An Inductive Learning Intervention to Improve News Veracity Discernment

Ariana Modirrousta-Galian[1,2], Tina Seabrooke[2], Yaniv Hanoch[3], Nicholas J. Kelley[2], Philip A. Higham[2]

[1]Department of Experimental Psychology, University College London, UK

[2]School of Psychology, University of Southampton, UK

[3]Business School, University of Wolverhampton, UK

Author Note

Ariana Modirrousta-Galian https://orcid.org/0000-0003-2925-2976

Tina Seabrooke https://orcid.org/0000-0002-4119-8389

Yaniv Hanoch https://orcid.org/0000-0001-9453-4588

Nicholas J. Kelley https://orcid.org/0000-0003-2256-0597

Philip A. Higham https://orcid.org/0000-0001-6087-7224

The data, analytic code, and materials needed to replicate this research are available on OSF (https://doi.org/10.17605/OSF.IO/5URX7). This research was preregistered (https://aspredicted.org/SHR_R5S; https://aspredicted.org/SLH_Y5S; https://aspredicted.org/34P_FQP). We have no conflicts of interest to disclose. Our work was supported by the Economic and Social Research Council South Coast Doctoral Training Partnership. The funding source had no other involvement other than financial support. We thank Chloe Lam for her assistance with survey development in Experiment 1.

Correspondence concerning this article should be addressed to Ariana Modirrousta-Galian, Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H OAP, United Kingdom. Email: a.modirrousta-galian@ucl.ac.uk

Abstract

Across three preregistered experiments (total *N* = 1,135), we tested whether an inductive learning (IL) intervention improved participants' news veracity discernment (i.e., ability to distinguish between true and false news). IL involves learning categories by observing or classifying exemplars. Therefore, in this research, IL involved observing true and false news headlines and classifying them as either "true" or "false", with immediate feedback on accuracy. In Experiment 1 (*N* = 214), the IL intervention significantly improved participants' news veracity discernment compared to control, but the Bayesian evidence was only anecdotal. In Experiment 2 (*N* = 483), we incorporated game-design elements into the IL intervention, including performance-contingent badges. Unexpectedly, the effect of the IL intervention decreased. We reasoned that, because the IL intervention involved easy-to-hard classification training, the provision of performance-contingent badges inadvertently made participants more aware of their declining performance. This awareness may have undermined their motivation to learn as the training progressed. Therefore, in Experiment 3 (*N* = 438), we implemented hard-to-easy classification training instead, and the IL intervention significantly improved participants' news veracity discernment, now with strong Bayesian evidence.

Keywords: inductive learning, category learning, misinformation, fake news, news headlines

Public Significance Statement

Inductive learning involves learning new concepts or categories by being exposed to exemplars of those concepts or categories. Previous research has shown that inductive learning can improve people's ability to distinguish between a wide variety of categories, such as different painting styles, bird families, French pronunciations, and even malignant versus benign skin lesions. This paper is the first to extend these findings to the domain of news media, showing that inductive learning can improve people's ability to distinguish between true and false news headlines.

An Inductive Learning Intervention to Improve News Veracity Discernment

On July 29, 2024, a tragic mass stabbing took place in Southport, England, resulting in the deaths of three young girls and injuries to eight children and two adults (Culley & Khalil, 2024). Although the attacker was a British citizen born in Cardiff, Wales, to Rwandan parents, false claims that he was a Muslim asylum seeker spread online (Lindsay & Grewar, 2024). These baseless rumors fueled violent race riots across England, during which rioters attacked hotels housing asylum seekers, mosques, people of color, and police officers (Kimathi, 2024). This incident is one of many that demonstrates the potentially dangerous consequences that can arise when people fail to distinguish between true and false news.

To mitigate the harmful effects of misinformation, researchers have developed various interventions designed to improve people's news veracity discernment. These interventions include providing media literacy tips for spotting false news (Guess et al., 2020; Hoes et al., 2024), administering training protocols based on fact-checkers' knowledge for detecting misinformation (Soetekouw & Angelopoulos, 2024), and teaching people about manipulation techniques that are thought to be common to false news (Roozenbeek & van der Linden, 2022; Roozenbeek et al., 2022). Many of these interventions operate under the assumption that equipping people with explicit knowledge about true and false news will improve their news veracity discernment. However, is such explicit instruction necessary? That is, can people learn to tell apart true and false news without researchers telling them how to do so?

We consider this question pertinent for three reasons. First, misinformation is constantly evolving in terms of content (e.g., to capitalize on trending topics), format (e.g., appearing as videos on social media rather than just traditional articles), and deceptive strategies (e.g., crafting false news so that it shares linguistic features with true news; Kuntur et al., 2025; Thakar & Bhatt, 2024). It is therefore possible for interventions to become obsolete when their focus is to provide specific markers of true and false news. Moreover, updating these interventions can be a lengthy process; researchers need to identify the up-to-date statistical markers of true and false news, establish how best to incorporate such

markers into an intervention, and test whether the intervention is successful (the last two steps may be iterative). Thus, interventions that focus on explicitly teaching people about the markers of true and false news may struggle to adapt quickly enough to the evolving misinformation landscape.

Second, the tips being provided by interventions may not be consistently discriminative of false news despite being presented as such. For example, consider the abovementioned interventions that teach people about manipulation techniques that are thought to be common to misinformation (Roozenbeek & van der Linden, 2022; Roozenbeek et al., 2022). These are called inoculation interventions, and one such manipulation technique taught by them is emotional language. However, as Pennycook et al. (in press) noted, although research suggests that false news generally contains more negative sentiment than true news (Carrasco-Farré, 2022), emotional language and news veracity are "certainly far from perfectly correlated" (p. 28). Indeed, Aslam et al. (2020) analyzed 141,208 news headlines about COVID-19 from reputable news sources (e.g., BBC and Reuters) and found that 52% evoked negative sentiment, 30% evoked positive sentiment, and only 18% were neutral. Consequently, presenting these manipulation techniques as "misinformation techniques" (Roozenbeek & van der Linden, 2022, p. 572) when they can be found in both true and false news might explain why some inoculation interventions have been shown to make people skeptical of all news with little or no effect on news veracity discernment (Graham et al., 2023; Modirrousta-Galian & Higham, 2023).

Third, people are generally good at discriminating between true and false news without training (Modirrousta-Galian et al., 2023, 2025; Pfänder & Altay, 2025). Therefore, it is worth exploring whether the knowledge provided by interventions is consistent with the knowledge participants already typically use to discern the veracity of news. Conceivably, building upon pre-existing knowledge may be more effective than encouraging the use of unfamiliar or counterintuitive rules. In one study, Modirrousta-Galian et al. (2025) asked 327 U.S. adults to rate the veracity of various true and false news headlines. After each veracity rating, participants indicated the decision strategy they used: guess, intuition, familiarity,

prior knowledge, rule, or other. Participants reported using describable rules only 6% of the time, even though their overall discrimination of true and false news headlines was well above chance. Therefore, providing explicit guidance to help people improve their news veracity discernment may have limited success as it requires them to use specific rules that they rarely report using in the first place.

In light of these considerations, we turn to inductive learning, which involves learning concepts or categories via exposure to multiple different exemplars of those concepts or categories (Kornell & Vaughn, 2018). This type of learning is fundamental to the human experience, from prehistoric hunter-gatherers learning to classify food as edible or poisonous, to modern-day radiologists learning to classify tumors as benign or malignant. Accordingly, learning to group similar objects into categories and being able to apply this knowledge to new instances is sometimes essential for our survival. Given the importance of inductive learning, a considerable amount of research has examined the different categories people can learn via exposure to exemplars. In this paper, we focus on two topical categories that have received relatively little attention with respect to inductive learning: true and false news. We use the terms true and false news to refer to true/accurate and false/inaccurate information (also known as misinformation; American Psychological Association, 2024), respectively.[1]

We created an inductive learning intervention to determine whether people can improve their news veracity discernment by observing and classifying examples of true and false news. This intervention did not involve any explicit guidance for discerning the veracity of information. It is noteworthy that a previous attempt at creating an inductive learning intervention to improve news veracity discernment was not successful (Modirrousta-Galian et al., 2023). However, this earlier effort consisted of a single experiment with less inductive

[1] Notably, there are various definitions of misinformation. Some definitions refer to misinformation simply as false, inaccurate, or incorrect information, as we have done in this paper (Adams et al., 2023; American Psychological Association, 2024; van der Meer & Jin, 2020). Others include misleading content, allowing misinformation to encompass half-truths or information that is mostly true but communicated deceptively (Allen et al., 2020; de Ridder, 2021; Southwell et al., 2022). Some definitions also consider intentionality, differentiating between misinformation that is intended to mislead and that which is not, the former being commonly referred to as disinformation (American Psychological Association, 2024; Fetzer, 2004; Søe, 2021).

training, and although typically efficacious category learning techniques were used to enhance the inductive learning process, factors that moderate the effectiveness of such techniques were not considered.[2] In contrast, this paper consists of three experiments (total *N* = 1,135) and, as will be demonstrated in subsequent sections, we considered the moderating factors of category learning techniques when designing the inductive learning intervention.

Overview of the Inductive Learning Intervention

The design of our inductive learning intervention was informed by the category learning literature. Specifically, we incorporated various category learning techniques to improve people's ability to distinguish between true and false news. There are important factors that moderate the effectiveness of category learning techniques. However, these moderating factors have not been investigated in the context of true and false news, so we employed multiple category learning techniques shown to work across different settings (see Table 1). These techniques included blocked observational learning to promote within-category comparisons and interleaved classification training to promote between-category comparisons. We also incorporated additional features in the intervention that have been shown to boost learning in the literature, including easy-to-hard and hard-to-easy scheduling of exemplars, category-level metacognitive judgments of learning (JOLs), learning context, and gamification.

[2] Modirrousta-Galian et al.'s (2023) inductive learning intervention included interleaving, classification training, and gamification (three out of the eight features listed in Table 1). The training used 42 headlines (compared to 64 in the current experiments), and the gamification was more rudimentary, with fixed achievement badges being awarded to all players regardless of their performance (rather than performance-contingent badges as in Experiments 2 and 3).

Table 1

Glossary for All the Key Features of the Inductive Learning Intervention

Key feature	Explanation
Category learning technique	
Blocking	A type of exemplar sequencing in which exemplars are grouped by category
Interleaving	A type of exemplar sequencing in which exemplars are intermixed across categories
Observational learning	A type of learning task involving the observation of labelled exemplars
Classification training	A type of learning task involving the assignment of category labels to exemplars
Additional feature	
Scheduling of exemplars	The ordering of exemplars during learning in terms of difficulty (e.g., easy to hard or hard to easy)
Category-level metacognitive JOLs	Predictions of future classification performance for novel exemplars of each learned category
Learning context	The provision of real-world examples to contextualize learning
Gamification	The use of game design elements in non-game contexts

Note. JOL = judgment of learning. The phrases “type of exemplar sequencing” and “type of learning task” are underlined to highlight the two classes of category learning techniques described in the text.

Category Learning Techniques and Their Moderators

In the current work, we focused on two widely studied classes of category learning techniques. The first is exemplar sequencing, which entails blocking (i.e., grouping exemplars by category) or interleaving (i.e., intermixing exemplars across categories). The second is type of learning task, which entails passive techniques (e.g., observing labelled exemplars, termed observational learning) or active techniques (e.g., assigning category labels to exemplars, termed classification training).

Exemplar Sequencing

Previous research has investigated whether interleaving or blocking is more beneficial for category learning (Carpenter & Mueller, 2013; Carvalho & Goldstone, 2014a, 2014b, 2015; Kornell & Bjork, 2008; Taylor & Rohrer, 2010; Wahlheim et al., 2011). In some cases, interleaving has proved superior to blocking, and two notable hypotheses have been put forward to explain this finding (Guzman-Munoz, 2017; Kornell & Bjork, 2008; Taylor &

Rohrer, 2010). The discriminative-contrast hypothesis argues that interleaving allows for comparisons between exemplars of different categories and therefore highlights their differences, promoting discrimination. In contrast, the spacing hypothesis argues that interleaving causes instances of the same category to be spread over time during learning (also known as spacing), which is known to improve memory (e.g., Higham et al., 2023).

Kang and Pashler (2012) provided evidence for the discriminative-contrast hypothesis over the spacing hypothesis. In Experiment 1, they showed that spreading blocked learning over time yielded worse performance than spreading interleaved learning over time, suggesting that the interleaving benefit could not be due to spacing. In Experiment 2, they showed that interleaving, but not blocking, was equally as effective as simultaneously presenting exemplars from different categories, suggesting that the interleaving benefit stems from promoting between-category comparisons.

However, the discriminative-contrast hypothesis cannot explain cases where blocking proved superior or equal to interleaving (Carpenter & Mueller, 2013; Carvalho & Goldstone, 2014a, 2014b, 2015). Therefore, Carvalho and Goldstone (2014a, 2014b, 2015) put forward the attentional-bias hypothesis, which argues that learners compare the current exemplar being studied to the previous exemplar. According to this hypothesis, if the two exemplars are from the same category, the learner's attention will focus on their similarities, but if the two exemplars belong to different categories, the learner's attention will focus on their differences. Thus, interleaving is thought to promote between-category comparisons, whereas blocking is thought to promote within-category comparisons. The effectiveness of blocking or interleaving therefore depends on the nature of the task. If classification is difficult because of low within-category similarity, blocking will be superior as it will help learners spot the shared features within a category. However, if classification is difficult because of high between-category similarity, interleaving will be superior as it will help learners spot the unique features between categories.

Type of Learning Task

Carvalho and Goldstone (2015) suggested that observational learning (i.e., where participants simply observe exemplars that have been labelled for them) promotes within-category comparisons since it requires attention to the features of an exemplar to determine why it is labelled as such. Conversely, they proposed that classification training (i.e., where participants must actively assign category labels to exemplars) promotes between-category comparisons since it requires attention to the diagnostic features of each category. In line with this, Carvalho and Goldstone suggested that observational learning is facilitated by blocking, which further emphasizes within-category comparisons, while classification training is facilitated by interleaving, which further emphasizes between-category comparisons.

As evidenced by the preceding discussion, the effectiveness of a category learning technique depends on the context. Therefore, it is crucial to determine the most effective category learning techniques across different contexts. Two key factors moderate the effectiveness of category learning techniques (Hughes & Thomas, 2021). The first is category similarity, that is, the degree to which exemplars share features in common with each other. The second is category type, that is, the degree to which category membership can be described verbally.

