Effects of Inductive Learning and Gamification on News Veracity Discernment

Ariana Modirrousta-Galian, Philip A. Higham, and Tina Seabrooke

University of Southampton

Author Note

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

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

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

Additional materials are available at: https://osf.io/dnfrh/. We have no conflicts of interest to disclose. Our work was funded by the Economic and Social Research Council South Coast Doctoral Training Partnership.

Correspondence concerning this article should be addressed to Ariana Modirrousta-Galian, Centre for Perception and Cognition, School of Psychology, University of Southampton, Highfield, Southampton, UK, SO17 1BJ. Email: amg1g17@soton.ac.uk

Abstract

This pre-registered study tests a novel psychological intervention to improve news veracity discernment. The main intervention involved inductive learning (IL) training (i.e., practice discriminating between multiple true and fake news exemplars with feedback) with or without gamification. Participants ($N = 282$ Prolific users) were randomly assigned to either a gamified IL intervention, a non-gamified version of the same IL intervention, a no-treatment control group, or a Bad News intervention, a notable web-based game designed to tackle online misinformation. Following the intervention (if applicable), all participants rated the veracity of a novel set of news headlines. We hypothesized that the gamified intervention would be the most effective for improving news veracity discernment, followed by its non-gamified equivalent, then Bad News, and finally the control group. The results were analyzed with receiver operating characteristic curve analyses, which has previously never been applied to news veracity discernment. The analyses indicated that there were no significant differences between conditions and the Bayes factor indicated very strong evidence for the null. This finding raises questions about the effectiveness of current psychological interventions and contradicts prior research that has supported the efficacy of Bad News. Age, gender, and political leaning all predicted news veracity discernment.

Keywords: Online misinformation, fake news, inductive learning, category learning, gamification

Public Significance Statement

This pre-registered study proposes a novel psychological intervention informed by literatures on inductive learning and gamification for improving news veracity discernment. Although this intervention did not improve news veracity discernment, we highlight several correctable issues with the way in which inductive learning and gamification were applied, which could inform future research attempts to use them in this context. The present study also uses receiver operating characteristic curves to analyze results, which have never been applied to news veracity discernment despite their suitability.

Effects of Inductive Learning and Gamification on News Veracity Discernment

Consider a society in which objective facts hold less weight than the opinions of popular social media users, and in which expert knowledge is labelled deceitful whenever its content does not appeal to people's attitudes or personal opinions. Although this may seem like a fictional dystopian reality, modern philosophers and political scientists argue that this describes our society today. They refer to the current period, which is characterized by this shift in societal perspectives and behaviors, as the post-truth era (Lewandowsky et al., 2017; Spoelstra, 2020).

The emergence of the post-truth era may have been the chief driving force behind the present-day dissemination and acceptance of online misinformation, which refers to any false or misleading content shared on the internet (Allen et al., 2020). There are entire websites dedicated to spreading online misinformation, such as "Infowars" and "Now8news" (links are not provided due to malware risks), and an alarming number of individual social media users spread such misinformation as well, both intentionally and unintentionally. Indeed, 42.8% of UK adults who frequently shared news on their online platforms admitted to deliberately sharing false stories (Chadwick & Vaccari, 2019), and 38.2% of US adults reported accidentally shared fake news or information on social media (Watson, 2020).

The prevalence of online misinformation would be less problematic if people were able to identify and disregard it. Unfortunately, research suggests that the opposite is true; Pew Research Centre (2019) conducted a survey on US adults and found that 49% of them had shared a story online that they later realized contained false information. This illustrates people's inability to consistently discriminate between true and fake news, which can lead to harmful consequences. For example, it has made people doubt the credibility of certain scientific truths, such as the existence of climate change and the benefits of receiving vaccines (Burki, 2020; van der Linden et al., 2017). This unfounded skepticism could result in irreversible damage to our planet and resurgence of diseases that were once under control. These are just two of the many potentially dangerous outcomes that could stem from online misinformation and people's inability to identify it.

Discovering ways to prevent such damaging consequences is vital. One approach would be to tackle the spread of online misinformation, but this is a complex, multi-faceted issue that requires government action and policy change (Kyza et al., 2020). A second approach would be to develop an effective intervention to improve people's ability to identify fake news, which is certainly within the scope of psychological research. If people were taught how to recognize online misinformation, its harmful effects might be minimized, and the incentive to create such false content may decrease due to a lack of public confidence and support (Musgrove et al., 2018).

Various cognitive biases could interfere with this objective. For example, confirmation bias is the tendency for people to look for information that aligns with their prior beliefs (Nickerson, 1998). This is linked to the belief bias, which has two main suppositions: (a) people are more likely to accept belief-consistent information than belief-inconsistent information; and (b) people tend to judge belief-consistent arguments as more logical than belief-inconsistent arguments (Edwards & Smith, 1996; Lord et al., 1979). These two connected outcomes are likely driven by the bias blind spot and the myside bias. The bias blind spot refers to the belief that one is a more unprejudiced reasoner than everyone else, while the myside bias refers to the propensity for evaluating information with a predisposed preferentialism towards one's own prior beliefs (Stanovich et al., 2013; Pronin et al., 2002). These biases may explain why fake news has such long-lasting effects. That is, once people have accepted misinformation as true, they often continue to believe in it even after it has been corrected, which is termed the continued influence effect (Lewandowsky et al., 2012). For a more exhaustive list of relevant cognitive biases, see Britt et al. (2019).

Two different types of interventions have been proposed to limit the effects of such cognitive biases, namely a priori inoculation interventions and post hoc corrections. A priori inoculation interventions aim to prevent people from believing fake news they encounter in the future, while post hoc corrections aim to correct people's belief in fake news they encountered in the past. An example of an a priori inoculation intervention is Bad News (https://www.getbadnews.com/#intro), an online browser game developed through

collaboration between a team at the University of Cambridge and Dutch media platform DROG (Roozenbeek & van der Linden, 2019). It is currently one of the most notable psychological interventions for tackling online misinformation. The theoretical framework underlying its design is based on an inoculation metaphor, where pre-emptively teaching people the common techniques used to create fake news helps them develop "cognitive immunity" against the online misinformation they may encounter in the future (Roozenbeek & van der Linden, 2019, p. 1). In contrast, Brashier et al. (2021) provided an example of post hoc corrections, where they exposed headlines as fake after they had already been presented. This was more effective in improving news veracity discernment than providing the same information during or before their presentation. However, post hoc corrections are not always more beneficial than a priori inoculations. In fact, the evidence is mixed, as other studies have shown that a priori inoculations can improve news veracity discernment in the short term and increase intentions to vaccinate more than post hoc corrections (Grady et al., 2021; Jolley & Douglas, 2017).

We propose a novel a priori inoculation intervention with a post hoc correction component. The intervention is a priori as it provides training designed to prevent people from trusting online misinformation they encounter in the future by improving their news veracity discernment. Furthermore, it incorporates a type of post hoc correction during training by providing learners with immediate feedback on whether they correctly or incorrectly categorised true and fake news. The decision to include this particular post hoc correction was informed by research that demonstrated its ability to reduce the continued influence effect (Walter & Tukachinsky, 2019).

Ultimately, this paper has two main aims: to propose a theoretical framework that provides a method to improve people's ability to discriminate between true and fake news, and to test the effectiveness of a novel psychological intervention that uses this method. The proposed theoretical framework relies on two methodologies, namely inductive learning and gamification. Inductive learning involves learning to distinguish between different categories by classifying exemplars from those categories and, in most cases, receiving immediate

feedback on accuracy, while gamification entails using game design elements in non-game contexts (Birnbaum, 2013; Huotari & Hamari, 2016). Consequently, we used inductive learning and gamification in the present study to create a novel psychological intervention designed to improve discernment between true and fake news. To test the intervention's effectiveness, it was compared to a no-treatment control group in which participants did not take part in any training or intervention, a non-gamified version of the same intervention, and Bad News.

Misinformation is an umbrella term for various types of fake news. Some of these are categorized into concrete classifications, such as misinformation, which refers to false or misleading content, and disinformation, which refers to a subset of misinformation that is shared with the deliberate intent to deceive (Allen et al., 2020; Guess & Lyons, 2020). In contrast, others are presented along conceptual dimensions, such as the misinformation spreader's ontological position on truth and facts (i.e., realism vs. constructivism), typical rhetorical style (i.e., formal vs. informal), and primary audience (i.e., individual people vs. institutions; McCright & Dunlap, 2017). More research is needed on how different types of misinformation differ in terms of their psychological impacts (Ecker et al., 2022). Therefore, we will not distinguish between these different categories and dimensions throughout this paper.

However, the distinction between false and misleading content, both of which constitute misinformation, is more pertinent to this paper. This distinction highlights the fact that misinformation can exist on a continuum of falsity; it can include false content that is entirely fictitious, or misleading content that involves half-truths or information that is mostly true but communicated in an inaccurate or deceptive way (Mourão & Robertson, 2019). Crucially, this paper tackles the issue of news veracity discernment, which is the ability to discern the objective truth of news items. Seeing as the objective truth of misleading content is difficult to determine, here we focus on false content.

Inductive Learning

To incorporate inductive learning into the present study's psychological intervention, true and fake news were conceptualized as two separate categories, which one can learn to differentiate by observing exemplars. The rationale behind this framework is simple; by being exposed to exemplars, it is possible to learn the commonalities between items from the same category, as well as the discrepancies between items from different categories (Birnbaum et al., 2013). Although inductive learning has never been applied to true and fake news, it has proven effective in other circumstances. For example, it has improved people's ability to discriminate between different butterfly species, bird species, x-ray interpretations, and paintings from different artists (Birnbaum et al., 2013; Kornell & Bjork, 2008; Rozenshtein et al., 2016; Wahlheim et al., 2011; Yan et al., 2016).

The effectiveness of inductive learning is determined by whether it results in generalization, which refers to the ability to generalize learning from previously studied cases to novel cases (Kruschke, 2005). To illustrate this concept, consider someone who has undergone inductive learning with a set of true and fake news items. Suppose they have learned to discriminate between these specific cases, but when presented with a novel set of true and fake news items, they are unable to distinguish between them. This would suggest that the knowledge they have acquired from inductive learning is limited to certain instances and cannot be generalized to new instances. Therefore, this knowledge is limited aside from the unlikely event that they encounter the same set of previously learned items. To assess generalization, the items used during inductive learning should be different from those used to evaluate discernment between categories.

One aspect of inductive learning that influences its effectiveness is the order in which items are presented during learning, which is termed *sequencing* (Noh et al., 2016). The two most common forms of sequencing are blocking and interleaving. The former refers to presenting items from each category in separate blocks with infrequent category alternation, while the latter refers to presenting items from each category in an intermixed manner with frequent category alternation (Birnbaum et al., 2013). Several studies have suggested that interleaving results in better discrimination than blocking for high similarity categories, as it

highlights the differences between them (Carvalho & Goldstone, 2014; Kornell & Vaughn, 2018; Sana et al., 2017). These studies have also indicated that blocking results in better discrimination than interleaving for low similarity categories, as it emphasizes the commonalities between exemplars of the same category. Luo et al. (2020) reported that participants could only discriminate between true and fake news exemplars with 51% accuracy (i.e., near chance), which suggests that the true and fake news categories are of high similarity. If this is the case, true and fake news exemplars should be presented in an interleaved manner during inductive learning to improve discrimination between them. The proposed theoretical explanation for the effectiveness of interleaved sequencing is termed the *discriminative-contrast hypothesis*, which states that interleaving items from different categories allows for direct comparison between them and thus highlights their differences, promoting discrimination (Guzman-Munoz, 2017).

Another aspect of inductive learning that influences its effectiveness is whether feedback is provided on response accuracy. Research has demonstrated that feedback boosts learning, particularly when given immediately after a learner's response (Brosvic & Epstein, 2007). Wiklund-Hörnqvist et al. (2014) proposed two potential reasons for the positive effects of feedback in learning contexts. Firstly, they argued that it may act as an additional learning opportunity that prevents recurring retrieval failures. These refer to memory lapses, namely temporary failures to retrieve information from long-term memory (Miller, 2021). Secondly, they hypothesized that feedback may prevent erroneous learning, which can occur when an item is incorrectly categorized, and feedback is not provided to correct the error. Erroneous responses such as these are often encoded in a learner's long-term memory and thus repeated later (Roediger & Marsh, 2005). Also, as previously mentioned, immediate feedback is useful for mitigating certain cognitive biases, such as the continued influence effect (Walter & Tukachinsky, 2019).

Overall, several factors must be considered to make inductive learning as effective as possible. First, interleaved sequencing should be implemented to promote discriminative contrast between high similarity categories. Second, immediate feedback should be

provided on response accuracy to prevent retrieval failures, erroneous learning, and cognitive biases. Finally, to determine the efficacy of the inductive learning process, the generalization of knowledge from specific to novel cases should be tested. This can be done by ensuring that the items used during learning are different from those used to assess category discrimination. Accordingly, the first two factors were applied to the novel psychological intervention, and the last factor was applied to test its effectiveness. Therefore, the intervention presented participants with news items in an interleaved manner, which they categorized as either true or false and received immediate feedback on response accuracy. To assess participants' discernment and generalization after taking part in the intervention, they rated the reliability of a different set of news items.

