameter

	Control	Feedback	Domain	Irrelevant	IA Control × Irrelevant	IA Feedback × Domain
	BF_{10}	BF_{10}	BF_{10}	BF_{10}	BF_{10}	BF_{10}
Transfer learning effect	0.20 [0.19]	>10,000 [>10,000]	>10,000 [>10,000]	0.08 [0.08]	0.42 [0.37]	13 [11]
Hindsight effect	0.06 [0.05]	280 [268]	76 [80]	0.25 [0.23]	0.29 [0.25]	0.11 [0.07]

Note. Control = Control group, Feedback = Feedback group, Domain = Domain-information group, Irrelevant = Irrelevant-information group. IA = Interaction. OJ = Original Judgment, ROJ = Recall of Original Judgment, NJ = New Judgment.

Appendix C. Mapping Knowledge

As in Groß et al. (2023), in addition to metric knowledge, we also report results for another component of numerical domain knowledge, namely, mapping knowledge (i.e., knowledge of the ordering of objects in a domain; Brown & Siegler, 1993, 1996). Mapping knowledge is typically quantified by the rank-order correlation between estimates and actual values.

As in the original study, we expected that there would be no improvement in the rank-order correlation for the NJ relative to the OJ task (i.e., no transfer learning effect), as mapping knowledge usually only improves for the objects for which actual values are presented (Brown, 2002; Brown & Siegler, 1993; Groß et al., 2024). For the same reason, we expected to observe an improvement in the rank-order correlation for the ROJ relative to OJ task (i.e., a hindsight effect) in the feedback group only, as only this group receives actual values for the previously estimated objects.

For the analyses of mapping knowledge, we Fisher (r-to-z) transformed and then mean-centered the rank-order correlation. Unlike in the analyses of metric knowledge, we included cases of perfectly accurate judgment (i.e., OJ = actual value) and perfect OJ reproduction (i.e., ROJ = OJ), as they do not distort the computation of person-level rank-order correlations.

As in the original study, we used Bayesian linear-mixed effects models. We defined the rank-order correlation of each participant j as the criterion of the models. As fixed effects, we specified the task and further included random intercepts and random slopes for participants. We specified the following priors. For the intercept and slope parameters, we used `normal(0,0.5)`; for the standard deviation of the random effects (i.e., intercept and slope for participants) and the residual standard deviation, we used half-normal distributions with values > 0, `normal(0,0.5)`; for the correlations of the random effects, we used an LKJ prior distribution with $\eta = 2$. We compared a model including the fixed-effect predictor task (OJ versus NJ for transfer learning effects; OJ versus ROJ for hindsight effects) to a baseline model without the predictor.

Tables C1 and C2 present the results. The analyses of transfer learning effects yielded evidence for improved mapping knowledge in both the feedback and the domain-information group. While these findings are inconsistent with those of Groß et al. (2023), who did not find any transfer learning effects for mapping knowledge, they are in line with some previous research on seeding effects (Bröder et al., 2023; Brown & Siegler, 1996). As expected, the analyses of hindsight effects showed improved mapping knowledge in the feedback group. In addition, and in line with the observed transfer learning effect in the domain-information group, we also found a hindsight effect in the domain-information group.

While this finding deviates from the findings of Groß et al. (2023), it aligns well with the knowledge-updating account of hindsight bias, which predicts that if information elicits knowledge updating, as indicated by transfer learning effects, hindsight effects should co-occur. As expected, we found no hindsight or transfer learning effects for mapping knowledge in the control or irrelevant-information groups.

In sum, while some results deviated from the original study, the general pattern of results remained in line with the predictions of the knowledge-updating account of hindsight bias, with transfer learning effects consistently co-occurring with hindsight effects.

Table C1. Rank-Order Correlations Between Estimated and Actual Values

	Control		Feedback		Domain		Irrelevant	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
OJ Task	0.61	(0.13)	0.63	(0.10)	0.60	(0.14)	0.60	(0.14)
ROJ Task	0.62	(0.15)	0.70	(0.11)	0.63	(0.14)	0.60	(0.13)
NJ Task	0.64	(0.13)	0.69	(0.12)	0.65	(0.14)	0.61	(0.14)

Note. Shown are the means and standard deviations of the rank-order correlations by group and task. Control = Control group. Feedback = Feedback group. Domain = Domain-information group. Irrelevant = Irrelevant-information group. OJ = Original Judgment, ROJ = Recall of Original Judgment, NJ = New Judgment.