Category Similarity

Intuitively, lower within-category similarity makes learning harder, as it makes it difficult to learn the common features within a category. Likewise, higher between-category similarity also makes learning harder, as it makes it difficult to learn the features that distinguish between the categories.

In accordance with the attentional-bias hypothesis, the effectiveness of a category learning technique depends on the extent to which it focuses a learner's attention towards the most difficult part of the task (Hughes & Thomas, 2021; see Figure S1). Specifically, when task difficulty is driven by low within-category similarity, learning techniques should highlight within-category similarities (e.g., blocked observational learning). Conversely, when task difficulty is driven by high between-category similarity, learning techniques should highlight between-category differences (e.g., interleaved classification training). When task

difficulty is high across both dimensions (low within-category similarity and high between-category similarity), learning techniques should highlight both within-category similarity and between-category differences (e.g., blocked observational learning and interleaved classification learning). However, when task difficulty is low across both dimensions (high within-category similarity and low between-category similarity), the type of learning technique should not matter. There is substantial evidence for this moderating effect of category similarity on the effectiveness of category learning techniques (Carvalho & Goldstone, 2014a, 2014b, 2015; Zulkiply & Burt, 2013b).

Category Type

A common distinction within the category learning literature is between rule-based categories, which can easily be described verbally, and information-integration categories, which are difficult or impossible to describe verbally (Ashby et al., 1998; Ashby & Gott, 1988). For information-integration categories, multiple features from an exemplar must be examined holistically before it can be categorized, and this complex process typically cannot be described with verbal rules. This dual-process perspective is in accordance with the COmpetition between Verbal and Implicit Systems (COVIS) model, one of the most influential accounts of category learning, that assumes distinct explicit (rule-based) and implicit (information-integration) categorization systems (Ashby et al., 1998; Ashby et al., 2012). Note, however, that several studies have disputed the evidence in support of the COVIS model, casting doubt over its explanatory power (e.g., Edmunds et al., 2015, 2018, 2019; Newell et al., 2011).

The literature on the moderating effect of category type on the effectiveness of category learning techniques is mixed. Do and Thomas (2023) found that interleaving was more effective than blocking for both rule-based and information-integration learning. However, Noh et al. (2016) reported contrasting findings: They found that blocking was more effective for rule-based learning, while interleaving was more effective for information-integration learning. It is worth noting that both studies looked at observational learning, not classification training. Ashby et al. (2002) found that classification training (with feedback)

was more effective than observational learning for information-integration learning, but that the type of learning task had little effect on rule-based learning. In contrast, Edmunds et al. (2015) found that classification training was more effective than observational training for both information-integration and rule-based learning.

Inductive Learning of True and False News

True and false news are real, complex stimuli. In contrast, most of the aforementioned research on category similarity and category type used simple, artificial stimuli (e.g., sine wave gratings).[3] Therefore, the extent to which the results from this prior work will generalize to true and false news headlines is unknown. It is also difficult to determine the category type and the degree of category similarity of true and false news since comparable stimuli have rarely been examined in this context. Indeed, Hughes and Thomas (2021) encouraged future work to develop methods to measure the category similarity of real, complex stimuli, and urged research on category type to extend beyond simple, artificial stimuli (e.g., see Nosofsky & McDaniel, 2019). They also proposed that categories, especially complex ones, might not be binary, and that it is conceivable for them to be a mixture of both information-integration and rule-based categories.

Overview of the Experiments

Since it is difficult to determine the category similarity and category type of true and false news, we adopted a "kitchen sink" approach, employing various category learning techniques shown to work across different contexts. Across three experiments, we tested a novel inductive learning intervention that entailed blocked observational learning followed by interleaved classification training, which has been shown to improve category learning compared to either observational learning or classification training alone (Thai et al., 2015). Furthermore, the intervention involved category-level metacognitive judgments of learning (JOLs) that asked about the likelihood of correctly identifying true and false news headlines

[3] Additionally, previous research has typically focused on visual stimuli, whereas true and false news are arguably more conceptual in nature. We consider the distinction between visual and conceptual categories in more detail in the General Discussion.

in a later test. This type of judgment has been shown to enhance inductive learning (Cruz & Minda, 2024; Lee & Ha, 2019). These three features (blocked observational learning, interleaved classification training, and category-level metacognitive JOLs) remained part of the inductive learning intervention across the three experiments, but several others were changed or added to identify the most effective combination of features for improving news veracity discernment relative to a control. We discuss these changes and additions in more detail later. See Figure S2 for screenshots of key intervention features, including observational learning and classification training, for which we believe visual depiction aids understanding.

Experiment 1

In Experiment 1, we examined whether a novel inductive learning intervention improved participants' news veracity discernment compared to a no-treatment control. This intervention involved four key features that have been shown to boost learning: both observational learning and classification training (Carvalho & Goldstone, 2015; Thai et al., 2015), an easy-to-hard schedule (Jiménez-Sánchez et al., 2022; Roads et al., 2018), both blocked and interleaved sequencing (Carvalho & Goldstone, 2014a, 2014b, 2015), and category-level metacognitive JOLs (Cruz & Minda, 2024; Lee & Ha, 2019).

Method

Transparency and Openness

All data, analytic code, and materials needed to replicate this study are available on OSF (https://doi.org/10.17605/OSF.IO/5URX7). This study was preregistered (https://aspredicted.org/SHR_R5S). We obtained ethical approval to conduct this study from the University of Southampton Ethics Committee (65104.A3). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. Data were analyzed using R (Version 4.5.2).

Participants

Participants were recruited from Prolific (https://www.prolific.co/), an online data collection platform. We used Prolific because it provides a large pool from which to recruit

U.S. participants and has been shown to produce high data quality (Peer et al., 2022). We excluded four participants for straight lining (Stosic et al., 2024; see the "Deviations From the Preregistration" section), two participants for withdrawing consent, and three participants for timing out. The final sample consisted of 214 participants (104 males, 109 females, one who specified "prefer not to say", $M_{\text{age in years}}$ = 41.91, $SD_{\text{age in years}}$ = 12.54), 104 in the treatment condition and 110 in the control condition. The sample size was based on an a priori power analysis computed in G*Power (Faul et al., 2009) that indicated 210 participants were required to detect a medium-sized effect with a two-tailed independent samples *t*-test ($n \approx 210$, $d = 0.50$, $1-\beta = .95$, $\alpha = .05$). Given the limited existing data on inductive learning of true and false news, we chose a medium-sized effect ($d = .50$) to balance funding feasibility with practical significance for the first in a series of experiments. We applied several prescreening options on Prolific to select a sample of participants located in the United States, balanced in terms of sex, between the ages of 18–65 years,[4] fluent in English, with a Prolific approval rate between 90%–100%, and who had not taken part in our previous misinformation studies on Prolific. Participants were paid at a rate of £6.00 per hour. See Table S1 for the sample demographics in Experiment 1. Data were collected in 2023.

Materials

We used Chen et al.'s (2023, Study 3) set of 94 news headlines (47 true and 47 false), all of which consisted of a headline and an image. Our item selection method (see Method S1) resulted in 32 news headlines (16 true and 16 false) ordered from easy to hard for the first training block, 32 different news headlines (16 true and 16 false) ordered from easy to hard for the second training block, and 30 different, randomly selected news headlines (15 true and 15 false) for the test block. We omitted source information from all headlines to prevent participants from relying on this feature as a shortcut to making decisions about their veracity (e.g., rating headlines from mainstream news sources as true

[4] Despite setting this age range, one participant in the control condition indicated that they were 66 years old when asked at the end of the study. A 1-year difference in the upper age limit of the study caused by a single participant is unlikely to impact the results, so we decided to include this participant in the analysis.

and rating headlines from unknown news sources as false).[5] We also omitted subheadings from all headlines because they would introduce an additional informational cue beyond the two already present (headline and image), which are the focus of this study.

Design

A two-group experimental design, consisting of a treatment group and a control group, was used. The independent variables were the between-subjects condition that participants took part in (treatment vs. control) and news headline type (true vs. false), manipulated within-subjects. The primary dependent variable of interest was participants' veracity ratings of the 30 news headlines presented in the test block. The secondary dependent variables of interest were: (a) participants' classification performance for the 32 news headlines presented in the second training block of the treatment condition (see Results S1); (b) participants' eight category-level metacognitive JOLs made throughout the two training blocks of the treatment condition (see Results S1); and (c) participants' three commonalities and differences statements (one about the commonalities between true news, another about the commonalities between false news, and a third about the differences between true and false news) throughout the two training blocks of the treatment condition (see Results S4). The demographic variables of interest were age, gender, and political orientation since previous studies have found them to be associated with news veracity discernment (Calvillo et al., 2020; Calvillo et al., 2021; Modirrousta-Galian et al., 2023).

Procedure

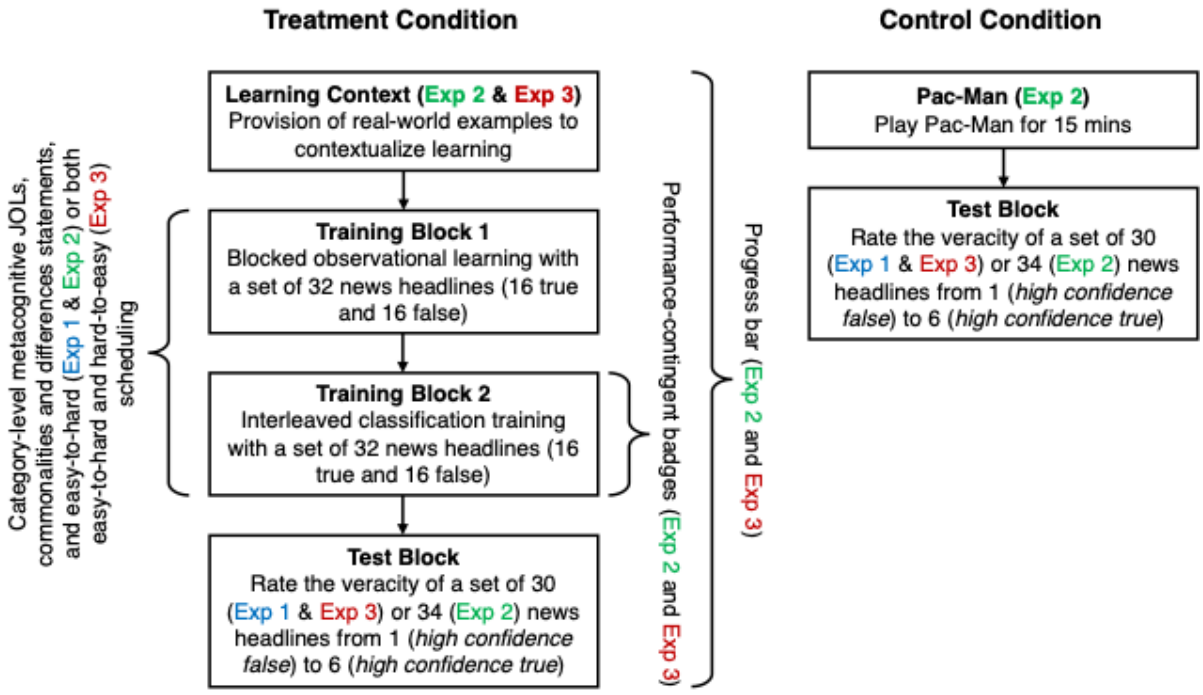
See Figure 1 for an overview of the procedure of each experiment. We applied device restrictions on Prolific, which recommended that participants access the experiment through a computer. Since this was a remote online study, participants could use any web

[5] Including source information could mean that participants simply learn which news sources are reliable and which are not, rather than learning to distinguish between true and false headlines based on their content (which is our aim). Improving veracity judgments based on headline content rather than source may be a more ambitious yet generalizable approach, as such improvements should apply regardless of whether source information is present or absent. In contrast, source-based learning would likely only be relevant when the source is present, which is not always the case in real-world situations, such as on social media where content is frequently shared without its original source.

browser or computer of their choosing. To navigate through the experiment, a mouse (or touchpad) and keyboard were necessary.

Figure 1

Overview of the Procedure of Experiments 1, 2, and 3



Note. JOL = judgment of learning. If individual experiments are not specified in parentheses immediately after a feature of the procedure is mentioned, it indicates that the feature was present in all experiments.

Participants were first shown a combined information sheet and consent form. To indicate that they had read and understood the form, were aged 18 or over, and agreed to take part in the study, participants clicked a button at the bottom of the web page. After this, they automatically entered full screen mode and were shown the following warning: “To minimize distractions, you have now entered full screen mode. Please do not exit full screen mode unless absolutely necessary. If you exit full screen mode too many times, we will not be able to use your data for our study”. If a participant exited full screen mode, they automatically re-entered full screen mode whenever they pressed the “next” button.

Participants were then asked for their Prolific IDs, which they provided by typing or pasting it into a textbox.

Subsequently, participants were given instructions for the task that followed, which depended on the condition they took part in. In the control condition, participants were immediately presented with the test block. This involved rating the veracity of 30 news headlines (15 true and 15 false) on a 6-point Likert scale from 1 (*high confidence false*) to 6 (*high confidence true*). The news headlines were presented individually and in a random order, and participants did not receive feedback on the accuracy of their ratings.

In the treatment condition, participants were initially presented with the first training block. This involved observational learning where participants were presented with 32 news headlines (16 true and 16 false) that were accurately labelled as true or false. The news headlines were presented individually, in an easy-to-hard order, and in a blocked sequence (eight true, eight false, eight true, eight false). Participants had to view each headline for a minimum of 2 s before being able to move on to the next one. After every eight news headlines, participants provided two category-level metacognitive JOLs: “What is the likelihood that you will be able to correctly indicate whether a news headline is true in a later test?” and “What is the likelihood that you will be able to correctly indicate whether a news headline is false in a later test?”. These two JOLs were presented in a random order, and participants responded on a slider scale from 0–100. At the end of the first training block, participants were shown all the true news headlines from that block and asked to type their commonalities. The same procedure was then repeated for the false news headlines. These two tasks were presented in a random order.