The Nature of Knowledge Acquired with Inductive Learning

Most research has not specified what type of knowledge is acquired with inductive learning procedures. Kornell and Bjork (2008), for example, demonstrated that participants could discern paintings created by different painters after an inductive learning procedure and asked participants a question about their learning following the test. However, that question pertained to whether participants believed blocked sequencing (all examples from one artist presented in succession, followed by all paintings by the next artist) or interleaved sequencing (intermixing different artists' paintings during training) was more beneficial to learning. There were no questions about how participants were able to make the discrimination (e.g., which features they used). Thus, in this and many other inductive learning studies, the nature of knowledge acquired with inductive learning is not known.

However, there is a large body of research on so-called "implicit learning" which may provide some answers to this question. In implicit learning studies, participants first study a sequence of exemplars as in inductive training procedures, but all the exemplars are consistent with a complex rule structure. In the classic version of these studies, the exemplars are consonant letter strings such as "MVXRT" that are generated by a finite-state grammar (e.g., Higham, 1997; Reber, 1967). After studying the exemplars, participants are informed about the rule structure and then shown new grammatical strings intermixed with

new non-grammatical strings and asked to discriminate between them. Accuracy on this

discrimination task is typically about 65%, well above chance (50%). However, when

participants are queried about the basis of their decisions, they cannot describe the rule

structure and sometimes describe bases that are inconsistent with their decisions. Implicit

learning has also been demonstrated with other complex rule structures such as rule sets

defined by feature conjunctions (e.g., Higham & Brooks, 1997; Neil & Higham, 2012, 2020),

the presence of an invariant digit in a sequence of four-digit numbers (Wright & Burton,

1995), and even the position of hour and minute hands on analogue clock faces (Bright &

Burton, 1994).

Potentially, an inductive learning procedure with true and fake news items could give

rise to tacit or implicit knowledge like that seen in implicit learning research. Alternatively, it

is possible that participants explicitly identify specific features that discriminate the two

categories, or explicitly base their decisions on global similarity between test items and

training items. In our view, although discovering the nature of any knowledge acquired is an

interesting and necessary line of inquiry, it is first necessary to demonstrate that inductive

training can produce learning with news headlines. Thus, our approach in the current study

was to select a large sample of true and fake news items from various sources on the

internet to use as materials for training and testing. We considered this to be the best

method to ensure that the news items we used in our research would be representative of

the news items people actually encounter on the internet.[1] Unlike previous research on Bad

News (e.g., Roozenbeek & van der Linden, 2019), we made no a priori assumptions about

the features that might or might not discriminate the items; rather, we left it to the

---

[1] Another strategy is to create a standardized scale of misinformation susceptibility that uses a limited set of items that generate similar responses across multiple studies. Maertens et al. (2022) adopted this strategy when developing the Misinformation Susceptibility Test (MIST). Our concern with the MIST (and other tests like it) is that although the items may perform similarly across studies, that does not guarantee that they are representative of the items people are likely to encounter on the internet. Furthermore, using a limited set of standardized items does not allow researchers to experimentally manipulate the items (as this would render them unstandardized), which may be important for understanding how participants discriminate between true and fake news. For example, researchers could present the same set of items to two different groups of participants but add a feature that is thought to be more common in fake news, such as grammatical errors, to one item set. If this manipulation leads to more "fake" responses, then researchers could conclude that grammatical errors distinguish fake from true news. A final limitation is that the standardized item set may "expire" relatively quickly as news quickly becomes obsolete, and it would therefore need to be continually updated.

participants to discover them. After all, if participants learn to discriminate the items but the learning is implicit, it is unlikely that the experimenters are in any better position than the participants to explicitly describe the differences between the item categories.

Gamification

To incorporate gamification into the present study's psychological intervention, game design elements were added. These refer to specific features that are shared by most games, such as points, badges, and rewards (Deterding et al., 2011). In general, game design elements are useful in capturing people's attention and encouraging them to reach their learning goals (Brauer et al., 2017). Such gamification techniques have been shown to improve motivation and engagement in a wide variety of learning contexts, including but not limited to computer programming, medical instruction, and second language learning (Boudadi & Gutiérrez-Colón, 2020; Dong et al., 2012; Pesare et al., 2016). However, when considering which game design elements to integrate into a learning environment, it is essential to investigate which are the most effective and thus worth using.

Through reviewing the gamification literature, we found four game design elements that consistently increased motivation and engagement during learning. The first was the use of eye-catching colors, which is thought to increase interest in learning (Ainley & Ainley, 2011; Pratama & Setyaningrum, 2018). The second was the inclusion of dialogue between the player and the game's narrator, which Toda et al. (2019) argued encouraged engagement by fostering interactivity. The third was the addition of a progress bar that advanced as a learner progressed through the learning experience. Progress bars are believed to increase engagement as they generate a sense of advancement towards the end goal, that is, completing the game (Sanchez et al., 2020). Finally, the fourth was the delivery of achievement badges, which are also thought to increase engagement by promoting a feeling of progression towards the final objective (Hakulinen et al., 2015).

To learn effectively, one must be willing to focus and work hard. Motivation and engagement incite such willingness, making them indispensable in learning environments (Filgona et al., 2020). Due to their effectiveness in increasing these essential driving forces,

the four abovementioned game design elements (i.e., eye-catching colors, dialogue between the player and the game's narrator, a progress bar, and achievement badges) were incorporated into the novel psychological intervention.

Bad News

The present study is not the first to use gamification in the context of online misinformation. In fact, gamified psychological interventions designed to improve people's ability at recognizing online misinformation are becoming increasingly prevalent, which explains the popularity of Bad News (Roozenbeek & van der Linden, 2019). This online browser game makes the player use different techniques to mass produce and spread online misinformation in a virtual world. These techniques include (a) impersonating people; (b) emotional language; (c) group polarization; (d) conspiracy theories; (e) discrediting opponents; and (f) trolling.

Several studies have provided evidence for the effectiveness of Bad News (Basol et al., 2020; Maertens et al., 2021; Roozenbeek et al., 2020; Roozenbeek & van der Linden, 2019). However, various methodological issues cast doubt over their findings. Firstly, one of these studies did not use a control group, which made it impossible to isolate the effects of playing Bad News on participants' ability to identify online misinformation. Secondly, they typically used pre-post designs in which participants rated the reliability of the same true and fake news headlines in the pre-test and the post-test. This introduced the confounding factor of memory, as participants may have recalled their pre-test ratings and used them to inform their post-test ratings (Knapp, 2016). Finally, all these studies used incomparable true and fake news items in the pre-test and the post-test. Specifically, the true news headlines had been reported (often extensively) by the mainstream media and were therefore likely to have been familiar to the participants (e.g., "President Trump wants to build a wall between the United States and Mexico"). In contrast, the fake news headlines were completely unfamiliar to participants, having been created by the researchers (e.g., The 8th season of #GameOfThrones will be postponed due to a salary dispute"). Therefore, participants' prior

knowledge might have acted as an additional confounding factor, influencing their ratings for the true news headlines but not the fake news headlines.

Notably, several of the more recent studies have attempted to address some of these criticisms. For example, Roozenbeek et al. (2020) aimed to tackle the issue of memory confounds caused by using the same true and fake news items in the pre-test and the post-test (i.e., item effects). They also addressed the issue that the implementation of a pre-test might affect post-test performance (i.e., testing effects), since prior experience with the testing procedure could be a cause of any observed effects. Roozenbeek et al. found evidence for item effects, but not testing effects. Furthermore, Maertens et al. (2021, Experiment 3) also tackled the issue of memory confounds but found no evidence for them. Nevertheless, these studies also involved some methodological concerns.

Despite finding evidence for item effects, when examining testing effects, Roozenbeek et al. (2020) did not keep the items consistent between experimental conditions.[2] Furthermore, Maertens et al. (2021) compared participants' reliability ratings for item sets that were unbalanced in their true-to-fake ratio (1:6 vs. 6:6), making the comparison between them unreliable, particularly for the true news items. Finally, both Roozenbeek et al. and Maertens et al. used a small number of items (ranging from seven to twelve) to assess participants' reliability ratings. This raises the concern that any observed effects are specific to those particular items. Consequently, although these studies provide evidence that some of the issues associated with pre-post designs may not be as relevant as previous research has suggested, replication studies with larger sets of balanced news items that are kept consistent between experimental conditions would help to confirm such findings.

These concerns led us to scrutinize the results of prior research that analyzed the effects of Bad News. We did this by plotting pre-post differences as a function of pre-test

---

[2] It is important to recognize that Roozenbeek et al. (2020) acknowledged this issue in the notes section of their paper. They did not consider it to be a major limitation since they conducted an analysis only with the shared items between all conditions and found similar results. However, this does not completely nullify the issue.

reliability ratings across four studies (see Figure 1; Basol et al., 2020; Maertens et al., 2021; Roozenbeek et al., 2020; Roozenbeek & van der Linden, 2019). These studies concluded that Bad News improved people's ability to identify online misinformation because participants generally rated fake news items, but not true news items, as more unreliable after playing the game compared to before. However, Figure 1 suggests that pre-post differences depended more on pre-test reliability ratings rather than the objective veracity of news items. Specifically, only items with medium pre-test reliability ratings were rated as less reliable after playing Bad News; surprisingly, items with very low or very high pre-test reliability ratings were rated as *more* reliable after playing Bad News. Therefore, the efficacy of this intervention appeared to be dependent on pre-test assessments of reliability, which challenges both the findings of prior research and the effectiveness of Bad News.

Figure 1

Pre-Post Differences Plotted as a Function of Pre-Test Reliability Ratings for the True and Fake News

Items in Basol et al. (2020), Maertens et al. (2021), Roozenbeek et al. (2020), and Roozenbeek and

van der Linden (2019)
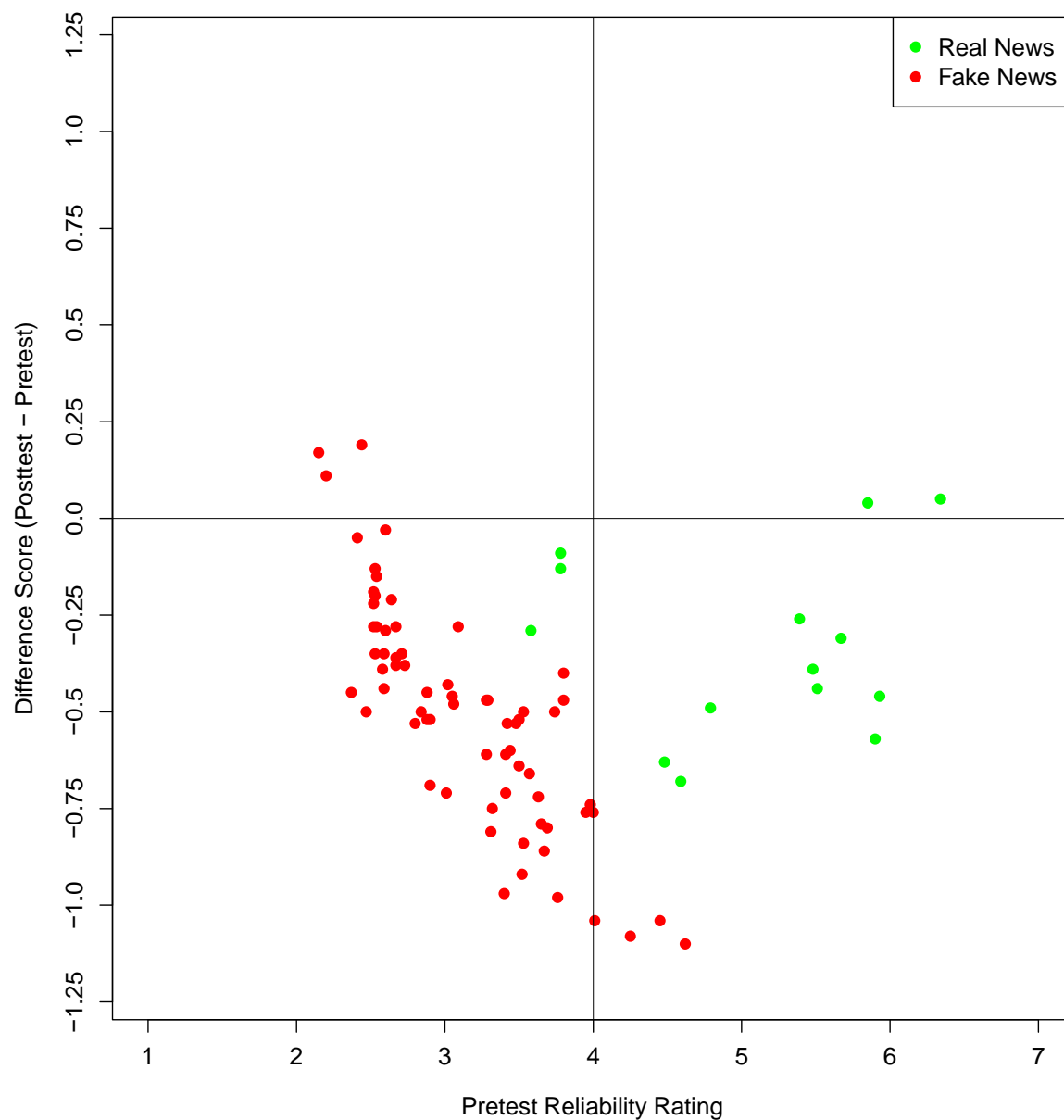
Note. Negative versus positive difference scores plotted on the y-axis indicate that, compared to pre-

test, playing Bad News decreased versus increased reliability ratings, respectively. For replication

purposes, it must be noted that data were limited to the inoculation group in Basol et al. (2020), the

pre-post group of Experiment 2 in Roozenbeek et al. (2020), and T1 and T2 of all three experiments

in Maertens et al. (2021). It is also worth mentioning that most studies that have supported the

effectiveness of Bad News have used the same or mostly the same news items. In line with this, many of the news items are shared between these four studies.