Table C2. Improvements in Mapping Knowledge

	Control	Feedback	Domain	Irrelevant
	BF_{10}	BF_{10}	BF_{10}	BF_{10}
Transfer learning effect (OJ vs. NJ)	0.70	3	8	0.05
Hindsight effect (OJ vs. ROJ)	0.08	>10,000	5	0.03

Note. Control = Control group. Feedback = Feedback group. Domain = Domain-information group. Irrelevant = Irrelevant-information group. OJ = Original Judgment, ROJ = Recall of Original Judgment, NJ = New Judgment.

Appendix D. Investigating Anchoring Effects in the Irrelevant-Information Group

In the main article, we reported analyses examining whether presenting actual values framed as irrelevant information (i.e., longitudes of European cities) elicited a hindsight effect. From the finding that no such effect occurred, we concluded that hindsight judgments were not influenced by anchoring-and-adjustment processes. This conclusion is based on the assumption that processes of anchoring and adjustment might be elicited by the presentation of the full set of relabeled actual values. However, it is possible that exposure to the presented values might have influenced the ROJs via anchoring-and-adjustment processes in other ways. Specifically, participants could have been influenced by the central tendency (i.e., the mean or median) of the presented values, or by the last items presented. To explore these possibilities, we calculated the mean and median $|\Delta|$—the absolute difference between each estimate and the mean or median of the actual values presented during the information phase—for both the OJ and the ROJ task. We specified the same priors as for the main analyses, and chose the *weakly informative* prior for the slope parameter (see Appendix B). We applied the same analytical steps as in the main analysis; that is, we compared a full model that included the main effect of task to a baseline model that did not. The results indicated that the ROJs were not influenced by either the mean ($BF_{10} = 0.20$) or the median ($BF_{10} = 0.32$) of the presented values. Additionally, we tested for recency effects by calculating the recency mean and median $|\Delta|$ for the last five values presented to each participant. Again, no effects on the ROJs were found, for either the recency mean $|\Delta|$ ($BF_{10} = 0.15$) or the recency median $|\Delta|$ ($BF_{10} = 0.16$).

Appendix E. Replicability of Study Results

Table E1. Detailed Comparison of Effects in the Replication and Original Study (Groß et al., 2023, Exp. 2) Regarding their Bayes Factors and Replication Assessment

Effect	Original BF_{10}	Replication BF_{10}	Replication Assessment
Transfer Learning Effects			
ME Feedback	> 10,000	> 10,000	Replicated
ME Domain	> 10,000	> 10,000	Replicated
ME Irrelevant	0.07	0.08	Replicated
ME Control	0.01	0.19	Replicated
IA Feedback*Domain	< 0.01	11	Not Replicated
IA Irrelevant*Control	< 0.01[a]	0.37	Not Replicated
Hindsight Effects			
ME Feedback	> 10,000[a]	268	Replicated
ME Domain	88	80	Replicated
ME Irrelevant	< 0.01	0.23	Replicated
ME Control	0.01	0.25	Replicated
IA Feedback*Domain	< 0.01	0.07	Replicated
IA Irrelevant*Control	<0.01	0.05	Replicated

Note. Overview of the results of the analyses of the main effects (ME) and interaction effects (IA) in both the replication study and the original study, with their associated Bayes Factors (BF), as well as an evaluation of whether the effects observed in the replication study replicate those reported in the original study. Replication = The present replication study, Original = Original study (Experiment 2 in Groß et al. (2023)). Replicated = Effect replicated; Bayes Factors in both studies support the same conclusion (BF > 3 for an effect, or BF $> 1/3$ for no effect), Not Replicated = Effect not replicated; Bayes Factors provide evidence for opposing conclusions or provide only weak evidence (BF $> 1/3$ and < 3). ME = Main effect of task (OJ vs. ROJ/NJ) for each group, IA = Interaction between task (OJ vs. ROJ/NJ) and group. OJ = Original Judgment, ROJ = Recall of Original Judgment, NJ = New Judgment. Feedback = Feedback group, Domain = Domain-information group, Irrelevant = Irrelevant-information group, Control = Control group.

[a] Not reported in Groß et al. (2023); computed based on data available on https://osf.io/va2jf/.

Supplementary Materials

Peer Review Communication

Download: https://collabra.scholasticahq.com/article/147499-hindsight-bias-through-knowledge-updating-a-conceptual-replication-of-gross-et-al-2023/attachment/310925.docx?auth_token=P1k9BjlQI16MEr-TTh3n

Cover Letter

Download: https://collabra.scholasticahq.com/article/147499-hindsight-bias-through-knowledge-updating-a-conceptual-replication-of-gross-et-al-2023/attachment/310926.pdf?auth_token=P1k9BjlQI16MEr-TTh3n