Participants then took part in the second training block, which was the same as the first except for four differences. First, a different set of 32 news headlines (16 true and 16 false) was used, although they were also presented individually and in an easy-to-hard order. Second, headlines were presented in an interleaved rather than a blocked sequence. We determined the interleaved sequence by using Random.org (https://www.random.org/) to systematically randomize the order of a list made up of 16 true and 16 false news headlines

until no more than three headlines of the same type appeared consecutively. The pre-determined interleaved sequence was then used for all participants. Third, classification training was implemented where participants classified each news headline by clicking either a “true” or a “false” button, and they were given feedback on whether their classifications were correct after each headline. The order of the “true” and “false” response options was randomized for each question to avoid potential response order effects. Lastly, at the end of the second training block, participants were shown all the true and false news headlines from that block and asked to type any differences they noticed between them. Following the second training block, participants took part in the test block, which was identical to that of the control condition.

After the test block, participants in both conditions were asked for their age (typed answer), gender (“male”, “female”, “other”, “prefer not to say”), and political orientation (1 = *very left-wing* to 7 = *very right-wing*). Finally, participants received a written debriefing.

Deviations From the Preregistration

Four participants exhibited straight lining in the test block, giving all items either “true” (4, 5, or 6) or “false” (1, 2, or 3) ratings. We consequently excluded these four participants from the analysis. Two false news headlines from the test block showed a floor effect, receiving a mean veracity rating of 1 in the control condition. One true news headline from the test block closely resembled a true news headline from the first training block. Both reported the same event (“Rep. Jeff Fortenberry resigns after being found guilty of lying to FBI” and “Nebraska congressman to resign after being found guilty of lying to FBI”) and featured an image of the same individual. We consequently removed these three test block news headlines from the analysis. These non-preregistered participant and item exclusions did not affect the overall results.

Regarding the sampling design, we randomly chose which condition (treatment or control) to run first. Once our desired sample size took part, we closed that condition and collected data for the remaining one, excluding participants who had previously participated. This sampling design deviated from the preregistration, where we stated that participants

would be randomly assigned to one of the two conditions. This deviation occurred because Prolific does not currently support random assignment of participants to conditions that have different completion times and payments. There was only a 1-day delay in data collection between conditions, and three Pearson's chi-squared tests showed no significant differences between the two groups in terms of age, $\chi^2(4, N = 214) = 8.45$, $p = .077$, gender, $\chi^2(2, N = 214) = 1.19$, $p = .552$, or political orientation, $\chi^2(6, N = 214) = 3.28$, $p = .773$.

Lastly, we conducted non-preregistered analyses that are clearly labelled as such in the manuscript and in the Supplemental Materials (see Results S1 and S4).

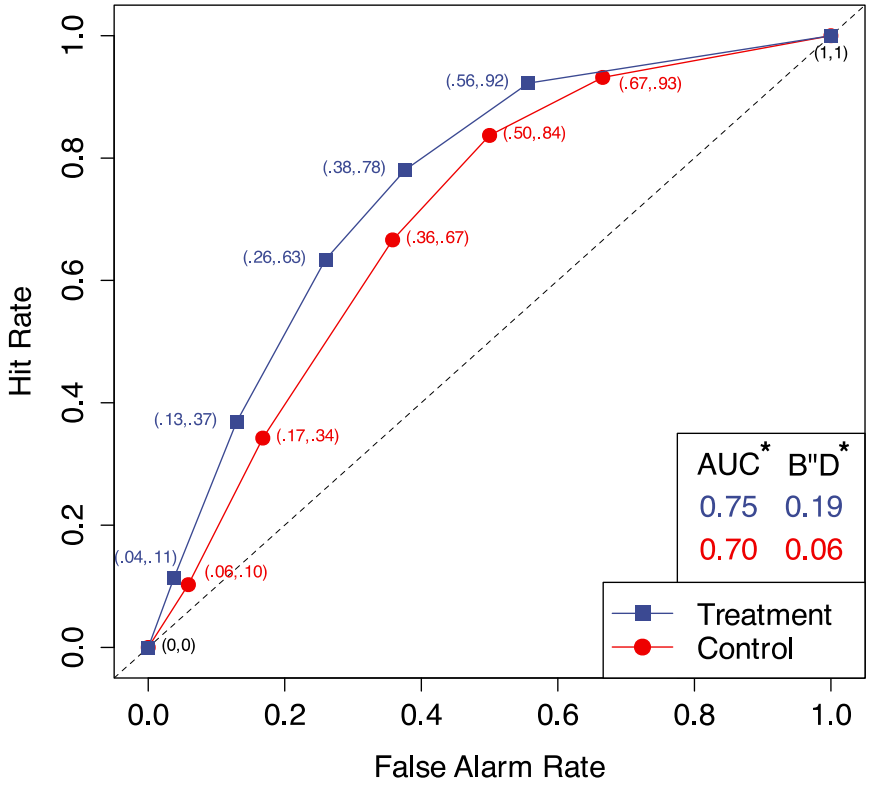
Results

Receiver Operating Characteristic (ROC) Analysis

Receiver operating characteristic (ROC) analysis allows for discrimination and response bias to be measured separately (Higham & Higham, 2019). In the context of this paper, discrimination refers to the ability to distinguish between true and false news (i.e., news veracity discernment), while response bias refers to the overall tendency to rate all news as true or false. To separate these two distinct aspects of performance, we conducted ROC analysis on participants' veracity ratings in the test block (see Figure 2).

Figure 2

ROC Curves for the Treatment Condition and Control Condition in Experiment 1



Note. ROC = receiver operating characteristic. * indicates that the intervention had a significant effect ($p < .05$) on the measure (either AUC [discrimination] or *B"D* [response bias]). The false alarm rate and hit rate, in that order, are shown beside each point.

ROC curves are useful for visualizing discrimination and response bias. Chance-level discrimination is used as a reference point and corresponds to the straight diagonal line drawn from the bottom-left corner to the top-right corner of the ROC space. The further the ROC curve bows away from the diagonal towards the top-left portion of the ROC space, the better the discrimination. ROC points that cluster towards the bottom-left versus top-right portion of the ROC curve represent conservative (overall tendency to rate all news as false) versus liberal (overall tendency to rate all news as true) responding, respectively.

For each participant, we calculated the area under the curve (AUC; a measure of discrimination) using the trapezoidal rule (Pollack & Hsieh, 1969), as well as *B"D* (a measure of response bias; Donaldson, 1992), using loglinear-corrected hit and false alarm rates

(Stanislaw & Todorov, 1999), for each point on the ROC curve. The AUC ranges from 0 to 1, with .50 representing chance-level discrimination and 1 representing perfect discrimination. *B″D* ranges from −1 to 1, with −1 representing an extreme liberal response bias, 0 representing no response bias, and 1 representing an extreme conservative response bias. We conducted Welch's independent-samples *t*-tests to compare the mean AUC values between the treatment and control conditions, as well as the mean *B″D* values (collapsed over ROC points) between the two conditions. For detailed discussions on signal detection theory (SDT), ROC analysis, AUC, *B″D*, and why these methods and measures are optimal for misinformation research, see Modirrousta-Galian and Higham (2023) and Higham et al. (2024).

The ROC curves for the treatment condition and control condition are shown in Figure 2. A Welch's independent-samples *t*-test revealed that the AUC values in the treatment condition (M = .75, SD = .16) were significantly greater than the AUC values in the control condition (M = .70, SD = .16), with anecdotal Bayesian evidence for the alternative hypothesis, $t(210.38) = 2.12$, $p = .035$, 95% CI [0.00, 0.09], $d = 0.29$, $BF_{10} = 1.22$. A Welch's independent-samples *t*-test revealed that the *B″D* values in the treatment condition (M = .19, SD = .30) were significantly greater than the *B″D* values in the control condition (M = .06, SD = .31), with strong Bayesian evidence for the alternative hypothesis, $t(212) = 3.20$, $p = .002$, 95% CI [0.05, 0.21], $d = 0.44$, $BF_{10} = 16.78$. As shown in Figure 2, the false alarm rates (FARs; the proportion of false news participants inaccurately categorized as true) in the treatment condition were considerably smaller than those in the control condition, whereas the hit rates (HRs; the proportion of true news participants accurately categorized as true) barely differed across the two conditions. This pattern of results indicates that, overall, there were lower veracity ratings in the treatment condition than in the control condition, which led to the difference in *B″D* values.

Linear Mixed-Effects Models

To examine demographic effects, we ran linear mixed-effects models with the lme4 R package (Version 1.1.37; Bates et al., 2015). We did not have strong predictions for the

interactions, so we adopted an exploratory model-building strategy to obtain the best-fitting and most parsimonious model. To do this, we used a step-up strategy to build the fixed effects structure first and the random effects structure second, and then a step-down strategy to trim the fixed-effects structure (for more information on step-up and step-down strategies, see Martínez-Huertas et al., 2022; Ryoo, 2011; Thrane et al., 2018; West et al., 2007). We started with an intercept-only model that included veracity ratings as the dependent variable, headline type (true vs. false) and condition (treatment vs. control) as fixed effects (with an interaction term), and participant number and headline number as random effects. Then, we forward fitted the model by adding terms individually. We tested whether the addition of terms significantly increased the model fit with likelihood ratio tests. Following Pinheiro and Bates (2000) and Yan et al. (2014), we used maximum likelihood estimates when comparing models with different fixed effects structures, and restricted maximum likelihood estimates when comparing models with different random effects structures. Terms that significantly improved the model's fit ($p < .05$) without introducing convergence or singularity issues were retained.

First, we added political orientation (mean-centered), age (mean-centered), and gender as fixed effects, in that order. This order was informed by research suggesting that political orientation has the greatest influence on belief in false news, followed by age and then gender (Gupta et al., 2023). Only political orientation improved the model fit. We then added all possible two-way interactions, and then all possible three-way interactions. Only the two-way interaction between political orientation and headline type improved the model fit. Once we had the maximal fixed effects structure, we added all possible random slopes for the fixed effects included in the model. The random slope for headline type, which was always included first, was added to the participant random structure. The random slopes for condition and political orientation were added, in that order, to the headline random structure. Since political orientation was a continuous variable, it was first added as an uncorrelated random slope and then as a correlated random slope.

The model that converged and resulted in better fit compared to the intercept-only model included the slope for headline type in the participant random structure as well as the slope for condition and the uncorrelated slope for political orientation in the headline random structure. Once we had the maximal fixed and random effects structures, we planned to backward fit the model by removing terms individually from the fixed effects structure, starting with the highest order non-significant interactions. However, all the highest order interactions were significant, and all the main effects were both significant and qualified by higher order interactions. Therefore, we did not remove any terms, and this marked the end of our model building process. All the variance inflation factors (VIFs) of the final model were smaller than five, which suggests that there was no multicollinearity among predictor variables (Shrestha, 2020).

The results from the final model are shown in Table 2. The first significant highest order interaction was the two-way interaction between headline type and condition (see Panel A of Figure 3). This interaction showed that there was no statistically significant difference in veracity ratings for true news between conditions (pairwise comparison: p = .544), but participants in the treatment condition gave significantly lower veracity ratings for false news than participants in the control condition (pairwise comparison: p = .004). The second significant highest order interaction was the two-way interaction between headline type and political orientation (see Panel B of Figure 3). This interaction showed that there was no statistically significant difference in veracity ratings for true news across participants with differing political orientations (pairwise comparisons: ps > .05), but more right-wing participants gave significantly higher veracity ratings for false news (pairwise comparisons: ps < .001; see the Note of Figure 3 for the levels specified for political orientation).

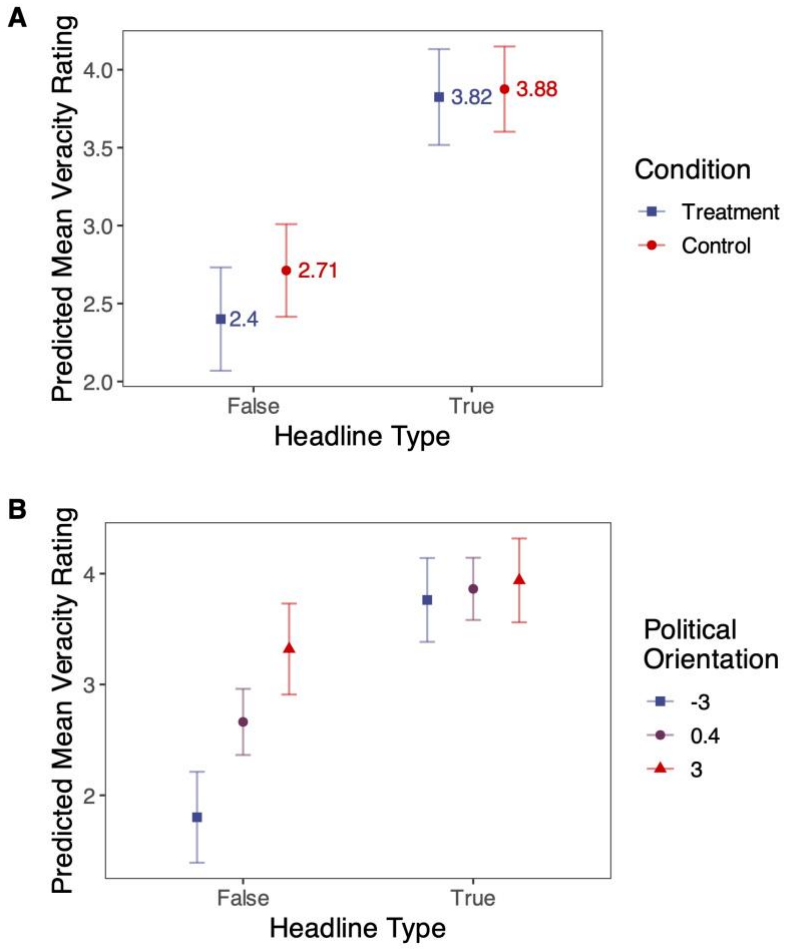
Table 2

Results From the Final Linear Mixed-Effects Model in Experiment 1

Fixed effects	*b*	*SE*	*t*	*df*	*p*
Intercept	2.71	0.15	17.91	37.46	<.001
type [true]	1.16	0.20	5.72	32.92	<.001
condition [treatment]	−0.31	0.11	−2.90	154.53	.004
political orientation	0.25	0.05	5.23	46.70	<.001
type [true] × condition [treatment]	0.26	0.13	2.04	122.03	.044
type [true] × political orientation	−0.22	0.06	−3.50	38.64	.001

Figure 3

Significant Highest Order Interactions From the Final Linear Mixed-Effects Model in Experiment 1



Note. Panel A depicts the significant two-way interaction between headline type and condition. Panel B depicts the significant two-way interaction between headline type and political orientation. Veracity ratings were made on a scale from 1 (*high confidence false*) to 6 (*high confidence true*). Political orientation was measured on a scale from 1 (*very left-wing*) to 7 (*very right-wing*). Political orientation was mean centered and, for ease of interpretation, represented by three equally spaced values rounded to "nice" numbers over their range (see "xlevels" in the documentation for the effects R package [Version 4.2.4]; https://cran.r-

project.org/web/packages/effects/effects.pdf). These values were −3, 0.4, and 3, where −3 can be interpreted as left-wing, 0.4 as centrist, and 3 as right-wing.