There is a simple explanation as to why items with very low pre-test reliability scores (i.e., 2) yielded positive difference scores; for them to have negative difference scores, participants would have had to assign them a post-test reliability rating at the lower limit of the scale (i.e., 1). However, items with very high pre-test reliability scores (i.e., 6) also yielded positive difference scores, even though they could have been assigned much lower post-test reliability ratings (i.e., anything under 6). This suggests that Bad News did not increase skepticism for all news items, only those that participants assigned medium pre-test reliability ratings and were thus unsure about to begin with, regardless of their objective veracity. Nonetheless, the bounded nature of the scale does not take away from the main point of the figure; Bad News had the same effect on true and fake news items as long as they were not subject to floor or ceiling effects.

In summary, most studies that have reported that Bad News improves people's ability to identify online misinformation have had several methodological issues. Although more recent papers have addressed some of these problems, there is arguably still scope for improvement. Furthermore, a graphical analysis of the results of these prior investigations (Figure 1) suggests that when people are not certain about the veracity of news items to begin with, Bad News may simply make them generally more skeptical of news items, regardless of whether they are true or fake (see Modirrousta-Galian & Higham, 2022 for further discussion on this issue). Reduced belief in truth (e.g., that COVID-19 booster vaccines are important) can sometimes be more damaging than increased belief in falsehoods (e.g., that 5G towers cause COVID-19). Consequently, we have included Bad News in our paper to test its effectiveness in a methodologically robust experiment and thus clarify its impact.

Present Study

The present study's research question is the following: Can inductive learning and/or gamification improve people's ability to discriminate between true and fake news? To address this query, a novel psychological intervention was created that incorporated these two factors. It involves categorizing interleaved examples of true and fake news and receiving immediate accuracy feedback in a game-like environment. This gamified intervention was compared to a no-treatment control group, a non-gamified version of the same intervention, and Bad News. We hypothesized that the novel gamified intervention would be the most effective in improving discernment between novel true and fake news items, followed by its non-gamified equivalent, then Bad News, and finally the control group. A method from signal detection theory (SDT), namely receiver operating characteristic (ROC) analysis, was used to assess discrimination free from response bias (Higham & Higham, 2019). Despite its ability to separate these two distinct aspects of performance, this particular analytic technique has previously never been applied to news veracity discernment.

Method

Transparency and Openness

All data, analytic code, and materials needed to replicate this study are available in our additional materials at https://osf.io/dnfrh. The present study, including its hypothesis and analysis plan, was preregistered on the AsPredicted website (https://aspredicted.org/2FK_XS6). We obtained ethical approval to conduct this research from the University of Southampton Faculty of Environmental and Life Sciences Ethics Committee (65104).

Participants

An a priori power analysis in G*Power 3.1 indicated that 280 participants were required across four groups to detect a small-to-medium effect size with a one-way ANOVA ($n \approx 280$, $f = .20$, $1-\beta = .80$, $\alpha = .05$). We slightly oversampled to account for exclusions and recruited 300 participants from an online academic research platform called Prolific. The participant pool was restricted to those fluent in English, between the ages of 18 and 65, and

residing in the USA at the time of the study. After excluding 18 participants for failing attention checks, the final sample consisted of 282 individuals. This included 128 females, 148 males, five individuals who identified as "other", and one individual who preferred not to disclose their gender. Participants were between the ages of 18 and 64 (M = 31.84, SD = 9.60) and were paid at a rate of £5.00 per hour. Of the 282 participants in the final sample, 72 were randomly assigned to the baseline condition and 70 were randomly assigned to each of the non-gamified, gamified, and Bad News conditions.

Materials and Design

Software Package and Data Storage

This experiment was created using a JavaScript library called jsPsych (de Leeuw, 2015). All participant data were stored on JATOS, a web-application that sets up private and secure servers for their users (Lange et al., 2015).

Training Phases

There were three different training phases, one for each of the non-gamified, gamified, and Bad News conditions. The training phase for the non-gamified condition included 42 news headlines, of which 21 contained true information and 21 contained false information. We did not omit source information from these items to make them comparable to news headlines that participants typically encounter online. The true news headlines were selected from mainstream news websites such as the BBC (https://www.bbc.co.uk/) and NPR (https://www.npr.org/sections/news/), while the fake news headlines were obtained from third-party fact-checking websites such as Snopes (https://www.snopes.com/) and Full Fact (https://fullfact.org/). These headlines were presented to participants in three separate blocks, each containing 14 headlines, of which seven contained true information and seven contained false information. The order of the blocks was fixed, whereas the order of the headlines within each block was randomly determined.

The training phase for the gamified condition was identical to that of the non-gamified condition aside from the four game design elements that were added. Firstly, more eye-catching colors were used, replacing the black and white theme in the non-gamified

condition with a dark grey, light blue, and white one. Secondly, a progress bar was added that indicated progression by gradually filling its empty space with a solid light blue bar. Thirdly, three achievement badges were awarded, one after each block. Finally, dialogue between the participant and the game's narrator was included, which did not provide any further information about the task, but simply fostered interactivity. For example, participants were presented with the greeting "Hello!" at the start of the game, to which they replied by selecting one of two options, namely "Hi!" or "What's going on?".

Finally, the training phase for the Bad News condition consisted of playing Bad News, which made participants mass produce and spread online misinformation in a virtual world. This was done by using six different techniques that are commonly used in fake news dissemination: (a) impersonating people; (b) emotional language; (c) group polarization; (d) conspiracy theories; (e) discrediting opponents; and (f) trolling. An achievement badge was awarded for every technique that was used to create and share fake news, and the game ended after all six badges were obtained. For more information on Bad News, see Roozenbeek and van der Linden (2019).

Test Phase

The test phase, which was present in all conditions, included 36 news headlines that were different from those used in the training phase. Of these, 18 contained true information and 18 contained false information, and all of them had been shared on social media in the past. They were obtained from Brashier et al. (2021), who pilot tested them and determined that they were all comparable in length, similarly familiar to US participants, and equally balanced in terms of the political viewpoint they favored (i.e., conservative vs. liberal). We omitted the source information from these items to prevent participants from simply using source to determine veracity. Two attention checks that looked like news headlines but stated "please move the slider to certainly true" were also included in the test phase. The order of the headlines and attention checks was randomly determined.

Survey Questions

There were five different survey questions. The first three were presented in all conditions and asked participants for their age, gender, and political orientation. The fourth was only presented in the Bad News condition and asked participants for a completion code. The fifth was only presented in the non-gamified, gamified, and Bad News conditions and asked participants about their willingness to replay the training phase.

Experimental Design and Research Variables

A between-subject, multi-group experimental design was used as participants were randomly assigned to one of four conditions: baseline, non-gamified, gamified, and Bad News. The independent variable was the condition that participants were randomly assigned to. The primary dependent variable was participants' veracity ratings of the 36 news headlines presented in the test phase. The secondary dependent variables were (a) participants' veracity ratings of the 42 news headlines presented in the non-gamified and gamified training phases; and (b) participants' willingness to replay the non-gamified, gamified, and Bad News training phases. Age, gender, and political orientation were measured as demographic variables.

Procedure

Device restrictions were applied on Prolific, which suggested that participants access the experiment through a computer. Since this was a remote online study, participants could use any web browser and computer of their choosing. To navigate through the experiment, a mouse (or touchpad) and keyboard were necessary.

Before starting the study, participants were shown a combined information sheet and consent form. Since the experiment did not involve deception, the information on this form was accurate and provided participants with a description of the study and its aims. To indicate that they had read the form and agreed to provide informed consent, participants had to click a button at the bottom of the web page. Once they did this, they were presented with a warning: "Please do not refresh the page or hit the back button on your browser at any point throughout this study. It will cancel the experiment and you will not be compensated". After this, participants were asked for their Prolific IDs, gender, age, and

political orientation. To enter their Prolific IDs and age, participants simply typed their answers into a textbox. To indicate their gender, participants chose between 4 options: "male", "female", "other", and "prefer not to say". Finally, to indicate their political orientation, participants rated themselves on a scale that ranged from 1 (*very left-wing*) to 7 (*very right-wing*).

After answering the demographic questions, participants in all conditions were presented with the following information: "Online misinformation is false or inaccurate information spread on the internet. It can have damaging effects on society, so it is important to be able to discriminate between true and false information online". Participants were then given clear instructions for the task that followed, which depended on the condition they had been randomly assigned to. In the baseline condition, participants were immediately presented with the test phase, which required participants to rate the veracity of 36 news headlines by moving a slider on a scale that ranged from 1 (*certainly false*) to 7 (*certainly true*; see Figure 2). Participants received no feedback for their ratings in the test phase.
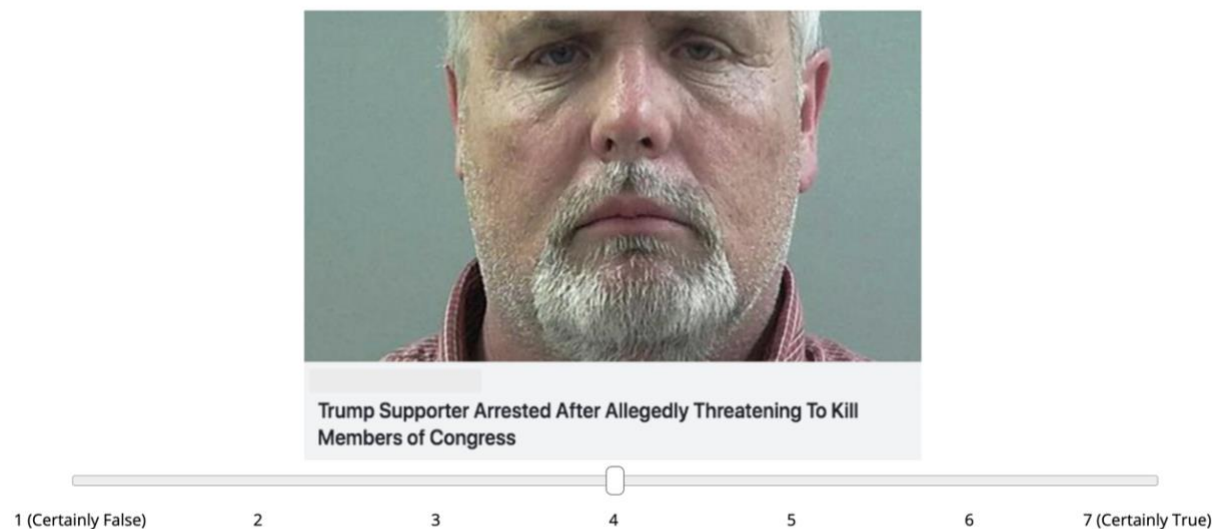
Figure 2

Example of a Fake News Item (Top) and a True News Item (Bottom) Presented in the Test Phase

In the non-gamified, gamified, and Bad News conditions, participants were first presented with their corresponding training phase. The training phase for the non-gamified and the gamified conditions required participants to rate the veracity of 42 news headlines

by clicking either a "true" or a "false" button. These headlines were presented to participants in three separate blocks, each containing 14 headlines, to give participants small breaks and prevent tedium. Participants were given feedback on whether their ratings were correct after each headline and were also given a total score out of 14 after finishing each block. The training phase for the Bad News condition consisted of playing the Bad News game.[3] After the training phase, participants in the Bad News condition had to type in a completion code that was only given to them after they had finished the game. They also had to indicate whether they would replay the training phase in their own time by clicking either a "yes" or a "no" button, which was also required from participants in the non-gamified and gamified conditions. After this, participants in these three conditions were finally presented with the test phase.

The time it took to complete the experiment depended on the condition participants were randomly assigned to. The baseline condition, the non-gamified and gamified conditions, and the Bad News condition each generally took around 10 mins, 15 mins, and 25 mins, respectively. However, since progression was entirely self-paced at every stage of each condition, the completion time varied greatly between participants. Finishing the test phase marked the end of the study for each condition, after which all participants were debriefed and redirected back to the Prolific website.

Data Analysis Plan

Preregistration and Statistical Software

All statistical analyses were carried out using R, and aside from the exploratory item-level analysis, they were all were preregistered on the AsPredicted website (https://aspredicted.org/2FK_XS6).

Main Analysis

---

[3] Considering the substantial number of studies conducted on Prolific that involved playing Bad News, prior experience with the game perhaps should have been controlled for. This could have been done by adding a question prior to playing Bad News that asked participants if they had played the game before, and we recommend this for future research.

SDT was used to analyze the data from this experiment. SDT is used when two different types of stimuli are being discriminated, which in the context of this study are the true and fake news headlines presented in the test phase (Stanislaw & Todorov, 1999). To quantify sensitivity, namely the ability to discriminate between different stimuli, several hit rates and false alarm rates were calculated for each participant (Velden & Clark, 1979). The hit rate was defined as the proportion of true news items that participants accurately categorized as true, and the false alarm rate was defined as the proportion of fake news items that participants inaccurately categorized as true. Since ratings were made on a scale, the hit rate and false alarm rate were obtained by treating each level of the scale (i.e., 1, 2, 3 …) as a cut-off point, specifically a point on the scale that corresponds to hypothetical yes/no criteria (Mandrekar, 2010).

Ratings in the test phase were measured on a scale that ranged from 1 (*certainly false*) to 7 (*certainly true*). In this case, a "yes" response indicated that a headline was being categorized as true, while a "no" response indicated that a headline was being categorized as false. So, if 4 was the cut-off point, participants that rated true news headlines 4 or above would be categorizing these headlines as true, and these responses would be called hits. In contrast, participants that rated fake news headlines 4 or above would also be categorizing these headlines as true, and these responses would be called false alarms. This was done for each point on the scale, which resulted in seven hit rates and seven false alarm rates for every participant.