Discussion

Experiment 1 showed that the inductive learning intervention improved participants' news veracity discernment. This enhanced discrimination was primarily driven by better false news detection. We observed this finding with both ROC analysis and linear mixed-effects model analysis. However, the Bayesian evidence for the alternative hypothesis in the ROC analysis was only anecdotal. We also found that more right-wing participants gave higher veracity ratings for false news.

Experiment 2

In Experiment 2, we again examined whether the inductive learning intervention improved participants' news veracity discernment, but this time compared the treatment condition to a time-matched control rather than a no-treatment control. We also attempted to boost the discrimination effect of the intervention by adding two more elements that have been shown to enhance learning: learning context, which involves providing real-world situations that underpin the learning material (Bransford et al., 2000; Prince & Felder, 2006), and gamification, which involves incorporating game design elements (Mazarakis & Bräuer, 2023; Sailer & Homner, 2020).

Method

Transparency and Openness

All data, analytic code, and materials needed to replicate this study are available on OSF (https://doi.org/10.17605/OSF.IO/5URX7). This study was preregistered (https://aspredicted.org/SLH_Y5S).[6] We obtained ethical approval to conduct this study from the University of Southampton Ethics Committee (65104.A4). We report how we determined

[6] The linked preregistration (https://aspredicted.org/SLH_Y5S) is part of a bundle that contains another almost identical preregistration (https://aspredicted.org/582x-34dy.pdf). The almost identical preregistration was based on an earlier version of Experiment 2 that we never collected data for and is therefore obsolete.

our sample size, all data exclusions (if any), all manipulations, and all measures in the study. Data were analyzed using R (Version 4.5.2).

Participants

We excluded three participants for exiting full screen mode more than five times, one participant for straight lining (Stosic et al., 2024; see the "Deviations From the Preregistration" section), one participant for withdrawing consent, and two participants for timing out. The final sample consisted of 483 participants (231 males, 239 females, seven who specified "other", six who specified "prefer not to say", $M_{\text{age in years}}$ = 39.88, $SD_{\text{age in years}}$ = 12.22), 241 in the treatment condition and 242 in the control condition, recruited from Prolific. The sample size was based on an a priori power analysis in G*Power (Faul et al., 2009) that indicated 484 participants were required to detect a small-to-medium-sized effect (d = 0.30, based on the AUC analysis in Experiment 1) with a one-tailed independent samples t-test ($n \approx 484$, d = 0.30, $1-\beta$ = .95, α = .05; see the "Deviations From the Preregistration" section). We applied the same prescreening options on Prolific and paid participants at the same rate as in Experiment 1. See Table S2 for the sample demographics in Experiment 2. Data were collected in 2023.

Materials

For the training blocks, we used the same news headlines as in Experiment 1. For the test block, we used a different set of 373 news headlines (188 true and 185 false) sent to us by Gordon Pennycook. Our item selection method (see Method S2) resulted in 34 news headlines (17 true and 17 false), all of which were in the same format as those from Experiment 1 (i.e., they consisted of a headline and an image with no source information or subheading). Also consistent with Experiment 1, the test block news headlines were not ordered by difficulty; only the training block news headlines were.

Design

As in Experiment 1, a two-group experimental design, consisting of a treatment group and a control group, was used. However, since the conditions were time-matched in Experiment 2, we ran them simultaneously and hence participants were truly randomly

assigned to the conditions. The variables of Experiment 2 were identical to those of Experiment 1 except for two differences. First, the primary dependent variable of interest was participants' veracity ratings of the new set of 34 news headlines presented in the test block of Experiment 2, compared to the old set of 30 news headlines presented in the test block of Experiment 1. Second, an additional secondary dependent variable of interest was introduced in Experiment 2, namely participants' responses to a question at the end of the treatment condition that asked participants whether they focused more on learning individual news headlines (memorization) or on developing a rule for why news headlines were true or false (rule abstraction) during training. This question was adapted from Little and McDaniel (2015), which found that self-reported learning orientation (memorization vs. rule abstraction) predicted responses in a category-learning task comparable to that of Experiments 1 and 2 (but see Results S2).

Procedure

See Figure 1 for an overview of the procedure of each experiment. The procedure of Experiment 2 was identical to that of Experiment 1 except for six differences. First, in the control condition, participants played Pac-Man for 15 min before being presented with the test block to match the completion time to that of the treatment condition. We used Pac-Man because it is a straightforward game that requires minimal explanation, similar to other games (e.g., Tetris) that have been used as controls for gamified misinformation interventions in previous work (e.g., Roozenbeek & van der Linden, 2022). Second, in the treatment condition, before the first training block, participants read an explanation for at least 30 s about why it is important to distinguish between true and false news. This learning context is available on OSF (https://doi.org/10.17605/OSF.IO/5URX7). Third, in the treatment condition, a progress bar was added to incorporate gamification. Fourth, in the treatment condition, after every eight headlines in the second training block, participants were given a total score out of the number of headlines they rated up until that point and a badge depending on that score, also to incorporate gamification. Fifth, in both the treatment and control conditions, the test block involved rating the veracity of a new set of 34 news

headlines. Sixth, in the treatment condition, after the test block, participants were asked: "During the training blocks, were you more focused on trying to learn the individual news headlines, or trying to develop a rule for why news headlines were either true or false?". Participants responded on a 7-point Likert scale from 1 (*solely memorization*) to 7 (*solely rule*).

Deviations From the Preregistration

The preregistered power analysis contained two errors. First, the effect size should have been $d = 0.29$ (obtained from the AUC analysis in Experiment 1), rather than $d = 0.30$. Second, the *t*-test should have been two-tailed rather than one-tailed. We therefore reran the a priori power analysis in G*Power (Faul et al., 2009), this time specifying $d = .29$ rather than .30 and a two-tailed rather than one-tailed independent samples *t*-test. It revealed that with a sample size of 484, we would have .89 power to detect an effect size of $d = 0.29$. Regarding exclusions, one participant was excluded because they exhibited straight lining, giving all items in the test block a rating of 1. Five false news headlines from the test block showed a floor effect, receiving a mean veracity rating of 1 in the control condition. We consequently removed these five false news headlines from the analysis. These non-preregistered participant and item exclusions did not affect the overall results. Lastly, we conducted non-preregistered analyses that are clearly labelled as such in the manuscript and in the Supplemental Materials (see Results S2).

Results

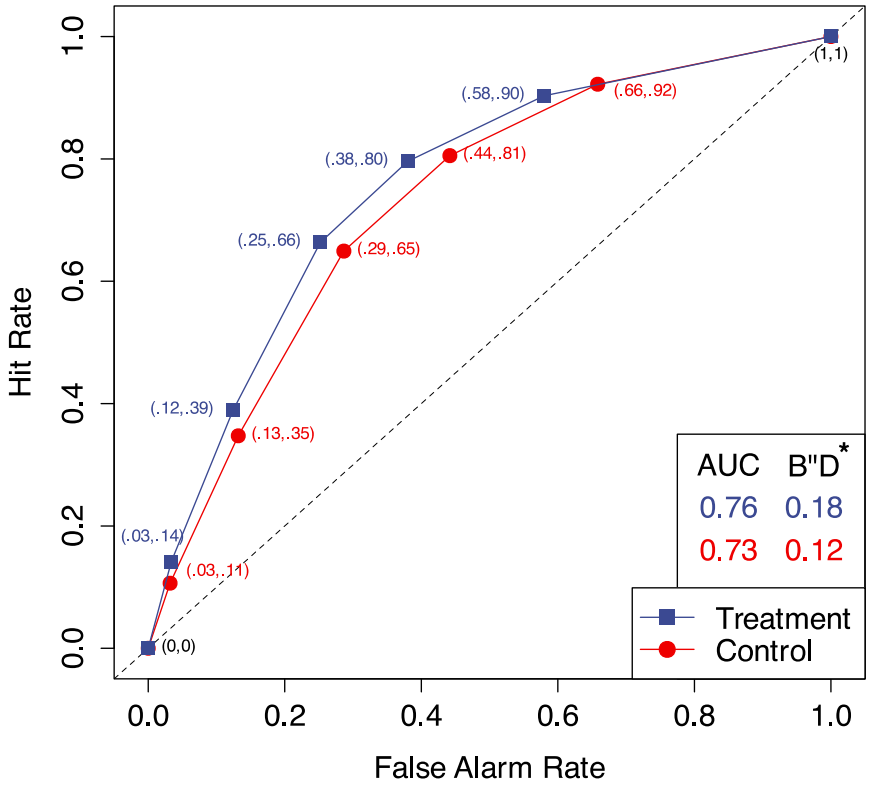
Receiver Operating Characteristic (ROC) Analysis

The ROC curves for the treatment condition and control condition are shown in Figure 4. A Welch's independent-samples *t*-test revealed that the AUC values in the treatment condition ($M = .76$, $SD = .15$) were not significantly different from the AUC values in the control condition ($M = .73$, $SD = .14$), with anecdotal Bayesian evidence for the null hypothesis, $t(478.64) = 1.66$, $p = .098$, 95% CI [-0.00, 0.05], $d = 0.15$, $BF_{10} = 0.38$. A Welch's independent-samples *t*-test revealed that the *B"D* values in the treatment condition ($M = .18$, $SD = .30$) were significantly greater than the *B"D* values in the control condition ($M = .12$, *SD*

= .31), although the Bayesian analysis revealed anecdotal evidence for the null hypothesis, $t(480.70) = 2.04$, $p = .041$, 95% CI [0.00, 0.11], $d = 0.19$, $BF_{10} = 0.76$. As shown in Figure 4, the FARs were smaller in the treatment condition than in the control condition. The opposite was true for the HRs, although the difference in FARs was considerably greater. Therefore, compared to the control condition, the greater *B"D* values in the treatment condition (i.e., more conservative responding) was mostly attributable to lower veracity ratings given to false news (i.e., better false news detection).

Figure 4

ROC Curves for the Treatment Condition and Control Condition in Experiment 2



Note. ROC = receiver operating characteristic. * indicates that the intervention had a significant effect ($p < .05$) on the measure (either AUC [discrimination] or *B"D* [response bias]). The false alarm rate and hit rate, in that order, are shown beside each point.

Linear Mixed-Effects Models

We followed the same exploratory model-building strategy and started with the same initial model as in Experiment 1. Consequently, we forward fitted the fixed effects structure, and the best fitting model included the main effect of political orientation, the two-way interaction between political orientation and headline type, and the three-way interaction between political orientation, headline type, and condition. We then forward fitted the random effects structure, and the best fitting model included the slope for headline type in the participant random structure and the slopes for condition and political orientation in the headline random structure. Once we had the maximal fixed and random effects structures, we planned to backward fit the model by removing the non-significant three-way interaction between political orientation, headline type, and condition. However, this resulted in convergence issues. Therefore, we did not remove any terms, and this marked the end of our model building process. All the VIFs of the final model were smaller than five.

The results from the final model are shown in Table 3. The first significant highest order interaction was the two-way interaction between headline type and condition (see Panel A of Figure 5). This interaction showed that there was no statistically significant difference in veracity ratings for true news between conditions (pairwise comparison: p = .296), but participants in the treatment condition gave significantly lower veracity ratings for false news than participants in the control condition (pairwise comparison: p = .013). The second significant highest order interaction was the two-way interaction between headline type and political orientation (see Panel B of Figure 5). This interaction showed that there was no statistically significant difference in veracity ratings for true news across participants with differing political orientations (pairwise comparisons: ps > .05), but more right-wing participants gave significantly higher veracity ratings for false news (pairwise comparisons: ps < .001; see the Note of Figure 5 for the levels specified for political orientation).

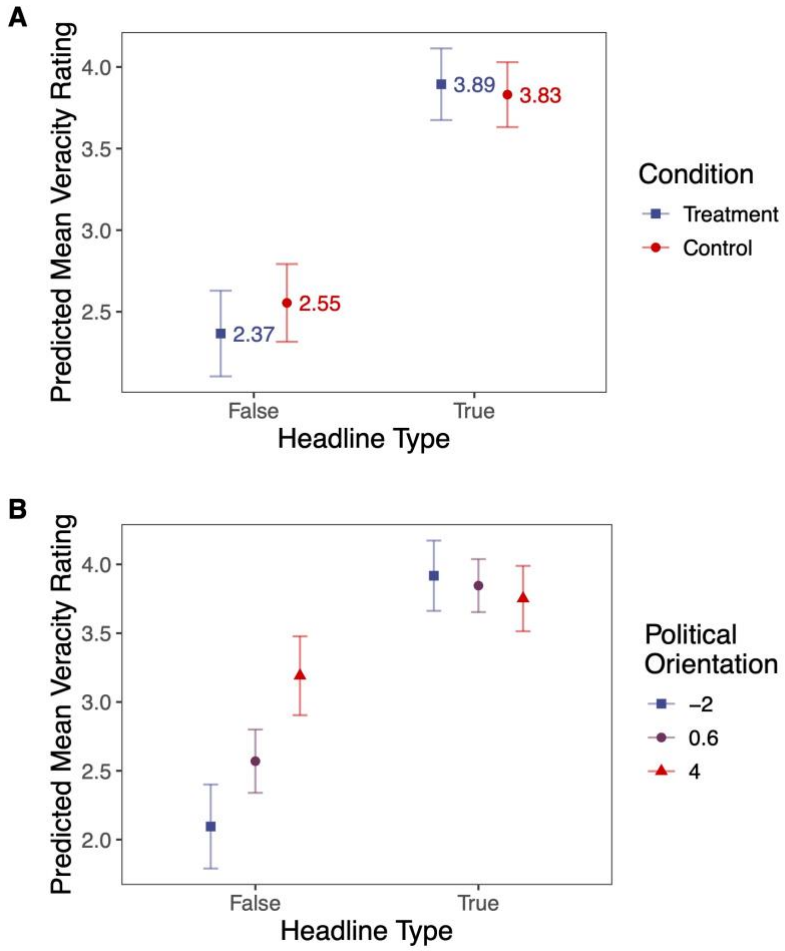
Table 3

Results From the Final Linear Mixed-Effects Model in Experiment 2

Fixed effects	*b*	*SE*	*t*	*df*	*p*
Intercept	2.55	0.12	21.02	35.43	<.001
type [true]	1.28	0.16	8.23	32.25	<.001
condition [treatment]	−0.19	0.08	−2.50	180.05	.013
political orientation	0.18	0.04	4.78	95.83	<.001
type [true] × condition [treatment]	0.25	0.09	2.95	119.19	.004
type [true] × political orientation	−0.18	0.05	−3.93	70.78	<.001
condition [treatment] × political orientation	0.00	0.04	0.00	479.02	.999
type [true] × condition [treatment] × political orientation	−0.06	0.04	−1.40	478.98	.163

Figure 5

Significant Highest Order Interactions From the Final Linear Mixed-Effects Model in Experiment 2



Note. Panel A depicts the significant two-way interaction between headline type and condition. Panel B depicts the significant two-way interaction between headline type and political orientation. Veracity ratings were made on a scale from 1 (*high confidence false*) to 6 (*high confidence true*). Political orientation was measured on a scale from 1 (*very left-wing*) to 7 (*very right-wing*). Political orientation was mean centered and, for ease of interpretation, represented by three equally spaced values rounded to “nice” numbers over their range (see

"xlevels" in the documentation for the effects R package [Version 4.2.4]; https://cran.r-project.org/web/packages/effects/effects.pdf). These values were −2, 0.6, and 4, where −2 can be interpreted as left-wing, 0.6 as centrist, and 4 as right-wing.

Subgroup AUC Analysis (Not Preregistered)

The ROC analysis suggested that treatment participants did not show better discrimination than control participants in the test block. One potential reason for this is that participants' classification performance decreased across the second training block (see Results S2). As evidenced by their category-level metacognitive JOLs (see Results S2), participants were aware of their deteriorating performance, which may have had a demotivating effect on learning. To test this possibility, we reanalyzed the test block data but included only the participants from the treatment condition who did not show a major deterioration in classification performance during the second training block.

Accordingly, we calculated each treatment participant's proportion correct score (i.e., the proportion of news headlines classified correctly) for every eight news headlines in the second training block. Next, we calculated the difference in proportion correct score between the last eight and the first eight news headlines for each participant, which indicated how their performance varied from the beginning to the end of the second training block. Positive scores indicated an improvement across training, while negative scores indicated a deterioration across training. The distribution of proportion correct score differences was: −.625, −.500, −.375, −.250, −.125, .000, .125, .250, and .375. We then excluded participants in the bottom third of proportion correct score differences (−.625, −.500, or −.375), resulting in a subgroup of 165 (out of 241) treatment participants.[7]

Rerunning the Welch's independent-samples *t*-test on AUC values revealed that news veracity discernment in the treatment condition (M = .77, SD = .14) was now significantly better than in the control condition (M = .73, SD = .14), with moderate Bayesian

[7] We originally excluded participants with negative proportion correct score differences, but this resulted in a subgroup of only 21 participants.

evidence for the alternative hypothesis, $t(350.17) = 2.76$, $p = .006$, 95% CI [0.01, 0.07], $d = 0.28$, $BF_{10} = 4.34$. This result suggests that the inductive learning intervention was effective for participants who did not experience a considerable decline in performance during training.

Discussion

Experiment 2 showed that the inductive learning intervention improved participants' news veracity discernment, although the improvement was not apparent in all measures. Consistent with Experiment 1, this enhanced discrimination was driven by better false news detection. Although we observed this finding only descriptively with the ROC analysis, it was statistically significant with linear mixed-effects model analysis. It is noteworthy, however, that the Bayesian evidence for the null hypothesis in the ROC analysis was merely anecdotal. Therefore, the addition of learning context and gamification did not boost the effectiveness of the intervention. We also found that more right-wing participants gave higher veracity ratings for false news, which replicates the analogous result from Experiment 1.

Experiment 2 Follow-Up (Not Preregistered)

We conducted a non-preregistered, exploratory follow-up to Experiment 2, where we examined the longer-term effects of the inductive learning intervention on news veracity discernment and on a different but related discrimination task. Accordingly, we recontacted participants 31 days after Experiment 2 and asked them to take part in a follow-up that consisted of two sections: one focusing on misinformation and the other on fraud. The misinformation section involved the test block from Experiment 2, and the fraud section involved a real and fake website discrimination task from Kelley et al. (2023). The method and results of the Experiment 2 follow-up are available on OSF (https://doi.org/10.17605/OSF.IO/5URX7).

To summarize the key findings, the Experiment 2 follow-up showed that the inductive learning intervention did not improve participants' news veracity discernment after 1 month. This is not surprising considering that the immediate effects of our intervention were weak, and that both the effects of inductive learning and misinformation interventions have been

shown to decay considerably over time (Guess et al., 2020; Zulkiply & Burt, 2013a). Furthermore, participants demonstrated a more liberal response bias in the follow-up compared to Experiment 2. Since the test block was repeated across Experiment 2 and the follow-up with the same news headlines, the illusory truth effect—the finding that repeated exposure to information increases its perceived truth, regardless of its objective veracity (Henderson et al., 2022)—can explain this result (Modirrousta-Galian et al., 2025). We also found that the inductive learning intervention did not have any longer-term effects on a real and fake website discrimination task.

Experiment 3

In Experiment 3, we again examined whether the inductive learning intervention improved participants' news veracity discernment compared to a no-treatment control. We also attempted to boost the discrimination effect of the intervention by implementing two changes to the second training block of the treatment condition (i.e., the classification training). The first change involved applying a hard-to-easy (rather than easy-to-hard) schedule during classification training. The rationale behind this change was that with a hard-to-easy schedule, performance-contingent badges should make participants aware that their performance is increasing throughout training, which should positively impact their motivation to learn. This reasoning stemmed from our supposition that the combination of an easy-to-hard schedule and performance-contingent badges during classification training in Experiment 2 had a negative effect on motivation and may have contributed to the small effects of the intervention in that experiment. The second change involved counterbalancing the order of the "true" and "false" response options so that half of the participants saw "true" above "false" while the other half saw "false" above "true" (rather than randomizing the order for each question). The rationale behind this change was that randomly switching the response locations of button presses (as was done in Experiments 1 and 2) can impair procedural processes, which can in turn interfere with certain types of learning (Ashby et al., 2003; Maddox et al., 2004, 2010; Spiering & Ashby 2008).

Method

Transparency and Openness

All data, analytic code, and materials needed to replicate this study are available on OSF (https://doi.org/10.17605/OSF.IO/5URX7). This study was preregistered (https://aspredicted.org/34P_FQP). We obtained ethical approval to conduct this study from the University of Southampton Ethics Committee (65104.A4). We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. Data were analyzed using R (Version 4.5.2).

Participants

We excluded two participants for exiting full screen mode more than five times, three participants for straight lining (Stosic et al., 2024), one participant for withdrawing consent, and three participants for timing out. The final sample consisted of 438 participants (197 males, 230 females, six who specified "other", five who specified "prefer not to say", $M_{\text{age in years}}$ = 38.88, $SD_{\text{age in years}}$ = 11.16), 219 in the treatment condition and 219 in the control condition, recruited from Prolific. The sample size was based on an a priori power analysis in G*Power (Faul et al., 2009) that indicated 434 participants were required to detect a small-to-medium-sized effect (obtained from the subgroup AUC analysis in Experiment 2) with a two-tailed independent samples *t*-test ($n \approx 434$, $d = 0.27$, $1-\beta = .80$, $\alpha = .05$; see the "Deviations From the Preregistration" section). We applied the same prescreening options on Prolific and paid participants at the same rate as in Experiments 1 and 2. See Table S3 for the sample demographics in Experiment 3. Data were collected in 2024.

Materials

For the training blocks, we used the same news headlines as in Experiments 1 and 2. For the test block, we used the non-excluded news headlines from Experiments 1 and 2, which were 31 true news headlines (14 from Experiment 1 and 17 from Experiment 2) and 26 false news headlines (14 from Experiment 1 and 12 from Experiment 2). However, five false news headlines from the test block showed a floor effect, receiving a mean veracity rating of 1 in the control condition. We consequently removed these five false news headlines from the analysis.

Design

As in Experiments 1 and 2, a two-group experimental design was used consisting of a treatment group and a control group. As in Experiment 2, true random assignment to conditions was employed. We achieved this despite the different completion times (and thus payments) for each condition by first collecting the Prolific IDs of 700 participants. We then randomized the order of these Prolific IDs and invited the first half to take part in the control condition and the second half to take part in the treatment condition. To expedite the data collection process, we collected the Prolific IDs of 791 additional participants and repeated the process.

The variables of Experiment 3 were identical to those of Experiment 2 except for two differences. First, the primary dependent variable of interest was participants' veracity ratings of a different set of 30 news headlines (15 true and 15 false, randomly selected for each participant from a set of 57 news headlines). Second, the memorizer versus rule abstractor question was replaced with a different question that asked participants to rate their familiarity with the news headlines presented in the test block on a 6-point Likert scale from 1 (*not at all familiar*) to 6 (*very familiar*). The memorizer versus rule abstractor question was dropped in this experiment because it was not associated with any other variable in Experiment 2 (see Results S2). The new familiarity question aimed to explore the potential interaction between headline familiarity and the effect of the inductive learning intervention (see Results S3).

Procedure

See Figure 1 for an overview of the procedure of each experiment. The procedure of Experiment 3 was identical to that of Experiment 2 except for four differences. First, in the control condition, participants were presented with the test block but did not play Pac-Man beforehand due to funding constraints. Second, in the second training block of the treatment condition, the order of "true" and "false" response options was counterbalanced so that half of the participants saw "true" above "false" while the other half saw "false" above "true". Third, in both the treatment and control conditions, the test block involved rating the veracity

of a different set of 30 news headlines (15 true and 15 false, randomly selected for each participant from a set of 57 news headlines). Fourth, in the treatment condition, instead of being asked about memorization versus rule-abstraction after the test block, participants were asked: "Overall, how familiar were you with the news headlines in the test phase (had you seen or heard about them before)?". Participants responded on a 6-point Likert scale from 1 (*not at all familiar*) to 6 (*very familiar*).

Deviations From the Preregistration

The preregistered power analysis contained an error: The effect size should have been $d = 0.28$ rather than $d = 0.27$. We therefore reran the a priori power analysis in G*Power (Faul et al., 2009), this time specifying $d = .28$ rather than .27. It revealed that with a sample size of 434 and alpha of .05, we would have .83 power to detect an effect size of $d = 0.28$ with a two-tailed independent samples *t*-test, which is greater than the power we aimed to achieve (.80). Lastly, we conducted non-preregistered analyses that are clearly labelled as such in the manuscript and in the Supplemental Materials (see Results S3).

Results

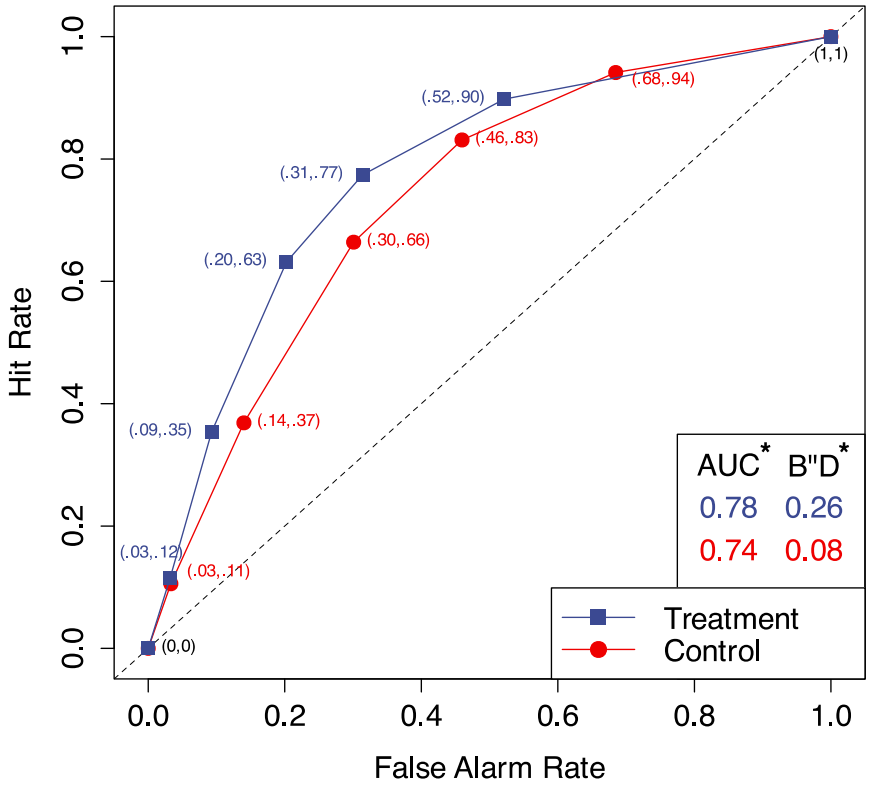
Receiver Operating Characteristic (ROC) Analysis

The ROC curves for the treatment condition and control condition are shown in Figure 6. A Welch's independent-samples *t*-test revealed that the AUC values in the treatment condition ($M = .78$, $SD = .12$) were significantly greater than the AUC values in the control condition ($M = .74$, $SD = .14$), with strong Bayesian evidence for the alternative hypothesis, $t(431.03) = 3.16$, $p = .002$, 95% CI [0.02, 0.06], $d = 0.30$, $BF_{10} = 12.92$. A Welch's independent-samples *t*-test revealed that the *B"D* values in the treatment condition ($M = .26$, $SD = .29$) were significantly greater than the *B"D* values in the control condition ($M = .08$, $SD = .25$), with extreme Bayesian evidence for the alternative hypothesis, $t(428.07) = 7.04$, $p < .001$, 95% CI [0.13, 0.24], $d = 0.67$, $BF_{10} = 960{,}135{,}143$. As shown in Figure 6, both the FARs and HRs were smaller in the treatment condition than in the control condition, but the difference in FARs was considerably greater. Therefore, compared to the control condition, the greater *B"D* values in the treatment condition (i.e., more conservative responding) were

mostly attributable to lower veracity ratings given to false news (i.e., better false news detection) as well as a slightly lower veracity ratings given to true news (i.e., slightly worse true news detection).