Each participant's hit rates were plotted as a function of their false alarm rates to create individual ROC curves. The area under the curve (AUC) was calculated for every participant, which provided a measure of sensitivity. This was done using the trapezoidal rule formula obtained from Pollack & Hsieh (1969) shown in Equation 1.

$$AUC = 0.5 \sum_{k=0}^{n} (HR_{k+1} + HR_k)(FAR_{k+1} - FAR_k) \tag{1}$$

In this equation, *k* denotes the different criteria plotted on the ROC curve, *n* represents the total number of criteria, *HR* signifies the hit rate, and *FAR* indicates the false alarm rate. A

one-way ANOVA with 4 levels (one for each condition) was used to compare participants'

AUC values. To evaluate the evidence for both the null hypothesis (i.e., condition does not

influence AUC values) and the alternative hypothesis (i.e., condition does influence AUC

values), the Bayes factor was calculated. This was done using the anovaBF function from

the BayesFactor R package (Version 0.9.12-4.2; Morey & Rouder, 2018), which uses the

Jeffreys-Zellner-Siow default prior. Although the scaling factor for the prior can be adjusted,

we used the default $r = .5$ to provide a common frame of reference that allows for

comparisons between studies (Keysers et al., 2020). For a detailed discussion on default

Bayes factor tests for ANOVA designs, see Rouder et al. (2012). The Bayes factor was

interpreted through the discrete evidence categories proposed by Jeffreys (1961) and their

corresponding interpretations adapted by Lee and Wagenmakers (2013; see Table 1).

Table 1

Bayes Factor Evidence Categories According to Jeffreys (1961) and Their Corresponding

Interpretations Adapted by Lee and Wagenmakers (2013)

BF_{10}	Interpretation
>100	Extreme evidence for H1
30–100	Very strong evidence for H1
10–30	Strong evidence for H1
3–10	Moderate evidence for H1
1–3	Anecdotal evidence for H1
1	No evidence
1/3–1	Anecdotal evidence for H0
1/10–1/3	Moderate evidence for H0
1/30–1/10	Strong evidence for H0
1/100–1/30	Very strong evidence for H0
<1/100	Extreme evidence for H0

Note. BF_{10} quantifies the empirical evidence in favor of the alternative hypothesis.

To examine response bias, the average hit and false alarm rates across participants in each condition were calculated. This calculation resulted in seven hit rates and seven false alarm rates for each condition, which were plotted to create four ROC curves. To carry out descriptive statistics, the average AUC value was calculated for the full sample and for each condition.

Batailler et al. (2021) also used SDT to analyze true and fake news discrimination. Specifically, they reanalyzed data from two previous studies, Pennycook et al. (2018) and Pennycook and Rand (2019). However, their SDT reanalyses were limited to computing single-point indices of discrimination (d') and bias (c), which carry the strong assumption of equal-variance Gaussian underlying evidence distributions. To our knowledge, our study is the first in the context of research on fake news to use more powerful ROC analysis, which does not make the same strong assumptions. Indeed, Batailler et al. recommended using ROC analysis in future research.

Unlike Batailler et al. (2021), most prior work has simply compared participants' mean reliability ratings by condition for true and fake news items (e.g., Roozenbeek & van der Linden, 2019). Comparisons of mean ratings cannot reveal separate effects of an intervention on discrimination and response bias in the same way that ROC analysis can. Nonetheless, for consistency with previous literature, we have also conducted an analysis of mean ratings to facilitate comparisons between the results of our research and those of previous studies. Contrary to prior research (Basol et al., 2020; Maertens et al., 2021; Roozenbeek et al., 2020), this analysis showed that the condition participants were randomly assigned to had no significant effect on their mean reliability ratings of either true or fake news items. Please see our additional materials at https://osf.io/dnfrh/ for more details.

Exploratory Analyses

Demographic Analysis. Pearson correlations were conducted to examine the association between participants' AUC values (high AUC scores indicate better discrimination) and their age, political orientation, and gender. These were carried out on the

full sample to determine the general relationships between these demographic variables and the ability to discriminate between true and fake news headlines.

Training Phase Analysis. A Pearson's chi-squared test was conducted to determine whether participants' willingness to replay the training phase depended on the condition to which they were randomly assigned. Furthermore, SDT was also adopted to analyze the data from the non-gamified and gamified training phases. However, since they involved binary responses, ROC curves could not be plotted. In their place, each participant's hit rate, false alarm rate, and corrected hit rate were calculated. The hit rate was the proportion of true news items that participants correctly categorized as "true", and the false alarm rate was the proportion of fake news items that participants incorrectly categorized as "true". The corrected hit rate is another sensitivity index in SDT that is appropriate for discrimination indicated by binary responses (Stanislaw & Todorov, 1999). It is calculated by subtracting the false alarm rate from the hit rate. The change in participants' corrected hit rate across the three training blocks was examined, as well as how this change differed between the two conditions. To do this, a 3 (training blocks) x 2 (conditions) mixed model ANOVA was conducted.

Item-Level Analysis. The differences in average reliability ratings between each experimental condition and the baseline condition were plotted as a function of the average reliability ratings for the baseline condition. This was done to compare the item-level effects of each intervention to that of the baseline. To determine whether these difference scores varied significantly between experimental groups, two one-way between-subjects ANOVAs were conducted. The first ANOVA compared the difference scores for fake news items, and the second ANOVA compared the difference scores for true news items.

Results

Main Analysis

SDT analysis was conducted on participants' veracity ratings for the 36 news headlines presented in the test phase. The means and standard deviations for the AUC values in the full sample and in each condition are shown in Table 2. To test the present

study's hypothesis, a one-way between-subjects ANOVA was conducted to compare the effect of condition (i.e., baseline, non-gamified, gamified, or Bad News) on participants' ability to discriminate between true and fake news headlines (i.e., AUC values). Results showed that the main effect of condition on AUC values was not significant, $F < 1$. Furthermore, the Bayes factor indicated very strong evidence in favor of the null hypothesis (no effect of condition on AUC values), $BF_{10} = 0.026$ (please see our additional materials at https://osf.io/dnfrh/ for the full posterior distributions).[4] This indicates that the condition participants were randomly assigned to had no significant effect on their ability to discriminate between true and fake news headlines. The ROC curves for each condition are shown in Figure 3.

Table 2

Means and Standard Deviations of AUC Values in the Full Sample and in Each Condition

AUC values	Full sample	Baseline condition	Non-gamified condition	Gamified condition	Bad News condition
M	.80	.79	.79	.81	.79
SD	.14	.14	.12	.15	.15

Note. AUC = Area under the curve.

---

[4] If a study is underpowered, the Bayes factor will be indeterminate, showing equal or almost equal odds for both the null and the alternative hypotheses (see Rosenfeld & Olson, 2021 and Schönbrodt & Wagenmakers, 2018). Therefore, the fact that our analysis yielded a Bayes factor that indicated very strong evidence for the null hypothesis suggests that our study was sufficiently powered.

Figure 3

ROC Curves for the Baseline, Non-Gamified, Gamified, and Bad News Conditions
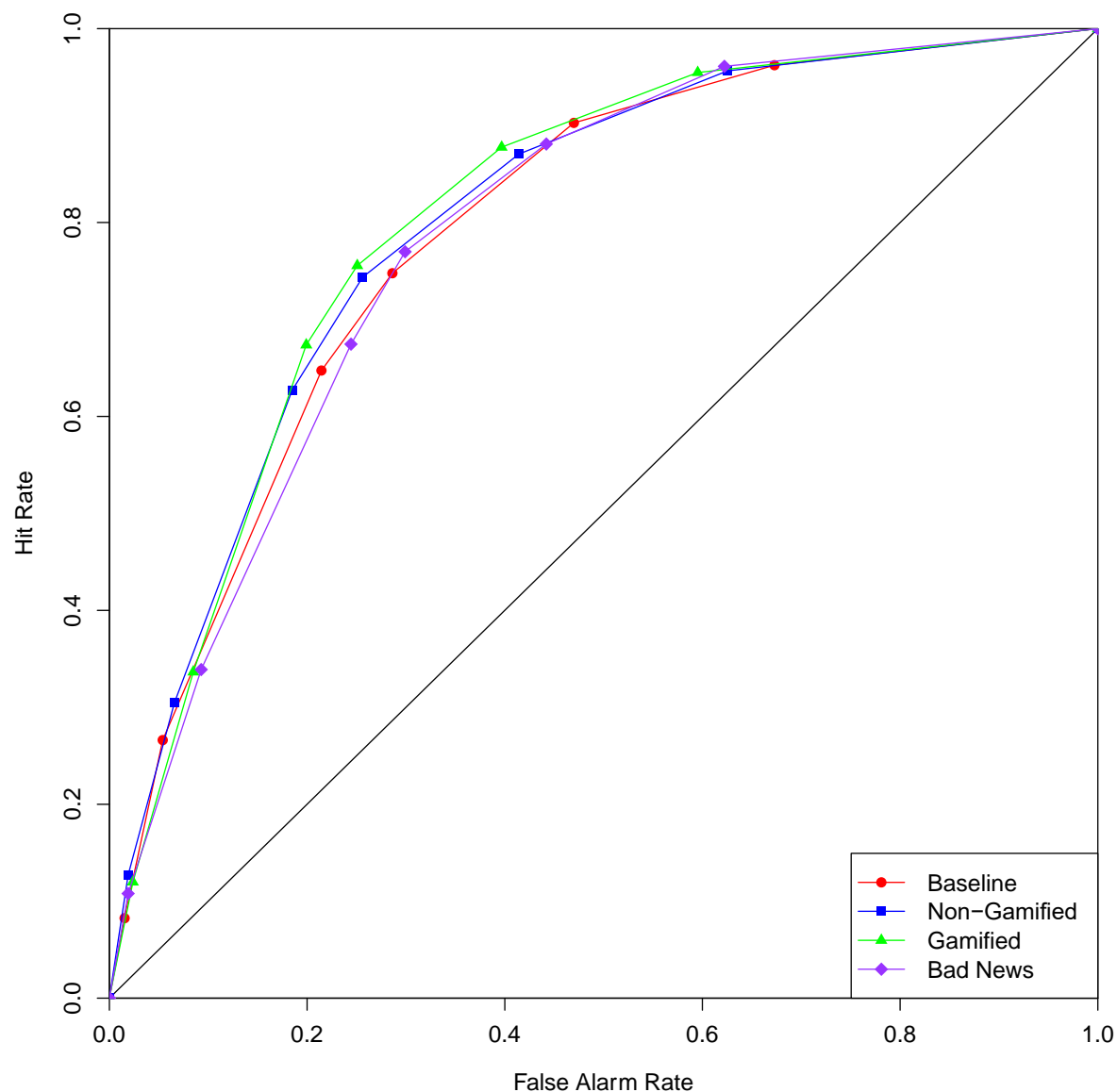
Note. ROC = Receiver operating characteristic curve. The hit rates are the proportion of true news items assigned each level of the reliability scale or higher, with the highest scale values in the bottom-left portion of the curves. The false alarm rates are the analogous proportions for fake news items. All curves meet at the (1, 1) point because all items are assigned scale value 1 or higher due to the nature of the rating task.

Exploratory Analyses

Demographic Analysis

The means and standard deviations for age and political orientation in the full sample and in each condition are shown in Table 3, and the frequencies of each gender option in the full sample and in each condition are shown in Table 4. To determine whether participants' age, political orientation, or gender were associated with their AUC values, three Pearson correlations were conducted on the full sample. Participants' age and AUC values were weakly negatively correlated, $r(280) = -.12$, $p = .048$, which suggests that being younger was associated with a greater ability to discriminate between true and fake news headlines. Participants' political orientation and AUC values were moderately negatively correlated, $r(280) = -.34$, $p < .001$. Since political orientation was measured on a scale that ranged from 1 (*very left-wing*) to 7 (*very right-wing*), this result indicates that being more left-wing was associated with a greater ability to discriminate between true and fake news headlines. Finally, participants' gender and AUC values were weakly positively correlated, $r(274) = .14$, $p = .016$, which suggests that being female was associated with a greater ability to discriminate between true and fake news headlines. The scatterplots representing these correlations are shown in Figure 4.