Figure 6

ROC Curves for the Treatment Condition and the Control Condition in Experiment 3



Note. ROC = receiver operating characteristic. * indicates that the intervention had a significant effect ($p < .05$) on the measure (either AUC [discrimination] or *B"D* [response bias]). The false alarm rate and hit rate, in that order, are shown beside each point.

Linear Mixed-Effects Models

We followed the same exploratory model-building strategy and started with the same initial model as in Experiments 1 and 2. Consequently, we forward fitted the fixed effects structure, and the best fitting model included the main effect of political orientation, the two-way interaction between political orientation and headline type, and the three-way interaction between political orientation, headline type, and condition. We then forward fitted the

random effects structure, and the best fitting model included the slope for headline type in the participant random structure. Once we had the maximal fixed and random effects structures, we backward fitted the model by removing the non-significant three-way interaction between political orientation, headline type, and condition, which resulted in better fit. This marked the end of our model building process. All the VIFs of the final model were smaller than five.

The results from the final model are shown in Table 4. The first significant highest order interaction was the two-way interaction between headline type and condition (see Panel A of Figure 7). This interaction showed that participants in the treatment condition gave significantly lower veracity ratings for false news (pairwise comparison: $p < .001$) and true news (pairwise comparison: $p = .005$) than participants in the control condition. The second significant highest order interaction was the two-way interaction between headline type and political orientation (see Panel B of Figure 7). This interaction showed that there was no statistically significant difference in veracity ratings for true news across participants with differing political orientations (pairwise comparisons: ps > .05), but more right-wing participants gave significantly higher veracity ratings for false news (pairwise comparisons: ps < .001; see the Note of Figure 7 for the levels specified for political orientation).

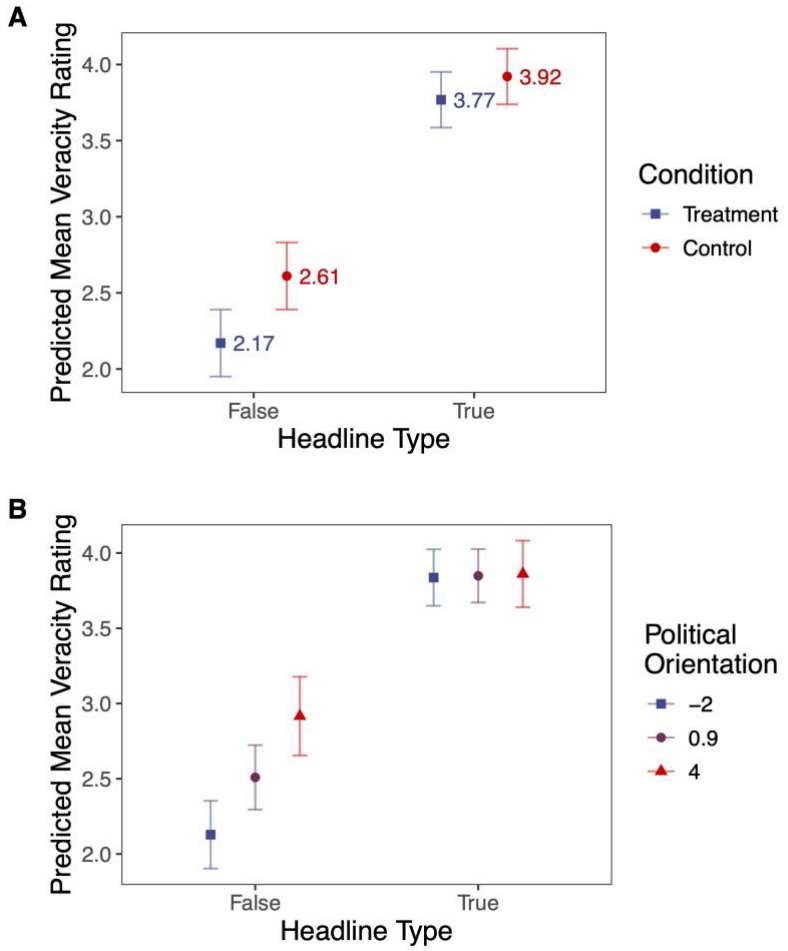
Table 4

Results From the Final Linear Mixed-Effects Model in Experiment 3

Fixed effects	*b*	*SE*	*t*	*df*	*p*
Intercept	2.61	0.11	23.26	67.87	< .001
type [true]	1.31	0.14	9.11	65.23	< .001
condition [treatment]	−0.44	0.06	−7.10	435.89	< .001
political orientation	0.13	0.02	6.67	436.62	< .001
type [true] × condition [treatment]	0.29	0.07	3.84	435.82	< .001
type [true] × political orientation	−0.13	0.02	−5.36	436.24	< .001

Figure 7

Significant Highest Order Interactions From the Final Linear Mixed-Effects Model in Experiment 3



Note. Panel A depicts the significant two-way interaction between headline type and condition. Panel B depicts the significant two-way interaction between headline type and political orientation. Veracity ratings were made on a scale from 1 (*high confidence false*) to 6 (*high confidence true*). Political orientation was measured on a scale from 1 (*very left-wing*) to 7 (*very right-wing*). Political orientation was mean centered and, for ease of interpretation, represented by three equally spaced values rounded to "nice" numbers over their range (see "xlevels" in the documentation for the effects R package [Version 4.2.4]; https://cran.r-project.org/web/packages/effects/effects.pdf). These values were −2, 0.9, and 4, where −2 can be interpreted as left-wing, 0.9 as centrist, and 4 as right-wing.

Discussion

Experiment 3 showed that the inductive learning intervention improved participants' news veracity discernment. We observed this finding with both ROC analysis and linear

mixed-effects model analysis. Moreover, the Bayesian evidence for the alternative hypothesis in the ROC analysis was strong (compared to anecdotal in Experiment 1), which supports the robustness of the intervention's effect. It is important to note that the inductive learning intervention also elicited a slight conservative response bias. However, the decrease in FARs (false news) far outweighed the decrease in HRs (true news). We also found that more right-wing participants gave higher veracity ratings for false news, which replicates the analogous result from Experiments 1 and 2.

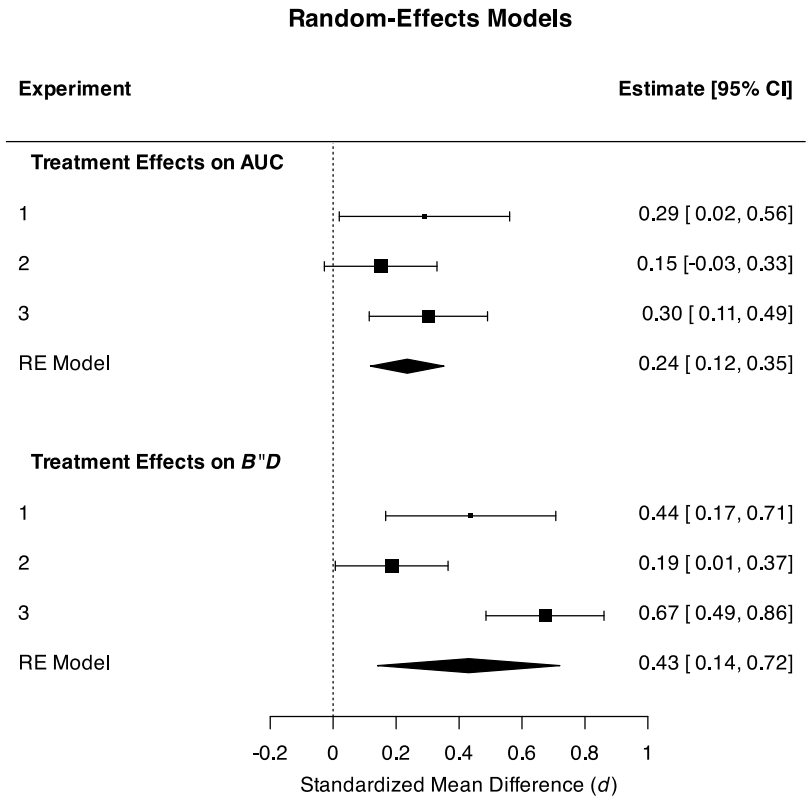
Mini Meta-Analysis

We carried out a mini meta-analysis to determine the overall discrimination and response bias effects of the inductive learning intervention across Experiments 1, 2, and 3 (total N = 1,135). We fitted two random-effects models, one on the treatment-control AUC effect sizes and the other on the treatment-control *B″D* effect sizes. These two models compared the meta-analytic effect size estimate to the null effect size (d = 0). We fitted random-effects models rather than fixed-effects models because of the heterogeneity between our three experiments (e.g., different participant pools, items, and treatment versions; Borenstein et al., 2010). We conducted the meta-analysis in R (Version 4.5.2) with the metafor package (Version 4.8.0; Viechtbauer, 2010).

The results of the mini meta-analysis are illustrated in Figure 8. The random-effects model on treatment-control AUC effect sizes showed significantly greater AUC values in the treatment condition compared to the control (d = 0.24, 95% CI [0.12, 0.35], SE = 0.06, z = 3.95, p < .001). Similarly, the random-effects model on treatment-control *B″D* effect sizes showed significantly greater *B″D* values in the treatment condition compared to the control (d = 0.43, 95% CI [0.14, 0.72], SE = 0.15, z = 2.92, p = .004). These results suggest that there were treatment-based effects on both discrimination and response bias, whereby the inductive learning intervention improved news veracity discernment but also elicited more conservative responding.

Figure 8

Two Random-Effects Models Fitted Separately on Treatment-Control AUC and B"D Effect Sizes



Note. RE = random effects. The error bars represent 95% confidence intervals. The size of the points represents the weight given to each study. The two outcomes are measured on different scales: AUC ranges from 0 to 1, while *B"D* ranges from –1 to 1. The effect size, confidence interval, and standard error for response bias are each approximately double those for discrimination, mirroring the twofold range of its measurement scale.

Supplemental Study

We conducted a preregistered supplemental study, where we examined whether the inductive learning intervention still improved participants' news veracity discernment after removing two key elements: category labels during observational learning and feedback during classification training. Our aim was to test whether exposure to news headlines, without any indication of their veracity, was sufficient to improve discrimination. Our rationale for conducting this supplemental study was that previous work has found discernment improvements from pre-test to post-test even in control conditions, in which participants completed an unrelated task or no task between tests (see Figure 12 from Modirrousta-

Galian & Higham, 2023). Since the headlines in the control conditions in Experiments 1–3 were not exposure-matched to the treatment conditions, we sought to rule out the possibility that mere exposure was driving the observed intervention effects. To test this possibility, we employed a one-group quasi-experimental pre-post design. The intervention followed the procedure as Experiment 3's treatment condition, but we removed key inductive learning elements that allow participants to learn by observing the category labels (either presented with the news items during observational learning or as feedback following classification training). The method and results of the supplemental study are available on OSF (https://doi.org/10.17605/OSF.IO/5URX7).

To summarize the key findings, the supplemental study showed that the stripped-down inductive learning intervention did not improve participants' news veracity discernment. In fact, discrimination was descriptively lower in the posttest compared to the pretest. Furthermore, the intervention elicited more conservative responding. That is, participants gave significantly lower veracity ratings to both true and false news in the posttest compared to the pretest. Together, these results suggest that mere exposure to news headlines is not sufficient to improve discrimination, and they point towards category labels during observational learning and feedback during classification training as potentially critical components of inductive learning, offering a promising direction for future research. It is also worth noting that the discernment improvements in the pre-post control conditions reported by Modirrousta-Galian & Higham (2023) were small, with four out of the five control conditions analyzed containing zero in the 95% CI. Moreover, other factors may have contributed to the null discernment and increased conservative responding effects in this supplemental study, as it was not a pure control. Participants viewed and classified 62 news headlines, albeit with no indication of their veracity, and provided JOLs, which may have prompted greater reflection on task performance and, in turn, increased skepticism.

General Discussion

Key Findings

Taken together, the results of our three experiments indicated that our novel inductive learning intervention improved participants' ability to discriminate between true and false news. This enhanced discrimination was primarily driven by better false news detection. Additionally, in all three experiments, being more right-wing predicted poorer false news detection. This relationship between conservatism and false news susceptibility in U.S. samples is supported by previous work (Garrett & Bond, 2021; Gupta et al., 2023; Modirrousta-Galian et al., 2023, 2025). There is also evidence for this association in other countries, such as Spain (Gómez Calderón et al., 2023) and India (Gupta et al., 2023), suggesting that this may be a widespread phenomenon.

Effect of Inductive Learning on Discrimination and Response Bias

Our inductive learning intervention improved true and false news discrimination but also elicited a tendency for participants to respond conservatively (overall lower ratings of truth). Notably, the risks of not believing true news can sometimes be just as harmful as believing false news. For example, individuals who stormed the U.S. Capitol on January 6, 2021, had not only fallen for false news claiming that the 2020 presidential election was rigged in favor of Joe Biden, but also disregarded true news maintaining that there was no evidence of widespread voter fraud, as confirmed by the U.S. attorney general and Justice Department (Perez & Cole, 2020; Singman, 2020). Therefore, an overall decrease in truth ratings warrants discussion.

Importantly, this tendency for our intervention to cause participants to respond conservatively is qualitatively different from an analogous tendency observed with gamified inoculation interventions (e.g., Bad News and Go Viral!). Modirrousta-Galian and Higham (2023), for example, reanalyzed the available data at the time using ROC analysis and found that gamified inoculation interventions also resulted in conservative responding, a finding that has subsequently been replicated with newer research (Graham et al., 2023). However, the main difference between inoculation interventions and those reported here is that the former interventions had *only* an effect on response bias; there was no reliable effect on discrimination (except in studies that used poor true news items that were at ceiling; e.g.,

Roozenbeek & van der Linden, 2019).[8] Conversely, our interventions produced *both* more conservative responding *and* better discrimination.

This difference is important because it means the underlying models of performance are also likely to differ between the two types of intervention. Specifically, in gamified inoculation studies where an intervention only produces more conservative responding, the likely scenario is either that: (1) there is a comparable reduction in the subjective truth of *both* true and false items while the criteria participants use to judge truth are static (dual distribution shift); or (2) the subjective truth of the news items is unaffected while the criteria used to judge truth shift to a more conservative position (criteria shift). Although both scenarios will result in no change to discrimination and more conservative responding, they have different psychological interpretations (see Modirrousta-Galian & Higham, 2023, pp. 2432–2434). In contrast, the more conservative responding in the current experiments was most likely due to the intervention selectively reducing the subjective truth of false news *only*. Across the experiments, ratings of true news remained relatively impervious to the interventions, while ratings to false news decreased, a data pattern consistent with a *single* distribution shift. Critically, however, if the subjective truth of false news decreased while everything else remained the same, there would be both an increase in discrimination (true and false news become easier to distinguish), and a decrease in the overall magnitude of truth ratings when all the items are considered. This overall reduction in truth ratings would be reflected in the bias measure (*B"D*) as more conservative responding.

Thus, the data suggest that our intervention differed from others in that it had selective effects, targeting belief in false news while leaving belief in true news relatively intact. As has been argued previously (e.g., Higham et al., 2024; Modirrousta-Galian & Higham, 2023; Seabrooke et al., 2026), the specificity of an intervention is a critical factor when determining its efficacy. The fact that our intervention had specific effects also renders

[8] Note, however, that when Bad News was updated to include pre- and post-intervention "feedback exercises"—which appear similar to inductive learning training—whereby participants rate the reliability of various news items and receive feedback, the intervention was found to improve discrimination (Leder et al., 2024).

some potential interpretations of conservative responding insufficient. For example, one explanation of more conservative responding is that simply exposing people to false news primes skepticism, causing truth ratings to decrease (e.g., Clayton et al., 2020; Graham et al., 2023; Modirrousta-Galian & Higham, 2023). As we have argued, the more conservative responding observed in our experiments was not likely due to increased skepticism but was more likely a by-product of our intervention specifically reducing belief in false news.

Effect of Inductive Learning Between Experiments

While it may be tempting to compare results across the three experiments and attribute any differences to the distinct features incorporated in each version of the intervention, this would be conjecture. Although our experiments were similar, they were conducted at different times and on different samples, making it difficult to draw causal conclusions from analyzing them together. For example, it might be tempting to conclude that hard-to-easy scheduling is more effective than easy-to-hard scheduling during classification training for improving discrimination of true and false news, because the intervention was (slightly) more effective in Experiment 3 than in Experiments 1 and 2. However, to causally establish whether such specific features of the intervention affected classification performance and JOLs during training, and subsequently—or consequently—performance at test, it would be necessary to run experiments in which the inductive learning intervention differs in the presence or absence of such features.

Inductive Learning of Visual and Conceptual Categories

The core difference between true and false news headlines lies in the veracity of the information they convey. That is, true news headlines present true information, and false news headlines present false information. Veracity, in this context, is conceptual, as the difference between truth and falsehood is rooted in the underlying concepts or ideas they present. For example, a news headline claiming that vaccines contain trackable microchips is false precisely because the notion it presents is untrue. In light of these considerations, we consider true and false news headlines to be primarily conceptual categories—even though news headlines often include an image, which introduces a visual component.

Most of the inductive learning literature has focused on visual categories, such as painting styles (Kornell & Bjork, 2008), bird species (Wahlheim et al., 2011), or rock types (Do & Thomas, 2023). This includes much of the research covered in the Introduction that served as a basis for the design of our inductive learning intervention. The suitability of using research on visual categories to inform studies on conceptual categories can be called into question. We note, however, that our inductive learning intervention was successful, suggesting at least some overlap between visual and conceptual category learning. In addition, there are some notable exceptions in the literature that also found inductive learning to be successful for non-visual categories. For instance, people have been found capable of inductively learning French words with different pronunciation rules (Carpenter & Mueller, 2013) and scenarios depicting different fallacies or biases (Motz et al., 2023).

In the context of misinformation research, recent work suggests that inductive learning may be particularly effective for visual categories. Seabrooke et al. (2025) created an inductive learning intervention (comparable to that of this paper) designed to improve participants' ability to distinguish between real and artificial-intelligence (AI)-generated faces. Their intervention was successful and produced very large effect sizes (largest AUC effect size: d = 1.82), far exceeding those obtained in this paper (largest AUC effect size: d = 0.30). The difference between real and AI-generated faces lies in specific visual features. For example, AI-generated faces are more likely to be highly proportional than real faces (Miller et al., 2023). Repeated exposure to exemplars in inductive learning may facilitate the detection of these straightforward visual cues. In contrast, conceptual categories likely involve less clear-cut discriminative features and may require higher-level cognitive processes, such as reasoning that is prone to inferential errors and cognitive biases or the application of factual knowledge. Indeed, how does one know that vaccines do not contain trackable microchips? Without at least some knowledge of vaccines, it would be practically impossible to assess whether such information is true or false and inductive learning with other types of news is unlikely to help.

A Note on Effect Sizes

Although the effect size of our intervention on true and false news discrimination was small-to-medium by conventional standards (AUC meta-analytic effect size: d = 0.24), it is numerically comparable to, or even larger than, other interventions reported to improve news veracity discernment. For example, recent work combined psychological inoculation (teaching people about manipulation techniques that are thought to be common to misinformation) and accuracy prompts (drawing people's attention to the concept of accuracy), two of the most popular interventions against misinformation, into a single intervention (Pennycook et al., in press). This combined intervention yielded an average effect size of d = 0.20. Furthermore, Modirrousta-Galian and Higham's (2023) meta-analysis on gamified inoculation interventions reported a treatment effect (relative to control) of d = 0.08. Lu et al. (2023), who included both gamified and non-gamified inoculation interventions in their meta-analysis, reported a treatment effect (relative to control) of d = 0.20. Therefore, it appears that intervention effects on true and false news discrimination are modest at best within the current literature.

Limitations and Future Directions

There are several limitations with the three experiments we report in this paper. First, the average completion time between the treatment condition and the control condition varied in Experiment 1 (21 min vs. 7 min) and Experiment 3 (24 min vs. 6 min). This introduced potential confounds. For example, practice effects, where greater exposure to the task and stimuli facilitates learning, may have contributed more to improved performance in the treatment condition than in the control condition. Conversely, the longer duration of the treatment condition could have induced fatigue or boredom, potentially worsening performance.

To control for these potential effects, we ran a (non-preregistered) linear regression model with AUC values as the dependent variable, condition as the independent variable, and total study completion time (in seconds) as a covariate for each experiment. The only result that changed was in Experiment 2, where the AUC values in the treatment condition were now significantly—not just numerically—greater than those in the control condition (see

Results S4). Notably, Experiment 2 differed from Experiments 1 and 3 in that we time-matched the control condition by having participants play Pac-Man for 15 min. In summary, while we acknowledge the potential effects of time-on-task on performance, these appear minimal and unlikely to account for the observed treatment effects across our experiments.

Another limitation is the potentially limited scalability of our intervention on social media compared to less time-intensive approaches. For example, accuracy prompts, which simply prompt people to consider the accuracy of the information they encounter, can be seamlessly integrated into social media platforms and scaled to reach large audiences (Pennycook et al., 2020). In contrast, our intervention requires individuals to opt in to a 15-min experience, making it less scalable and more likely to draw in a self-selected subgroup of the population. At the same time, scalability considerations may be premature as the evidence we present in this paper is from controlled experiments with limited ecological validity. We tested the inductive learning intervention by asking participants to rate the veracity of individually presented headlines, which is not representative of how people typically engage with news in everyday life. Therefore, before scaling inductive learning interventions, it is pertinent to test them in more naturalistic settings and with more ecologically valid measures, such as post interactions on mock social media feeds (Butler et al., 2024). Moreover, our data do not allow us to identify which features of the inductive learning intervention are most effective or whether their combined effects are additive—information that would be essential for optimizing the training prior to larger scale deployment.

That said, the intervention in the supplemental study included most of the same elements as the procedure used in Experiment 3's treatment condition, including gamification elements, mere exposure to true and false news headlines, easy-to-hard and hard-to-easy exemplar scheduling, and category-level metacognitive JOLs. Nevertheless, participants did not show improved discernment of true and false news after completing the intervention. This suggests that the inductive learning elements of the intervention (e.g., category labels during observational learning and feedback during classification training)

drove the improvements observed in Experiments 1–3. However, Experiments 1–3 and the supplemental study were conducted at different time points with different samples, so this suggestion remains tentative.

An important consideration is that we have only tested our intervention with one type of misinformation: fact-checked false news headlines. These headlines represent a subset of the broader misinformation landscape. In reality, much of the misinformation circulating online is unlikely to be fact-checked given the time and resources required for such verification. To assess whether the intervention's effects extend beyond fact-checked false news, future research could examine its impact on other misinformation types, such as misleading but not demonstrably false content. Lastly, the long-term effects of our intervention are underexplored in this paper. The effects of both inductive learning and misinformation interventions have been shown to decay considerably over time (Guess et al., 2020; Zulkiply & Burt, 2013a), which makes long-term efficacy a plausible issue for our intervention that warrants further investigation. Overall, however, we view this paper as an important first step in examining the immediate effects of an inductive learning intervention on news veracity discernment, using the most extensively studied type of news stimuli in the literature (Pfänder & Altay, 2025).

Constraints on Generality

Our findings suggest that inductive learning can improve people's ability to distinguish between true and false news headlines. This effect was observed using fact-checked false news headlines, as determined by fact-checking websites (e.g., Snopes), and true news headlines from mainstream news sources (e.g., BBC), with source information removed from all headlines. Thus, we expect our results to be reproduced in studies where participants are presented with these types of news headlines. However, we do not have evidence to support that our findings will generalize to, for example, non-fact-checked false news headlines with source information included. Additionally, the samples we collected all consisted of participants residing in the United States. Therefore, we do not have evidence to support that our findings will generalize to samples from other countries. We have no

reason to believe that the results depend on other characteristics of the participants, materials, or context.

Conclusion

In summary, inductive learning can improve people's ability to discriminate between true and false news headlines. We replicated this finding to varying degrees across three preregistered experiments. Although this is not the first paper to use inductive learning for the purpose of improving true and false news discrimination (see Modirrousta-Galian et al., 2023), it is the first to provide evidence for its effectiveness in this context.

References

Adams, Z., Osman, M., Bechlivanidis, C., & Meder, B. (2023). (Why) is misinformation a problem? *Perspectives on Psychological Science*, *18*(6), 1436-1463. https://doi.org/10.1177/17456916221141344

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), Article eaay3539. https://doi.org/10.1126/sciadv.aay3539

American Psychological Association (2024, July). *Misinformation and disinformation*. https://www.apa.org/topics/journalism-facts/misinformation-disinformation

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, *31*(7), 1114–1125. https://doi.org/10.3758/bf03196132

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33–53. https://doi.org/10.1037/0278-7393.14.1.33

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*(5), 666–677. https://doi.org/10.3758/bf03196423

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2012). Formal approaches in categorization. In E. M. Pothos & A. J. Willis (Eds.), *COVIS* (pp. 65–87). Cambridge University Press. https://doi.org/10.1017/CBO9780511921322.004

Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*, *7*, Article 23. https://doi.org/10.1057/s41599-020-0523-3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Bransford, J. D., Brown, A. L., Cocking, R. R., Donovan, M. S., & Pellegrino, J. W. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). The National Academies Press https://doi.org/10.17226/9853

Butler, L. H., Lamont, P., Wan, D. L. Y., Prike, T., Nasim, M., Walker, B., Fay, N., & Ecker, U. K. H. (2024). The (Mis)Information Game: A social media simulator. *Behavior Research Methods*, *56*(3), 2376–2397. https://doi.org/10.3758/s13428-023-02153-x

Calvillo, D. P., Garcia, R. J. B., Bertrand, K., & Mayers, T. A. (2021). Personality factors and self-reported political news consumption predict susceptibility to political fake news. *Personality and Individual Differences*, *174*, Article 110666. https://doi.org/10.1016/j.paid.2021.110666

Calvillo, D. P., Ross, B. J., Garcia, R. J. B., Smelter, T. J., & Rutchick, A. M. (2020). Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science, 11*(8), 1119–1128. https://doi.org/10.1177/1948550620940539

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*(5), 671–682. https://doi.org/10.3758/s13421-012-0291-4

Carrasco-Farré, C. (2022). The fingerprints of misinformation: How deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions. *Humanities and Social Sciences Communications*, *9*, Article 162. https://doi.org/10.1057/s41599-022-01174-9

Carvalho, P. F., & Goldstone, R. L. (2014a). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology*, *5*, Article 936. https://doi.org/10.3389/fpsyg.2014.00936

Carvalho, P. F., & Goldstone, R. L. (2014b). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495. https://doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*(1), 281–288. https://doi.org/10.3758/s13423-014-0676-4

Chen, X., Pennycook, G., & Rand, D. (2023). What makes news sharable on social media? *Journal of Quantitative Description: Digital Media*, *3*, 1–27. https://doi.org/10.51685/jqd.2023.007

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Cruz, A., & Minda, J. P. (2024). Was that my cue? Reactivity to category-level judgments of learning. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* (4671–4678). https://escholarship.org/content/qt0nr5974b/qt0nr5974b_noSplash_cb69bbaae9a3a49cdf969db2d51627c3.pdf

Culley, J., & Khalil, H. (2024, July 29). *Southport stabbings – what we know about the attack*. BBC. https://www.bbc.co.uk/news/articles/cy68z9dw9e7o

de Ridder, J. (2021). What's so bad about misinformation? *Inquiry*, 1–23. https://doi.org/10.1080/0020174x.2021.2002187

Do, L. A., & Thomas, A. K. (2023). The underappreciated benefits of interleaving for category learning. *Journal of Intelligence*, *11*(8), Article 153. https://doi.org/10.3390/jintelligence11080153

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*(3), 275–277. https://doi.org/10.1037/0096-3445.121.3.275

Edmunds, C. E. R., Milton, F., & Wills, A. J. (2015). Feedback can be superior to observational training for both rule-based and information-integration category structures. *Quarterly Journal of Experimental Psychology*, *68*(6), 1203–1222. https://doi.org/10.1080/17470218.2014.978875