Table 3

Means and Standard Deviations of Age and Political Orientation in the Full Sample and in Each Condition

Demographic variable	Full sample	Baseline condition	Non-gamified condition	Gamified condition	Bad News condition
Age					
M	31.84	30	33.67	29.87	33.87
SD	9.60	9.16	10.02	9.15	9.48
Political orientation					
M	3.26	3.5	3.11	3.37	3.06
SD	1.64	1.45	1.57	1.76	1.76

Note. Political orientation was measured on a scale that ranged from 1 (*very left-wing*) to 7 (*very right-wing*).
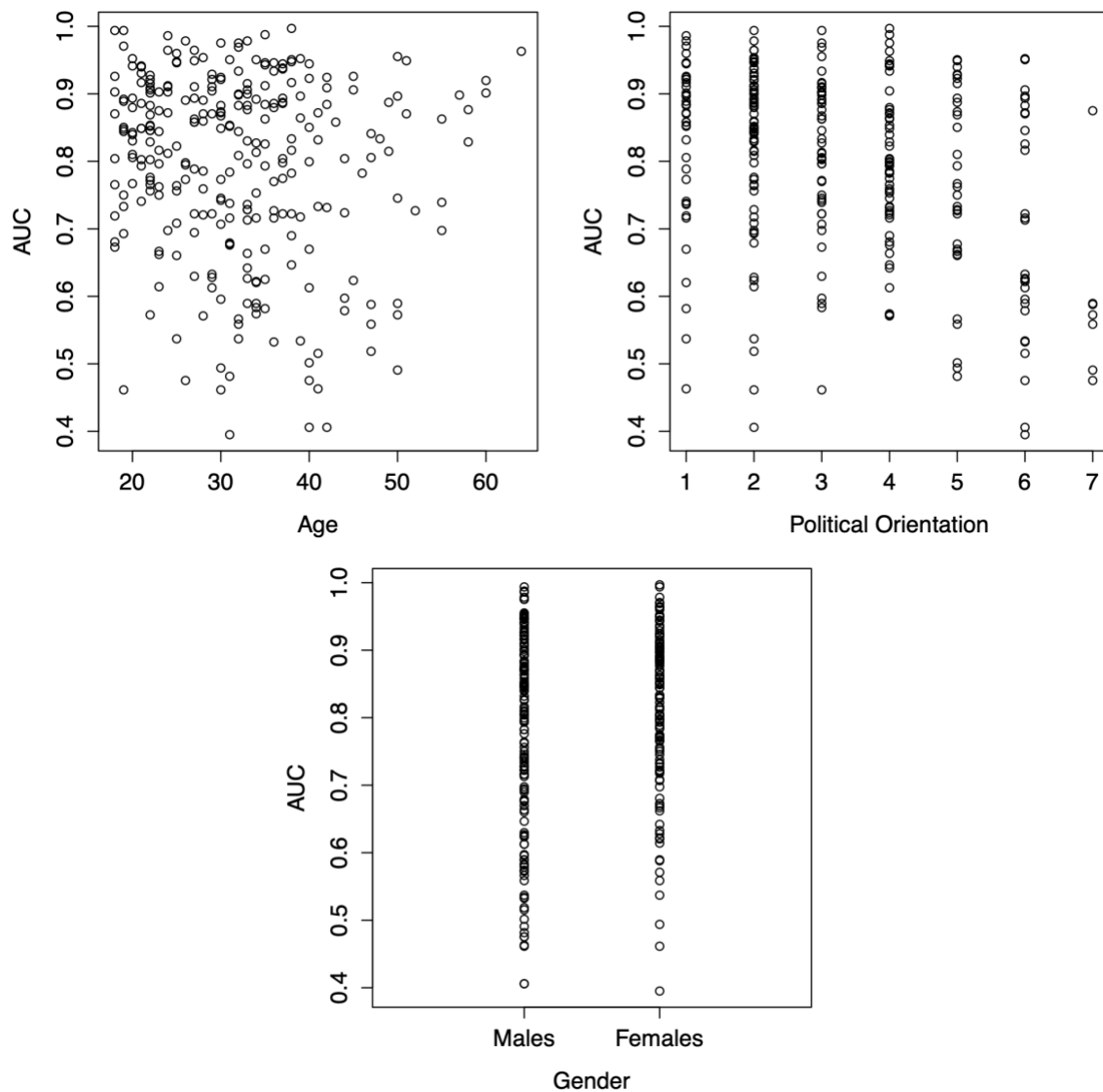
Table 4

Frequencies of Each Gender Option in the Full Sample and in Each Condition

Gender option	Full sample		Baseline condition		Non-gamified condition		Gamified condition		Bad News condition	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Male	148	52.48	40	55.56	30	42.86	37	52.86	41	58.57
Female	128	45.39	31	43.06	39	55.71	30	42.86	28	40.00
Other	5	1.77	1	1.39	1	1.43	2	2.86	1	1.43
PNTS	1	0.04	0	0	0	0	1	1.43	0	0

Note. PNTS = Prefer not to say.

Figure 4

Scatterplots for Age, Political Orientation, and Gender in the Full Sample

Training Phase Analysis

Participants' willingness to replay the training phase in the non-gamified, gamified, and Bad News conditions is shown in Table 5. A Pearson's chi-squared test revealed that willingness to replay was dependent on the condition to which participants were randomly assigned, $\chi^2(2, N = 282) = 9.38$, $p = .01$, V = .21. Specifically, willingness to replay was highest in the non-gamified condition, followed by the Bad News condition, and finally by the gamified condition.

Table 5

Contingency Table Showing the Number of Participants Willing Versus Unwilling to Replay the

Training Phase in the Non-Gamified, Gamified, and Bad News Conditions

Willingness to replay	Non-gamified condition		Gamified condition		Bad News condition	
	n	%	*n*	%	*n*	%
Willing to replay	48	68.57	30	42.86	39	55.71
Unwilling to replay	22	31.43	40	57.14	31	44.29

The means and standard deviations for the hit rates, false alarm rates, and corrected

hit rates in the non-gamified and gamified training phases are shown in Table 6. To establish

whether participants' performance changed across the three blocks in the non-gamified and

gamified training phases, and whether this change differed between the two conditions, a 3

(blocks) x 2 (conditions) mixed model ANOVA was conducted. The main effect of block on

corrected hit rates was significant, $F(2, 138) = 7.49$, $p < .001$, $\eta_p^2 = 0.10$, which indicates that

participants' performance worsened across the three blocks in the non-gamified and

gamified training phases. Post hoc paired *t*-tests with Bonferroni adjusted alpha levels of

.016 (.05/3) showed that performance in the first block ($M = .60$, $SD = .28$) was not

significantly different from performance in the second block ($M = .57$, $SD = .31$), $t(139) =$

1.34, $p = .18$, $d = 0.10$, 95% CI [-0.047, 0.25]. However, performance in the first block was

significantly better than performance in the third block ($M = .51$, $SD = .35$), $t(139) = 3.48$, $p <$

.001, $d = 0.27$, 95% CI [0.12, 0.43]. Performance in the second block was also significantly

better than performance in the third block, $t(139) = 2.50$, $p = .014$, $d = 0.17$, 95% CI [0.035,

0.31]. The main effect of condition on corrected hit rates was not significant, $F < 1$, which

indicates that performance across the three blocks did not differ significantly between the

non-gamified and gamified conditions. Finally, the interaction was not significant, $F < 1$.

Table 6

Means and Standard Deviations of the Hit Rates, False Alarm Rates, and Corrected Hit Rates in the

Non-Gamified and Gamified Training Phases

Training block	Non-gamified training phase			Gamified training phase		
	HR	FAR	CHR	HR	FAR	CHR
1	.77(.16)	.18(.17)	.59(.27)	.75(.19)	.14(.18)	.61(.29)
2	.74(.17)	.17(.23)	.57(.31)	.74(.20)	.17(.21)	.57(.32)
3	.72(.20)	.24(.23)	.48(.35)	.75(.19)	.21(.25)	.55(.35)

Note. HR = Hit rate. FAR = False alarm rate. CHR = Corrected hit rate. Standard deviations are

presented in parentheses.

Item-Level Analysis

To run an item-level analysis analogous to the one reported in Figure 1, the

differences in average reliability ratings between each experimental condition and the

baseline condition were plotted as a function of the average reliability ratings for the baseline

condition (see Figure 5). We did this analysis to compare the item-level effects of each

intervention to that of the baseline. Three important statistics can be observed from Figure 5:

(a) in the non-gamified condition, 17% (3/18) of fake items and 56% (10/18) of true items

were rated as more reliable than baseline; (b) in the gamified condition, 17% (3/18) of fake

items and 61% (11/18) of true items were rated as more reliable than baseline; and (c) in the

Bad News condition, 56% (10/18) of fake items 67% (12/18) of true items were rated as

more reliable than baseline. These statistics, along with visual inspection of Figure 5,

suggest that the non-gamified and gamified conditions improved discrimination at the item

level: Following the intervention, participants generally rated fake items as less reliable but

true items as more reliable than baseline. Note that this pattern was observed even for fake

items with low baseline reliability ratings. As the scale does not allow for much of a decrease

in reliability ratings for these items, it is particularly noteworthy that very few differences were

positive. In contrast, Bad News did not seem to improve discrimination at all, as it simply

biased participants towards indicating that items were more reliable regardless of their objective veracity.

Figure 5

Item-Based Scatterplots Showing the Effect of the Non-Gamified (Top-Left Panel), Gamified (Top-Right Panel), and Bad News (Bottom Panel) Interventions on Reliability Ratings as a Function of Baseline Reliability Rating
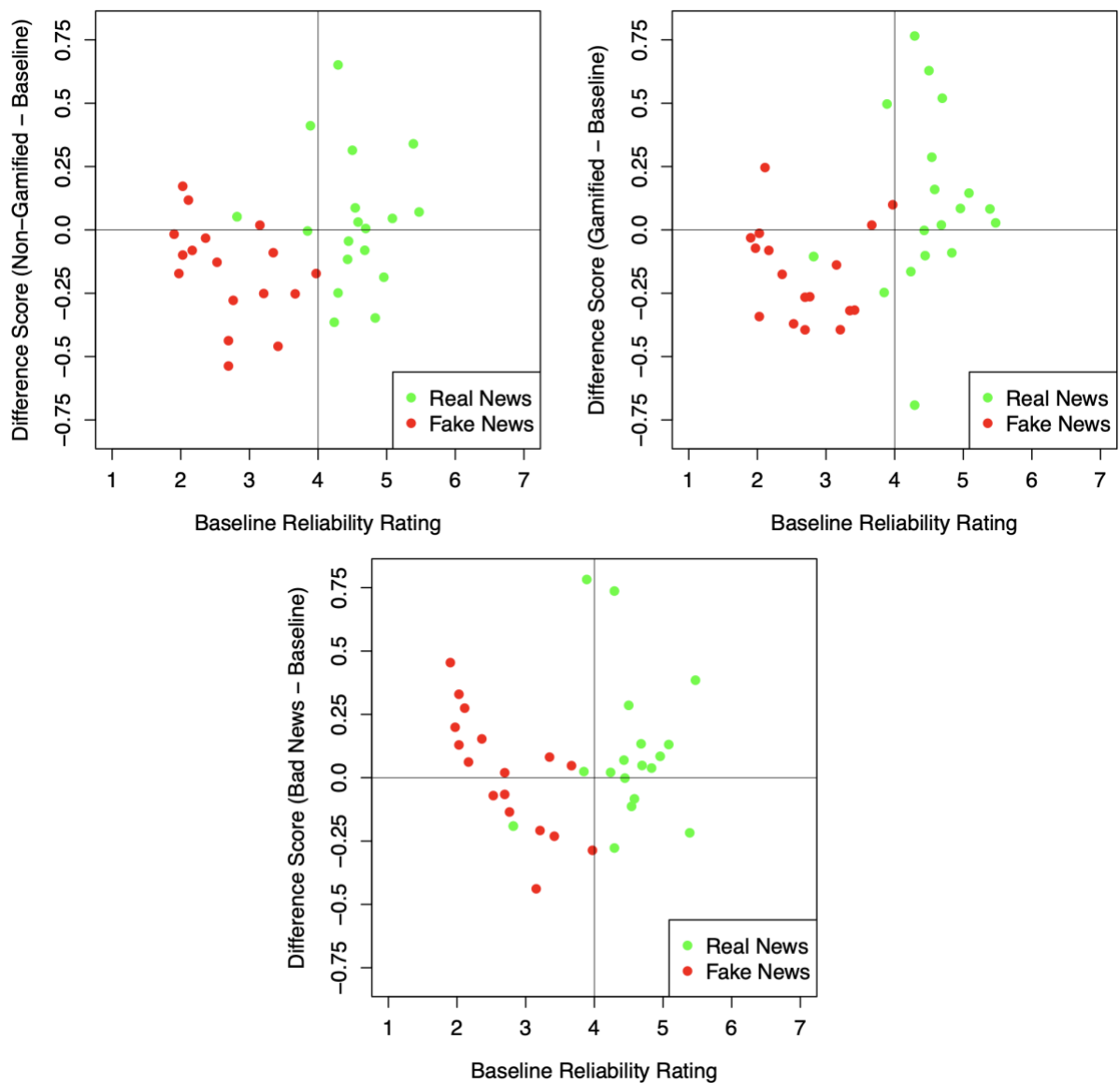
Note. Negative versus positive difference scores plotted on the y-axis indicate that, compared to baseline, an intervention decreased versus increased reliability ratings, respectively.

To determine whether these difference scores varied significantly between experimental groups, two one-way between-subjects ANOVAs were conducted. The first ANOVA compared the difference scores for fake news items and yielded a significant main effect of condition, $F(2, 51) = 3.74$, $p = .031$, $\eta_p^2 = 0.13$. Post hoc independent t-tests with Bonferroni adjusted alpha levels of .016 (.05/3) showed that the non-gamified ($M = -0.15$, $SD = 0.20$) and gamified ($M = -0.16$, $SD = 0.18$) conditions were not significantly different, $t(33.87) = 0.24$, $p = .81$, $d = 0.079$, 95% CI [-0.60, 0.76]. The non-gamified and Bad News ($M = 0.0069$, $SD = 0.23$) conditions were marginally significantly different, $t(33.07) = -2.15$, $p = .039$, $d = -0.72$, 95% CI [-1.41, -0.017], as were the gamified and Bad News conditions, $t(32.34) = -2.42$, $p = .021$, $d = -0.81$, 95% CI [-1.51, -0.10]. Thus, compared to the baseline condition: (a) participants in the non-gamified and gamified conditions tended to assign fake news items lower reliability ratings; (b) participants in the Bad News condition assigned fake news items reliability ratings that were (on average) the same; and (c) the difference between these difference scores when comparing the non-gamified condition versus the Bad News condition, and the gamified condition versus the Bad News condition, were both marginally significant. The second ANOVA examined the difference scores for true news items and yielded a non-significant main effect, $F < 1$. This indicates that difference scores for true news items remained constant across all experimental conditions.

Discussion

Results Summary

The present study yielded several noteworthy findings. Regarding the main analysis, there were no significant differences in true and fake news discrimination between any of the four conditions. This contradicted our hypothesis that the gamified intervention would be the most effective, followed by its non-gamified equivalent, then Bad News, and finally the control group. As for the exploratory analyses, there were four key results: (a) being younger, female, and left-wing was associated with better discrimination; (b) discrimination deteriorated across training blocks in the non-gamified and gamified training phases; (c) participants were most willing to replay the non-gamified, Bad News, and gamified training

phases, in that order; and (d) at the item level, the non-gamified and gamified interventions lowered reliability ratings to fake news items (relative to baseline) more than the Bad News intervention, but the effects were only marginally significant. There was no difference between the conditions for true news items.

The Application of Inductive Learning

A multitude of studies have supported the effectiveness of inductive learning for improving discernment between categories, particularly when factors such as immediate feedback and interleaved sequencing are applied (Brosvic & Epstein, 2007; Kornell & Vaughn, 2018; Sana et al., 2017). Therefore, the finding that the present study's novel psychological intervention was not effective in improving discrimination between true and fake news (barring some marginal effects in the item analysis) was unexpected. To determine the potential reasons for this unanticipated outcome, the relevant literature was examined further, participants' responses in the control group were scrutinized to determine their baseline discrimination ability, and the non-gamified and gamified training phases were analyzed to assess the inductive learning process they applied.

To begin with, studies that have used inductive learning to promote discernment between categories have typically implemented longer and more intensive training schedules. For example, Carvalho and Goldstone (2012) trained participants to discriminate between three different categories of shapes. The learning phase involved four training blocks, in which participants were presented with eight exemplars of each category two times. Therefore, participants were trained with a total of 32 exemplars of each category, all of which were repeated twice. In comparison, the present study's non-gamified and gamified interventions trained participants with 21 exemplars of each category, which were only shown once. This consideration is related to the more general concern of whether short interventions can have long-term effects on people's beliefs or biases, which are key determinants of news veracity discernment (Pennycook & Rand, 2021). In light of this issue, recent research has scrutinized the effectiveness of brief psychological interventions in learning contexts and found them to be less effective than previously thought, or not

effective at all (Kizilcec et al., 2020; Serra-Garcia et al., 2020). Consequently, lengthening and intensifying the inductive learning process, which can be done by adding more exemplars of each category and repeating their presentation more than once, may be necessary to improve discrimination between true and fake news. This course of action seems particularly appropriate given the results of the item-level analysis, which showed that participants in the non-gamified and gamified conditions generally rated true items as more reliable and fake items as less reliable compared to baseline. Although these effects were only marginally significant, lengthening and intensifying the learning process may increase them to the point that they are meaningful.

Notably, the finding that inductive learning did not improve discrimination between true and fake news raises two important considerations. The first is whether there are any features present in true and fake news that allow people to discriminate between them, and the second is whether inductive learning can enhance this discrimination. The fact that inductive learning did not achieve the latter in our experiment does not mean that discriminable features between true and fake news do not exist. Indeed, participants were quite adept at discriminating between true and fake news without any training in the control group, so there must be certain distinguishing features to which people are sensitive. If such features were absent, then the ROC curves would trace the chance diagonal, yielding an AUC of .5, which was not the case in our study (see Figure 3).

Furthermore, participants were quite accurate at discriminating between the true and fake news items in the non-gamified and gamified training phases as well as in each condition's test phase. Naturally, discrimination will depend on the particular stimuli chosen for training and testing. Nonetheless, these results support the idea that in some circumstances, people are quite adept at true and fake news discernment. This result stands in direct contrast to the previous literature showing that people's discrimination is only just above chance level (Luo et al., 2020). If people's news veracity discernment is actually quite good, the present study's assumption that true and fake news are high similarity categories can be challenged; since they are easily discriminated, they could be low similarity

categories. Accordingly, the most effective type of sequencing in this setting may be blocking rather than interleaving, as it would highlight the commonalities as opposed to the differences between true and fake news (Carvalho & Goldstone, 2014; Kornell & Vaughn, 2018; Sana et al., 2017). Critically, there is no prior research on this topic that can guide the decision to use either blocking or interleaving to enhance discrimination between true and fake news. So, to tackle this open-ended issue, future studies could use both types of sequencing, as learning both the commonalities and differences between categories would provide an integrated way of discriminating between them.

Finally, news veracity discernment deteriorated across training blocks in the non-gamified and gamified training phases. Specifically, participants' discrimination was significantly worse in the third block compared to the first and the second block. There are two possible explanations for this finding. The first is that participants became increasingly bored or fatigued as the training continued, which caused their mind to start wandering (Xu & Metcalfe, 2016). However, mind wandering has been shown to be reduced when the training material is easy and when interleaved sequencing is used (Eglington & Kang, 2017; Feng et al., 2013). As this was the case for the non-gamified and gamified training phases, it seems unlikely that this was the main issue that led to deterioration in discernment across training blocks, especially considering that they only lasted around 10 minutes each. The second explanation for this finding is thus more plausible, which is that the items in the third training block were simply more difficult than those in the preceding blocks. Since the training blocks were not counterbalanced, this possibility cannot be ruled out and serves as a limitation of the present study, albeit a minor one as training phase discrimination was not the main focus of the analysis. Nevertheless, this shortcoming should be corrected in future research, as it would allow for more concrete assumptions to be made about participants' decision-making patterns during training.

The Application of Game Design Elements

Prior research has supported the effectiveness of using gamification in learning environments, especially for promoting motivation and engagement (Boudadi & Gutiérrez-

Colón, 2020; Dong et al., 2012; Pesare et al., 2016). However, the present study did not demonstrate any benefits of using game design elements in non-game contexts. Most importantly, the gamified intervention did not improve discernment between true and fake news over its non-gamified equivalent. Although this could have been due to a more general issue, such as the inductive learning process that these two training phases applied, it is worth investigating how gamification may have contributed to this finding. Thus, the following question can be considered: Is it worth investing time and resources into gamifying a learning experience if it does not yield better results than a non-gamified one? To answer this query, it is necessary to determine whether gamification has any other benefits that make it worth using.

For this purpose, participants' willingness to replay each of the training phases was examined. Although this does not directly measure feelings like enjoyment, motivation, and engagement, it can serve as a general indication of them; if someone is motivated and engaged during learning, they will most likely enjoy the experience and thus be more inclined to say that they would take part in it again. Contrary to what was expected, participants were least willing to replay the gamified training phase. Therefore, not only did the gamified intervention generate the same results as its non-gamified counterpart, but participants were also less willing to replay it. This suggests that the specific game design elements used to gamify the present study's novel psychological intervention were essentially of no use.

The potential for gamification to have negative effects on learning or simply no effects at all has been corroborated by several studies. Some have shown that gamification can demotivate learners and thus lead to worse performance, while others have indicated that it can yield identical results to traditional learning experiences (de-Marcos et al., 2014; Hanus & Fox, 2015). Crucially, these studies applied gamification in a similar way as was done in the current research; the game design elements were added as a superficial afterthought that did not significantly change the learning experience. For example, in the present study's gamified training phase, achievement badges were provided regardless of

performance. Participants were aware of this since they were explicitly told by the game's narrator that they were being awarded badges for completing training blocks. Such completion-contingent rewards have been found to be less effective than performance-contingent rewards; they demotivate players by reducing their perceived autonomy, namely their perceived level of control over the game's outcomes (Jovanovic & Matejevic, 2014; Park et al., 2019). Therefore, it may be necessary make game design elements contingent on players' actions, as this can add weight to their in-game decisions and thus elicit motivation and engagement. Alternatively, they could be made seemingly contingent on players' actions even when they are not. For example, to make completion-contingent rewards appear performance-contingent, explanations as to why participants received them could be omitted.

The Design of Bad News

Various studies have concluded that Bad News is effective for improving people's ability to identify online misinformation (Basol et al., 2020; Maertens et al., 2021; Roozenbeek & van der Linden, 2019; Roozenbeek et al., 2020). However, several methodological issues cast doubt over their findings. To be specific, one of these studies lacked a control group, most of them adopted pre-post designs that introduced memory confounds (Knapp, 2016, but see Roozenbeek et al., 2020 and Maertens et al., 2021, Experiment 3, for evidence suggesting that these may not be as relevant as previous literature has suggested), and all of them used incomparable true and fake news items to assess participants' fake news detection. To elaborate on this last issue, the true news items were likely familiar to participants as they had been reported by the mainstream media, whereas the fake news items were completely unfamiliar to participants as they were created by researchers. Due to these drawbacks pertaining to the previous literature, it is difficult to conclude with certainty whether Bad News improves discernment between true and fake news.

Consequently, we tested the effectiveness of Bad News in a more methodologically robust experiment. The results showed that Bad News did not improve people's ability to

identify online misinformation. Although this could have been due to a more general issue, such as the overall design of the game, it is worth exploring the more specific and manageable factors that may have contributed to this finding. One such factor is that Bad News only presents examples of one category (i.e., fake news) rather than two categories (i.e., true and fake news). Most research that has successfully improved discernment between two categories has done so by presenting examples of both categories (Higgins & Ross, 2011; Pérez-Gay et al., 2017). This allows for within- and between-category comparisons, which have both been found to be indispensable for category learning (Hammer et al., 2008). The results of the item-level analysis can be seen as evidence for the importance of these comparisons. Compared to baseline, the non-gamified and gamified interventions, which incorporated such contrasts, showed a tendency to lower reliability ratings to fake news items at the item-level. Conversely, Bad News, which did not incorporate any contrasts, showed no such tendency (see Figure 5). Therefore, psychological interventions that aim to improve discernment between true and fake news should present both types of news items. This would allow for comparisons to be drawn within and between the two categories, which would in turn promote discrimination.

Critics may argue that Bad News was not designed to improve news discernment per se, but rather to train people to detect manipulative strategies in fake news. With such a goal in mind, discriminating between true and fake news is less important than ensuring that people successfully identify one or more manipulative strategies in fake news when they are present. Consequently, the fact that our test items were not selected or manipulated according to any of the six misinformation techniques presented in Bad News may be seen as a limitation of our study. After all, Bad News is designed to teach people about these specific techniques, so it might be expected that its efficacy can only be replicated with items that contain these features. However, our aim was to find interventions that work "in the wild", and we thus saw little value in selecting or manipulating items to ensure the success of any intervention. Furthermore, in our view, Bad News and our inductive learning procedure are both psychological interventions intended to protect internet users from the problems

associated with fake news. Regardless of the intervention or its proposed mechanisms, we believe it is of paramount importance to assess its *generality*. Some interventions may affect only the intended behavior, whereas others may affect the intended behavior as well as unintended behaviors. Depending on the behaviors, the latter could have serious consequences.

To illustrate this point by drawing an analogy to medicine, consider an intervention aimed at improving oncological diagnosis. A group of oncologists are provided with some training, and afterwards, they correctly identify tumors in cancer patients with enhanced frequency. Although this may seem like a successful intervention, it is only half the story. If, after the intervention, those same oncologists are also falsely identifying cancerous tumors in healthy patients at an enhanced rate, unnecessarily referring them for major surgery to remove internal organs, for weeks of unnecessary chemotherapy, or for a host of other debilitating but unnecessary cancer treatments, then the success of the intervention would be rightfully called into question.

To us, the oncology problem can be analogized to the fake news problem. If an intervention decreases belief in fake news but, at the same time, it decreases belief in true news, then its efficacy should be rightfully questioned too, as is the case with Bad News. However, it is important to clarify that Bad News causes participants to rate true news as less reliable, but not *unreliable*. The consequences of reduced reliability have not been empirically investigated and are thus open to conjecture. For example, one might argue that increased scepticism could lead to positive outcomes, such as higher willingness to fact-check information. At present, this is an empirical question that warrants further research. However, it is important to recognize that reduced belief in truth (e.g., that COVID booster vaccines are important) can sometimes be a worse problem than increased belief in falsehoods (e.g., that 5G towers cause COVID). Indeed, conspiracy theorizing has been likened to a "hybrid of skepticism" (Aupers, 2012, p. 30), highlighting the potential dangers of decreasing people's belief in truth or, in other words, increasing people's skepticism towards truth. Hence, we believe that to properly assess the efficacy of *any* intervention in the

context of the fake news problem, one must show that it reduces the likelihood of believing

fake news at the same time as showing belief in true news is unaffected (or even better, it is

increased). This is the reason we use discernment or discrimination as our main dependent

variable and why we would argue that ROC analysis is indispensable in this field (see also

Batailler et al., 2021). ROC analysis allows researchers to determine whether an intervention

has specific influences that affect discernment versus general influences that bias responses

to all types of news.

It is worth noting that participants were more willing to replay Bad News than the

gamified training phase. This could have resulted from the two performance-contingent

game design elements included in the former. The first involved a follower count that

denoted the number of virtual followers a player had gained, and the second involved a

credibility scale that indicated how credible a player was perceived to be by their virtual

followers. Both features depended entirely on the player's performance. Such dependency

typically increases a player's perceived control over the outcomes of the game, which in turn

usually increases their motivation to learn (Jovanovic & Matejevic, 2014; Park et al., 2019).