Edmunds, C. E. R., Milton, F., & Wills, A. J. (2018). Due process in dual process: Model-recovery simulations of decision-bound strategy analysis in category learning. *Cognitive Science, 42*(S3), 833–860. https://doi.org/10.1111/cogs.12607

Edmunds, C. E. R., Wills, A. J., & Milton, F. (2019). Initial training with difficult items does not facilitate category learning. *The Quarterly Journal of Experimental Psychology, 72*(2), 151–167. https://doi.org/10.1080/17470218.2017.1370477

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fetzer, J. H. (2004). Disinformation: The use of false information. *Minds and Machines*, *14*(2), 231–240. https://doi.org/10.1023/B:MIND.0000021683.28604.5b

Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, *7*(23), Article eabf1234. https://doi.org/10.1126/sciadv.abf1234

Gómez Calderón, B., Córdoba-Cabús, A., & López-Martín, Á. (2023). Las fake news y su percepción por parte de los jóvenes españoles: El influjo de los factores sociodemográficos. *Doxa Comunicación*, *36*, 19–42. https://doi.org/10.31921/doxacom.n36a1741

Graham, M. E., Skov, B., Gilson, Z., Heise, C., Fallow, K. M., Mah, E. Y., & Lindsay, D. S. (2023). Mixed news about the Bad News game. *Journal of Cognition*, *6*(1), 58. https://doi.org/10.5334/joc.324

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Gupta, M., Dennehy, D., Parra, C. M., Mäntymäki, M., & Dwivedi, Y. K. (2023). Fake news believability: The effects of political beliefs and espoused cultural values. *Information & Management*, *60*(2), Article 103745. https://doi.org/10.1016/j.im.2022.103745

Guzman-Munoz, F. J. (2017). The advantage of mixing examples in inductive learning: A comparison of three hypotheses. *Educational Psychology*, *37*(4), 421–437. https://doi.org/10.1080/01443410.2015.1127331

Henderson, E. L., Westwood, S. J., & Simons, D. J. (2022). A reproducible systematic map of research on the illusory truth effect. *Psychonomic Bulletin & Review, 29*(3), 1065–1088. https://doi.org/10.3758/s13423-021-01995-w

Higham, P. A., Fastrich, G. M., Potts, R., Murayama, K., Pickering, J. S., & Hadwin, J. A. (2023). Spaced retrieval practice: Can restudying trump retrieval? *Educational Psychology Review*, *35*(4), Article 98. https://doi.org/10.1007/s10648-023-09809-2

Higham, P. A., & Higham, D. P. (2019). New improved gamma: Enhancing the accuracy of Goodman–Kruskal's gamma using ROC curves. *Behavior Research Methods*, *51*(1), 108–125. https://doi.org/10.3758/s13428-018- 1125-5

Higham, P. A., Modirrousta-Galian, A., & Seabrooke, T. (2024). Mean rating difference scores are poor measures of discernment: The role of response criteria. *Current Opinion in Psychology*, *56*, Article 101785. https://doi.org/10.1016/j.copsyc.2023.101785

Hoes, E., Aitken, B., Zhang, J., Gackowski, T., & Wojcieszak, M. (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human Behaviour*. Advance online publication. https://doi.org/10.1038/s41562-024-01884-x

Hughes, G. I., & Thomas, A. K. (2021). Visual category learning: Navigating the intersection of rules and similarity. *Psychonomic Bulletin & Review*, *28*(3), 711–731. https://doi.org/10.3758/s13423-020-01838-0

Jiménez-Sánchez, A., Mateus, D., Kirchhoff, S., Kirchhoff, C., Biberthaler, P., Navab, N., González Ballester, M. A., & Piella, G. (2022). Curriculum learning for improved femur fracture classification: Scheduling data with prior knowledge and uncertainty. *Medical Image Analysis*, *75*, Article 102273. https://doi.org/10.1016/j.media.2021.102273

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Kelley, N. J., Hurley-Wallace, A. L., Warner, K. L., & Hanoch, Y. (2023). Analytical reasoning reduces internet fraud susceptibility. *Computers in Human Behavior*, *142*, Article 107648. https://doi.org/10.1016/j.chb.2022.107648

Kimathi, S. K. (2024, August 19). *For black Britons, UK riots leave lasting scars*. Reuters. https://www.reuters.com/world/uk/black-britons-uk-riots-leave-lasting-scars-2024-08-19/

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction?". *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., & Vaughn, K. E. (2018). In inductive category learning, people simultaneously block and space their studying using a strategy of being thorough and fair. *Archives of Scientific Psychology*, *6*(1), 138–147. https://doi.org/10.1037/arc0000042

Kuntur, S., Wróblewska, A., Paprzycki, M., & Ganzha, M. (2025). *Fake news detection: It's all in the data!* arXiv. https://doi.org/10.48550/arXiv.2407.02122

Leder, J., Schellinger, L. V., Maertens, R., van der Linden, S., Chryst, B., & Roozenbeek, J. (2024). Feedback exercises boost discernment of misinformation for gamified inoculation interventions. *Journal of Experimental Psychology: General, 153*(8), 2068–2087. https://doi.org/10.1037/xge0001603

Lee, H. S., & Ha, H. (2019). Metacognitive judgments of prior material facilitate the learning of new material: The forward effect of metacognitive judgments in inductive learning. *Journal of Educational Psychology, 111*(7), 1189-1201. https://doi.org/10.1037/edu0000339

Lindsay, M., & Grewar, C. (2024, August 9). *Social media misinformation 'fanned riots flames'*. BBC. https://www.bbc.co.uk/news/articles/c70jz2r4lp0o

Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition*, *43*(2), 283–297. https://doi.org/10.3758/s13421-014-0475-1

Lu, C., Hu, B., Li, Q., Bi, C., & Ju, X.-D. (2023). Psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *25*, Article e49255. https://doi.org/10.2196/49255

Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review*, *11*(5), 945–952. https://doi.org/10.3758/bf03196726

Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V., & Ashby, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research*, *74*(2), 219–236. https://doi.org/10.1007/s00426-009-0245-z

Martínez-Huertas, J. Á., Olmos, R., & Ferrer, E. (2022). Model selection and model averaging for mixed-effects models with crossed random effects for subjects and

items. *Multivariate Behavioral Research*, *57*(4), 603–619. https://doi.org/10.1080/00273171.2021.1889946

Mazarakis, A., & Bräuer, P. (2023). Gamification is working, but which one exactly? Results from an experiment with four game design elements. *International Journal of Human-Computer Interaction*, *39*(3), 612–627. https://doi.org/10.1080/10447318.2022.2041909

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science*, *34*(12), 1390-1403. https://doi.org/10.1177/09567976231207095

Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General*, *152*(9), 2411–2437. https://doi.org/10.1037/xge0001395

Modirrousta-Galian, A., Higham, P. A., & Seabrooke, T. (2023). Effects of inductive learning and gamification on news veracity discernment. *Journal of Experimental Psychology: Applied*, *29*(3), 599–619. https://doi.org/10.1037/xap0000458

Modirrousta-Galian, A., Higham, P. A., & Seabrooke, T. (2025). Wordless wisdom: The dominant role of tacit knowledge in true and fake news discrimination. *Journal of Applied Research in Memory and Cognition*, *14*(2), 231–240 https://doi.org/10.1037/mac0000151

Modirrousta-Galian, A. (2026). *An inductive learning intervention to improve news veracity discernment* [Data, code, and materials]. OSF. https://doi.org/10.17605/OSF.IO/5URX7

Modirrousta-Galian, A., Higham, P., Seabrooke, T., & Lam, C. (2023). *Kitchen sink study 1* [Preregistration]. AsPredicted. https://aspredicted.org/SHR_R5S

Modirrousta-Galian, A., Higham, P., & Seabrooke, T. (2023). *Kitchen sink study 2 (updated)* [Preregistration]. AsPredicted. https://aspredicted.org/SLH_Y5S

Modirrousta-Galian, A., Higham, P., & Seabrooke, T. (2024). *Kitchen sink study 3* [Preregistration]. AsPredicted. https://aspredicted.org/34P_FQP

Motz, B. A., Fyfe, E. R., & Guba, T. P. (2023). Learning to call bullsh*t via induction: Categorization training improves critical thinking performance. *Journal of Applied Research in Memory and Cognition, 12*(3), 310–324. https://doi.org/10.1037/mac0000053

Newell, B. R., Dunn, J. C., & Kalish, M. (2011). Systems of category learning: Fact or fantasy? In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 167–215). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-385527-5.00006-1

Noh, S. M., Yan, V. X., Bjork, R. A., & Maddox, W. T. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, *155*, 23–29. https://doi.org/10.1016/j.cognition.2016.06.007

Nosofsky, R. M., & McDaniel, M. A. (2019). Recommendations from cognitive psychology for enhancing the teaching of natural-science categories. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 21–28. https://doi.org/10.1177/2372732218814861

Seabrooke, T., Modirrousta-Galian, A., & Higham, P. A. (2026). Re-examining the bad news game: No evidence of improved discrimination of Indian true and fake news headlines. *Psychonomic Bulletin & Review*, *33*, Article 13. https://doi.org/10.3758/s13423-025-02827-x

Seabrooke, T., Pattni, M., & Higham, P. A. (2025). *Enhancing human detection of real and AI-generated hyperrealistic faces*. PsyArXiv. https://osf.io/preprints/psyarxiv/xjemh_v2

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Pennycook, G., Berinsky, A. J., Bhargava, P., Lin, H., Cole, R., Goldberg, B., Lewandowsky, S., & Rand, D. (in press). Technique-based inoculation and accuracy prompts must be combined to increase truth discernment online. *Nature Human Behavior*.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Perez, E., & Cole. D. (2020, December 1). *William Barr says there is no evidence of widespread fraud in presidential election*. CNN. https://edition.cnn.com/2020/12/01/politics/william-barr-election-2020/index.html

Pfänder J., & Altay, S. (2025). Spotting false news and doubting true news: A systematic review and meta-analysis of news judgements. *Nature Human Behavior*, *9*, 688–699. https://doi.org/10.1038/s41562-024-02086-1

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer.

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve and of d'e. *Psychological Bulletin, 71*(3), 161–173. https://doi.org/10.1037/h0026862

Prince, M. J., & Felder, R. M. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, *95*(2), 123–138. https://doi.org/10.1002/j.2168-9830.2006.tb00884.x

Roads, B. D., Xu, B., Robinson, J. K., & Tanaka, J. W. (2018). The easy-to-hard training advantage with real-world medical images. *Cognitive Research: Principles and Implications*, *3*. Article 38. https://doi.org/10.1186/s41235-018-0131-6

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*(1), Article 65. https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., & van der Linden, S. (2022). How to combat health misinformation: A psychological approach. *American Journal of Health Promotion*, *36*(3), 569–575. https://doi.org/10.1177/08901171211070958

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), Article eabo6254. https://doi.org/10.1126/sciadv.abo6254

Ryoo, J. H. (2011). Model selection with the linear mixed model for longitudinal data. *Multivariate Behavioral Research*, *46*(4), 598–624. https://doi.org/10.1080/00273171.2011.589264

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Singman, B. (2020, December 1). *Barr: DOJ yet to find widespread voter fraud that could have changed 2020 election*. Fox News. https://www.foxnews.com/politics/william-barr-doj-fbi-voter-fraud-2020-election

Soetekouw, L., & Angelopoulos, S. (2024). Digital resilience through training protocols: Learning to identify fake news on social media. *Information Systems Frontiers*, *26*(2), 459–475. https://doi.org/10.1007/s10796-021-10240-7

Southwell, B. G., Brennen, J. S. B., Paquin, R., Boudewyns, V., & Zeng, J. (2022). Defining and measuring scientific misinformation. *The ANNALS of the American Academy of Political and Social Science*, *700*(1), 98-111. https://doi.org/10.1177/00027162221084709

Søe, S.O. (2021). A unified account of information, misinformation, and disinformation. *Synthese*, *198*(6), 5929–5949. https://doi.org/10.1007/s11229-019-02444-x

Spiering, B. J., & Ashby, F. G. (2008). Response processes in information–integration category learning. *Neurobiology of Learning and Memory*, *90*(2), 330–338. https://doi.org/10.1016/j.nlm.2008.04.015

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers, 31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stosic, M. D., Murphy, B. A., Duong, F., Fultz, A. A., Harvey, S. E., & Bernieri, F. (2024). Careless responding: Why many findings are spurious or spuriously inflated. *Advances in Methods and Practices in Psychological Science*, *7*(1), 1–19. https://doi.org/10.1177/25152459241231581

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837–848. https://doi.org/10.1002/acp.1598

Thai, K.-P., Krasne, S., & Kellman, P. J. (2015). Adaptive perceptual learning in electrocardiography: The synergy of passive and active classification. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (2350–2355). https://kellmanlab.psych.ucla.edu/files/thai_krasne_kellman_2015.pdf

Thakar, H., & Bhatt, B. (2024). Fake news detection: recent trends and challenges. *Social Network Analysis and Mining*, *14*(1), Article 176. https://doi.org/10.1007/s13278-024-01344-4

Thrane, G., Murphy, M. A., & Sunnerhagen, K. S. (2018). Recovery of kinematic arm function in well-performing people with subacute stroke: A longitudinal cohort study. *Journal of NeuroEngineering and Rehabilitation*, *15*(1). https://doi.org/10.1186/s12984-018-0409-4

van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, *35*(5), 560–575. https://doi.org/10.1080/10410236.2019.1573295

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750–763. https://doi.org/10.3758/s13421-010-0063-y

West, B. T., Welch, K. B., & Galecki, A. T. (2007*). Linear mixed models: A practical guide using statistical software*. Chapman & Hall/CRC. https://www.academia.edu/37093545/LINEAR_MIXED_MODELS_A_Practical_Guide_Using_Statistical_Software

Yan, M., Zhou, W., Shu, H., Yusupu, R., Miao, D., Krügel, A., & Kliegl, R. (2014). Eye movements guided by morphological structure: Evidence from the Uighur language. *Cognition*, *132*(2), 181–215. https://doi.org/10.1016/j.cognition.2014.03.008

Zulkiply, N., & Burt, J. S. (2013a). Inductive learning: Does interleaving exemplars affect long-term retention? *Malaysian Journal of Learning and Instruction*, *10*, 133–155. https://doi.org/10.32890/mjli.10.2013.7655

Zulkiply, N., & Burt, J. S. (2013b). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*(1), 16–27. https://doi.org/10.3758/s13421-012-0238-9