Nevertheless, participants were still more willing to replay the non-gamified training phase

than Bad News. This may have been due to the completion-contingent rewards included in

the latter. These involved fixed achievement badges that were given to players irrespective

of their performance, which was also done in the gamified training phase. The use of this

type of reward could have reduced the benefits of the follower count and credibility scale,

since regardless of how many followers a player achieved or how credible they managed to

become, they received the same rewards. As a result, the completion-contingent rewards

most likely minimized the benefits of the performance-contingent game design elements,

reducing participants' perceived control and motivation to learn while playing Bad News.

The Demographic Predictors of News Veracity Discernment

Studies that have assessed the relationship between age and news veracity

discernment have generally found them to be significantly negatively correlated, which is

precisely what was found in the present study (Brashier & Schacter, 2020; Guess et al.,

2019; Pehlivanoglu et al., 2021). This finding may be due to an age-related decline in various cognitive abilities, such as executive control and attention (Verhaeghen & Cerella, 2002). The former involves processes like response inhibition, which denotes the ability to control automatic responses and replace them with conscious ones (Diamond, 2013). The latter refers to the ability to focus on either one specific stream of information (i.e., selective attention) or multiple streams of information (i.e., divided attention; Hahn et al., 2008). Deficits in cognitive functions such as these can make older adults more susceptible to false information, deception, and cognitive biases (Boyle et al., 2013; Jacoby, 1999). For example, Skurnik et al. (2005) found older adults were more vulnerable to the illusory truth effect, which refers to the phenomenon that repeating misinformation increases the likelihood that people will believe it (Fazio et al., 2019). In light of these findings, future studies should investigate how to tackle this age-dependent worsening of news veracity discernment, perhaps by developing certain psychological interventions tailored for older adults.

Research on the relationship between gender and news veracity discernment is scarce, and the findings on this topic are contradictory; some have shown no significant differences in performance between males and females, while others have shown females to perform significantly better than males (Roozenbeek & van der Linden, 2019; Roozenbeek et al., 2020). The present study supported the latter result as it found that being female was significantly correlated with better discrimination between true and fake news. However, due to the inconsistent results on this matter, further research is necessary to determine whether the present study's finding is robust.

Most studies that have investigated the link between political leaning and news veracity discernment show that right-wing individuals perform significantly worse than left-wing individuals (Michael & Breaux, 2021), which was replicated in the present study. This may be due to differences in analytic reasoning, which refers to rational, deliberate, and rule-based thinking (Witteman et al., 2009). This is opposite to intuitive reasoning, which refers to instinctive thinking governed by gut feelings (Pennycook et al., 2015). Crucially, research

suggests that Republicans have a lower propensity for analytic reasoning than Democrats, and that analytic reasoning is associated with better news veracity discernment (Pennycook & Rand, 2018, 2019). Therefore, it appears that right-wing individuals are worse at applying analytic reasoning, which in turn impairs their ability to discriminate between true and fake news. Accordingly, future research should explore the possibility of creating psychological interventions designed to promote analytic reasoning for the purpose of improving news veracity discernment.

ROC Analysis: Theoretical and Practical Considerations

ROC analysis can only be used in situations where it is possible to assume the objective states of the world. In the context of the current paper, this would be the assumption that true and fake news actually represent true and fake information, respectively. The need to assume the objective states of the world can be problematic in situations that do not follow such assumptions. For example, consider the following headline published by the Chicago Tribune in 2021: "A 'healthy' doctor died 2 weeks after receiving the COVID-19 vaccine; CDC is investigating why" (Boryga, 2021). Although the information in this headline is factual, it can be considered misleading as it suggests that the doctor died because of the vaccine, for which there is no evidence. Therefore, the decision regarding whether this headline falls into the category of true or fake news can be viewed as subjective. With headlines such as these, applying ROC analysis may be impractical when the objective is to discern truth from falsity.

Importantly, however, this issue is not unique to ROC analysis; it is a problem for any research that purports to prebunk or debunk fake news. Put simply, how can we improve people's ability to detect online misinformation if we cannot define online misinformation in the first place? In our view, there are two main solutions to this issue. The first is to select items that can be unambiguously defined as true or fake, which is what we attempted to do in the present paper. For example, it is categorically false that Nancy Pelosi's son was arrested for murder, and it is categorically true that Trump donated $1,000,000 to recovery efforts in Texas after a hurricane struck in 2017. The second solution would be to change

the aim of the intervention from discerning truth to discerning misleadingness. For example, researchers could select a number of factual but misleading headlines and determine whether inductive learning helps people discriminate these items from factual, non-misleading headlines. This would be an interesting avenue for future research.

Concluding Remarks

At present, there is a lack of research aimed at using cognitive learning principles to improve people's ability to discriminate between true and fake news. This is surprising considering the recent rise in scientific interest on the topic of online misinformation and people's inability to identify it. The current study aimed to fill this scientific lacuna by proposing a novel theoretical framework and psychological intervention designed to improve true and fake news discernment. The theoretical framework involves inductive learning and gamification, and the psychological intervention applies these concepts by presenting exemplars of both true and fake news in a game-like environment. Contrary to what was hypothesized, this novel intervention did not improve discernment between true and fake news.

Inductive learning and gamification have never been applied together before to tackle the issue of online misinformation, so despite the results of the present study, we are not immediately ruling them out as ineffective. Instead, we suggest that there are several correctable issues with the way in which the two concepts were applied. We have offered various changes that could be made to the inductive learning process and game design elements that were implemented in the novel psychological intervention. These are supported by the literature and could thus inform future research that attempts to incorporate these two concepts in this setting.

Overall, online misinformation is a relatively novel area of inquiry for psychological research. Therefore, studies that provide novel techniques for tackling fake news and test them in methodologically robust experiments, as was done in the present paper, will consistently lead to positive developments in the field. They will not only enhance our limited knowledge on the topic, but also contribute towards the goal of developing an effective

theoretical framework and psychological intervention for improving discernment between true and fake news. Achieving such objectives is vital, as they could reduce the damaging real-world consequences being caused by online misinformation and people's inability to identify it.

References

Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The

contribution of enjoyment to students' continuing interest in learning about science.

Contemporary Educational Psychology, *36*(1), 4–12.

https://doi.org/10.1016/j.cedpsych.2010.08.001

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

Aupers, S. (2012). 'Trust no one': Modernization, paranoia and conspiracy culture. *European

Journal of Communication*, *27*(1), 22–34.

https://doi.org/10.1177%2F0267323111433566

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news:

Gamified inoculation boosts confidence and cognitive immunity against fake news.

Journal of Cognition, *3*(1), Article 2. https://doi.org/10.5334/joc.91

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2021). A signal detection

approach to understanding the identification of fake news. *Perspectives on

Psychological Science*, *17*(1), 79–98. https://doi.org/10.1177/1745691620986135

Birnbaum, M. S. (2013). *Understanding and optimising the inductive learning of categories

and concepts* [Doctoral dissertation, University of California, Los Angeles].

eScholarship. https://escholarship.org/uc/item/8396m4g0

Birnbaum, M. S., Kornell, N., E. L. Bjork., & R. A. Bjork. (2013). Why interleaving enhances

inductive learning: The roles of discrimination and retrieval. *Memory and Cognition*,

41(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7

Bright, J. E. H., & Burton, A. M. (1994). Past midnight: Semantic processing in an implicit

learning task. *The Quarterly Journal of Experimental Psychology Section A*, *47*(1),

71–89. https://doi.org/10.1080/14640749408401144

Boryga, A. (2021, April 8). *A 'healthy' doctor died two weeks after getting a COVID-19

vaccine; CDC is investigating why*. Chicago Tribune.

https://www.chicagotribune.com/coronavirus/fl-ne-miami-doctor-vaccine-death-20210107-afzysvqqjbgwnetcy5v6ec62py-story.html

Boudadi, N. A., & Gutiérrez-Colón, M. (2020). Effect of gamification on students' motivation and learning achievement in second language acquisition within higher education: A literature review 2011–2019. *The EuroCALL Review*, *28*(1), 57–69. https://doi.org/10.4995/eurocall.2020.12974

Boyle, P. A., Yu, L., Wilson, R. S., Segawa, E., Buchman, A. S., & Bennett, D. A. (2013). Cognitive decline impairs financial and health literacy among community-based older persons without dementia. *Psychology and Aging*, *28*(3), 614–624. https://doi.org/10.1037/a0033103

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5), Article e2020043118. https://doi.org/10.1073/pnas.2020043118

Brashier, N. M., & Schacter, D. L. (2020). Aging in an era of fake news. *Current Directions in Psychological Science*, *29*(3), 316–323. https://doi.org/10.1177/0963721420915872

Brauer, S., Siklander, P., & Ruhalahti, S. (2017). Motivation in digital open badge-driven learning in vocational teacher education. *Journal of Professional and Vocational Education, 19*(3), 7–23. https://www.researchgate.net/publication/320711739_Motivation_in_Digital_Open_Badge-Driven_Learning_in_Vocational_Teacher_Education

Britt, M. A., Rouet, J.-F., Blaum, D., & Millis, K. (2019). A reasoned approach to dealing with fake news. *Policy Insights From the Behavioral and Brain Sciences*, *6*(1), 94–101. https://doi.org/10.1177%2F2372732218814855

Brosvic, G. M., & Epstein, M. L. (2007). Enhancing learning in the introductory course. *The Psychological Record*, *57*(3), 391–408. https://doi.org/10.1007/bf03395584

Burki, T. (2020). The online anti-vaccine movement in the age of COVID-19. *The Lancet Digital Health*, *2*(10), e504–e505. https://doi.org/10.1016/s2589-7500(20)30227-2

Carvalho, P. F., & Goldstone, R. L. (2012). Category structure modulates interleaving and

    blocking advantage in inductive category acquisition. *Proceedings of the Annual*

    Meeting of the Cognitive Science Society, *34*, 186–191.

    https://escholarship.org/uc/item/0nw0d367

Carvalho, P. F., & Goldstone, R. L. (2014). 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

Chadwick, A., & Vaccari, C. (2019). *News sharing on UK social media: Misinformation,*

    disinformation, and correction. Online Civic Culture Centre - Loughborough

    University. https://www.lboro.ac.uk/research/online-civic-culture-centre/news-

    events/articles/o3c-1-survey-report-news-sharing-misinformation/

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in

    a web browser. *Behavior Research Methods*, *47*(1), 1–12.

    https://doi.org/10.3758/s13428-014-0458-y

de-Marcos, L., Domínguez, A., Saenz-de-Navarrete, J., & Pagés, C. (2014). An empirical

    study comparing gamification and social networking on e-learning. *Computers &*

    Education, *75*, 82–91. https://doi.org/10.1016/j.compedu.2014.01.012

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to

    gamefulness. *Proceedings of the 15th International Academic MindTrek Conference*

    on Envisioning Future Media Environments, 9–15.

    https://doi.org/10.1145/2181037.2181040

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

    https://doi.org/10.1146/annurev-psych-113011-143750

Dong, T., Dontcheva, M., Joseph, D., Karahalios, K., Newman, M., & Ackerman, M. (2012).

    Discovery-based games for learning software. *Proceedings of the SIGCHI*

    Conference on Human Factors in Computing Systems, 2083–2086.

    https://doi.org/10.1145/2207676.2208358

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, *71*(1), 5–24. https://doi.org/10.1037/0022-3514.71.1.5

Eglington, L. G., & Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 475–485. https://doi.org/10.1016/j.jarmac.2017.07.005

Fazio, L., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin & Review*, *26*(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*(3), 586–592. https://doi.org/10.3758/s13423-012-0367-y

Filgona, J., Sakiyo, J., Gwany, D. M., & Okoronka, A. U. (2020). Motivation in learning. *Asian Journal of Education and Social Studies*, *10*(4), 16–37. https://doi.org/10.9734/ajess/2020/v10i430273

Grady, R. H., Ditto, P. H., & Loftus, E. F. (2021). Nevertheless, partisanship persisted: Fake new warnings help briefly, but bias returns with. Time. *Cognitive Research: Principles and Implications*, *6*, Article 52. https://dx.doi.org/10.1186%2Fs41235-021-00315-z

Guess, A. M., & Lyons, B. A. (2020). Misinformation, disinformation, and online propaganda. In N. Persily & J. A. Tucker (Eds.), *Social media and democracy: The state of the field, prospects and reform* (pp. 10–33). Cambridge University Press. https://doi.org/10.1017/9781108890960

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of

fake news dissemination on Facebook. *Science Advances*, *5*(1), Article eaau4586.

https://doi.org/10.1126/sciadv.aau4586

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

Hahn, B., Wolkenberg, F. A., Ross, T. J., Myers, C. S., Heishman, S. J., Stein, D. J., Kurup,

P. K., & Stein, E. A. (2008). Divided versus selective attention: Evidence for common

processing mechanisms. *Brain Research*, *1215*, 137–146.

https://doi.org/10.1016/j.brainres.2008.03.058

Hakulinen, L., Auvinen, T., & Korhonen, A. (2015). The Effect of achievement badges on

students' behavior: An empirical study in a university-level computer science course.

International Journal of Emerging Technologies in Learning (iJET), *10*(1), 18–29.

https://doi.org/10.3991/ijet.v10i1.4221

Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison

processes in category learning: From theory to behavior. *Brain Research*, *1225*, 102–

118. https://doi.org/10.1016/j.brainres.2008.04.079

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A

longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and

academic performance. *Computers & Education*, *80*, 152–161.

https://psycnet.apa.org/doi/10.1016/j.compedu.2014.08.019

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to

compare for what. *Proceedings of the Annual Meeting of the Cognitive Science

Society*, *33*. https://escholarship.org/uc/item/7zz764kx

Higham, P. A. (1997). Dissociations of grammaticality and specific similarity effects in

artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory,

and Cognition*, *23*(4), 1029–1045. https://doi.org/10.1037/0278-7393.23.4.1029

Higham, P. A., & Brooks, L. R. (1997). Learning the experimenter's design: Tacit sensitivity to the structure of memory lists. *The Quarterly Journal of Experimental Psychology Section A*, *50*(1), 199–215. https://doi.org/10.1080/027249897392297

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

Huotari, K., & Hamari, J. (2016). A definition for gamification: Anchoring gamification in the service marketing literature. *Electronic Markets*, *27*(1), 21–31. https://doi.org/10.1007/s12525-015-0212-z

Jacoby, L. L. (1999). Deceiving the elderly: Effects of accessibility bias in cued-recall performance. *Cognitive Neuropsychology*, *16*(3-5), 417–436. https://doi.org/10.1080/026432999380861

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

Jolley, D., & Douglas, K. M. (2017). Prevention is better than cure: Addressing anti-vaccine conspiracy theories. *Journal of Applied Social Psychology*, *47*(8), 459–469. https://doi.org/10.1111/jasp.12453

Jovanovic, D., & Matejevic, M. (2014). Relationship between rewards and intrinsic motivation for learning - Researches review. *Procedia - Social and Behavioral Sciences*, *149*, 456–460. https://doi.org/10.1016/j.sbspro.2014.08.287

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Psychological and Cognitive Sciences*, *117*(26), 14900–14905. https://doi.org/10.1073/pnas.1921417117

Knapp, T. R. (2016). Why is the one-group pretest–posttest design still used? *Clinical Nursing Research*, *25*(5), 467–472. https://doi.org/10.1177/1054773816666280

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

Kruschke, J. K. (2005). Category learning. In K. Lamberts & R. L. Goldstone (Eds.), *Handbook of Cognition* (pp. 183-201). Sage Publications Ltd. https://doi.org/10.4135/9781848608177

Kyza, E. A., Varda, C., Panos, D., Karageorgiou, M., Komendantova, N., Perfumi, S. C., Shah, S. I. H., & Hosseini, A. S. (2020). Combating misinformation online: Re-imagining social media for policy-making. *Internet Policy Review*, *9*(4). https://doi.org/10.14763/2020.4.1514

Lange, K., Kühn, S., & Filevich, E. (2015). Correction: "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS ONE*, *10*(7), Article e0134073. https://doi.org/10.1371/journal.pone.0134073

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modelling: A practical course*. Cambridge University Press. https://psycnet.apa.org/doi/10.1017/CBO9781139087759

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization:

    The effects of prior theories on subsequently considered evidence. *Journal of*

    Personality and Social Psychology, *37*(11), 2098–2109. https://doi.org/10.1037/0022-

    3514.37.11.2098

Luo, M., Hancock, J. T., & Markowitz, D. M. (2020). Credibility perceptions and detection

    accuracy of fake news headlines on social media: Effects of truth-bias and

    endorsement cues. *Communication Research*. Advance online publication.

    https://doi.org/10.1177/0093650220921321

Maertens, R., Götz, F. M., Schneider, C. R., Roozenbeek, J., Kerr, J. R., Stieger, S.,

    McClanahan, W. P., Drabot, K., & van der Linden, S. (2022). *The Misinformation*

    Susceptibility Test (MIST): A psychometrically validated measure of news veracity

    discernment. PsyArXiv. https://doi.org/10.31234/osf.io/gk68h

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term

    effectiveness of inoculation against misinformation: Three longitudinal experiments.

    Journal of Experimental Psychology: Applied, *27*(1), 1–16.

    https://psycnet.apa.org/doi/10.1037/xap0000315

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test

    assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316.

    https://doi.org/10.1097/jto.0b013e3181ec173d

McCright, A. M., & Dunlap, R. E. (2017). Combatting misinformation requires recognizing its

    types and the factors that facilitate its spread and resonance. *Journal of Applied*

    Research in Memory and Cognition, *6*(4), 389–396.

    http://dx.doi.org/10.1016/j.jarmac.2017.09.005

Michael, R. B., & Breaux, B. O. (2021). The relationship between political affiliation and

    beliefs about sources of "fake news." *Cognitive Research: Principles and*

    Implications, *6*(1), Article 6. https://doi.org/10.1186/s41235-021-00278-1

Miller, R. R. (2021). Failures of memory and the fate of forgotten memories. *Neurobiology of*

    Learning and Memory, *181*, Article 107426.

      https://doi.org/10.1016/j.nlm.2021.107426

Modirrousta-Galian, A., Higham, P. A., & Seabrooke, T. (2022, September 14). Effects of

    Inductive Learning and Gamification on News Veracity Discernment.

      https://doi.org/10.17605/OSF.IO/DNFRH

Modirrousta-Galian, A., & Higham, P. A. (2022). *How effective are gamified fake news*

    interventions? Reanalyzing existing research with signal detection theory. PsyArXiv.

      https://doi.org/10.31234/osf.io/4bgkd

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for

    common designs (Version 0.9.12-4.2) [R package].

      https://www.rdocumentation.org/packages/BayesFactor/versions/0.9.12-4.2

Mourão, R. R., & Robertson, C. T. (2019). Fake news as discursive integration: An analysis

    of sites that publish false, misleading, hyperpartisan and sensational information.

    Journalism Studies, *20*(14), 2077–2095.

      https://doi.org/10.1080/1461670X.2019.1566871

Musgrove, A. T., Powers, J. R., Rebar, L. C., & Musgrove, G. J. (2018). Real or fake?

    Resources for teaching college students how to identify fake news. *College &*

    Undergraduate Libraries, *25*(3), 243–260.

      https://doi.org/10.1080/10691316.2018.1480444

Neil, G. J., & Higham, P. A. (2012). Implicit learning of conjunctive rule sets: An alternative to

    artificial grammars. *Consciousness and Cognition*, *21*(3), 1393–1400.

      https://doi.org/10.1016/j.concog.2012.07.005

Neil, G. J., & Higham, P. A. (2020). Repeated exposure to exemplars does not enhance

    implicit learning: A puzzle for models of learning and memory. *Quarterly Journal of*

    Experimental Psychology, *73*(3), 309–329.

      https://doi.org/10.1177/1747021819873838

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises.

    Review of General Psychology, *2*(2), 175–220. https://doi.org/10.1037/1089-

    2680.2.2.175

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

Park, J., Kim, S., Kim, A., & Yi, M. Y. (2019). Learning to be better at the game:

    Performance vs. completion contingent reward for game-based learning. *Computers

    & Education*, *139*, 1–15. https://doi.org/10.1016/j.compedu.2019.04.016

Pehlivanoglu, D., Lin, T., Chi, K., Perez, E., Polk, R., Cahill, B. S., Lighthall, N. R., & Ebner,

    N. C. (2021). *Aging in an "infodemic": The role of cognition, affect, and experience on

    news veracity detection during the COVID-19 pandemic*. PsyArXiv.

    https://doi.org/10.31234/osf.io/3kgq9

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of

    analytic thinking. *Current Directions in Psychological Science*, *24*(6), 425–432.

    https://doi.org/10.1177/0963721415604610

Pennycook, G., & Rand, D. G. (2018). Cognitive reflection and the 2016 U.S. presidential

    election. *Personality and Social Psychology Bulletin*, *45*(2), 224–239.

    https://doi.org/10.1177/0146167218783192

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news

    is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*,

    39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive

    Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pérez-Gay, F., Thériault, C., Gregory, M., Sabri, H., Rivas, D., & Harnad, S. (2017). How

    and why does category learning cause categorical perception? *International Journal

    of Comparative Psychology*, *30*. https://doi.org/10.46867/ijcp.2017.30.01.01

Pesare, E., Roselli, T., Corriero, N., & Rossano, V. (2016). Game-based learning and

gamification to promote engagement and motivation in medical learning contexts.

Smart Learning Environments, *3*(1), Article 5. https://doi.org/10.1186/s40561-016-

0028-0

Pew Research Centre (2019). *Many Americans say made-up news is a critical problem that*

needs to be fixed [Fact sheet]. https://www.pewresearch.org/journalism/wp-

content/uploads/sites/8/2019/06/PJ_2019.06.05_Misinformation_FINAL-1.pdf

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC-curve of d'e.

Psychological Bulletin, *71*(3), 161–173.

https://psycnet.apa.org/doi/10.1037/h0026862

Pratama, L. D., & Setyaningrum, W. (2018, May 7–8). *Game-based learning: The effects on*

student cognitive and affective aspects [Paper presentation]. The 5th International

Conference on Research, Implementation, & Education of Mathematics and

Sciences, Yogyakarta, Indonesia.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self

versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369–381.

https://doi.org/10.1177/0146167202286008

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and*

Verbal Behavior, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Roediger, H. L., III., & Marsh, E. J. (2005). The positive and negative consequences of

multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, *31*(5), 1155–1159. https://doi.org/10.1037/0278-7393.31.5.1155

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological

resistance against online misinformation. *Palgrave Communications*, *5*, Article 65.

https://doi.org/10.1057/s41599-019-0279-9

Roozenbeek, J., Maertens, R., McClanahan, W., & van der Linden, S. (2020). Disentangling

item and testing effects in inoculation research on online misinformation: Solomon

revisited. *Educational and Psychological Measurement*, *81*(2), 340–362.

https://doi.org/10.1177/0013164420940378

Rosenfeld, J. P., & Olson, J. M. (2021). Bayesian data analysis: A fresh approach to power

issues and null hypothesis interpretation. *Applied Psychophysiology and*

Biofeedback, *46*(2), 135–140. https://doi.org/10.1007/s10484-020-09502-y

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes

factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374.

https://doi.org/10.1016/j.jmp.2012.08.001

Rozenshtein, A., Pearson, G. D. N., Yan, S. X., Liu, A. Z., & Toy, D. (2016) Effect of massed

versus interleaved teaching method on performance of students in radiology. *Journal*

of the American College of Radiology, *13*(8)*,* 979–984.

https://doi.org/10.1016/j.jacr.2016.03.031

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning

of cognitive concepts. *Journal of Educational Psychology*, *109*(1), 84–98.

https://doi.org/10.1037/edu0000119

Sanchez, D. R., Langer, M., & Kaur, R. (2020). Gamification in the classroom: Examining the

impact of gamified quizzes on student learning. *Computers & Education*, *144*, Article

103666. https://doi.org/10.1016/j.compedu.2019.103666

Schönbrodt, F. D., & Wagenmakers, E. -J. (2018). Bayes factor design analysis: Planning for

compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142.

https://doi.org/10.3758/s13423-017-1230-y

Serra-Garcia, M., Hansen, K. T., & Gneezy, U. (2020). Can short psychological interventions

affect educational performance? Revisiting the effect of self-affirmation interventions.

Psychological Science, *31*(7), 865–872.

https://doi.org/10.1177%2F0956797620923587

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims

become recommendations. *Journal of Consumer Research*, *31*(4), 713–724.

https://doi.org/10.1086/426605

Spoelstra, S. (2020). The truths and falsehoods of post-truth leaders. *Leadership*, *16*(6),

    757–764. https://doi.org/10.1177/1742715020937886

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

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and

    intelligence. *Current Directions in Psychological Science*, *22*(4), 259–254.

    https://doi.org/10.1177%2F0963721413480174

Toda, A. M., Klock, A. C. T., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., Bittencourt,

    I., Gasparini, I., Isotani, S., & Cristea, A. I. (2019). Analysing gamification elements in

    educational environments using an existing gamification taxonomy. *Smart Learning

    Environments*, *6*(1), Article 16. https://doi.org/10.1186/s40561-019-0106-1

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the

    public against misinformation about climate change. *Global Challenges*, *1*(2), Article

    1600008. https://doi.org/10.1002/gch2.201600008

Velden, M., & Clark, W. C. (1979). Reduction of rating scale data by means of signal

    detection theory. *Perception & Psychophysics*, *25*(6), 517–518.

    https://doi.org/10.3758/bf03213831

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of

    meta-analyses. *Neuroscience & Biobehavioral Reviews*, *26*(7), 849–857.

    https://doi.org/10.1016/s0149-7634(02)00071-4

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

Walter, N., & Tukachinsky, R. (2019). A meta-analytic examination of the continued

    influence of misinformation in the face of correction: How powerful is it, why does it

    happen, and how to stop it? *Communication Research*, *47*(2), 155–177.

    https://doi.org/10.1177/0093650219854600

Watson, A. (2020). *Share of people who have ever shared fake news or information on social media in the United States as of December 2020* [Fact Sheet]. https://www.statista.com/statistics/657111/fake-news-sharing-online/

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*(1), 10–16. https://doi.org/10.1111/sjop.12093

Witteman, C., van den Bercken, J., Claes, L., & Godoy, A. (2009). Assessing rational and intuitive thinking styles. *European Journal of Psychological Assessment*, *25*(1), 39–47. https://doi.org/10.1027/1015-5759.25.1.39

Wright, R. L., & Burton, A. M. (1995). Implicit learning of an invariant: Just say no. *The Quarterly Journal of Experimental Psychology Section A*, *48*(3), 783–796. https://doi.org/10.1080/14640749508401415

Xu, J., & Metcalfe, J. (2016). Studying in the region of proximal learning reduces mind wandering. *Memory & Cognition*, *44*(5), 681–695. https://doi.org/10.3758/s13421-016-0589-8

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918–933. https://doi.org/10.1037/xge0000177