How do people assess the truth of partially-familiar claims, relative to both familiar and unfamiliar claims?

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Philosophy in Psychology

at

The University of Waikato

by

Daniel McLennan

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2023

Abstract

People rate familiar claims as more true than unfamiliar claims, possibly because familiarity makes these claims feel easier to process. But little is known about how people assess the truth of claims that consist of a combination of familiar and unfamiliar information. It is important to know how people assess such "partially-familiar" claims because false claims often contain a familiar "kernel of truth" that may make the claim as a whole seem more believable. If this were the case, we would expect people to rate more partially-familiar claims true than unfamiliar claims, yet fewer than familiar claims. Across three experiments, we investigated how people assess the truth of partially-familiar claims, relative to both familiar and unfamiliar claims. We found that people classified more familiar claims as true than unfamiliar claims, and roughly the same number of partially-familiar claims true as unfamiliar claims. In other words, we replicated the known finding that people inflate their judgements of truth for familiar claims compared with unfamiliar claims; however, we found no evidence that a "kernel of truth" similarly inflates people's judgements. This research contributes to our understanding of how people decide whether a claim is true.

Acknowledgements

I am grateful for the advice, mentorship and supervision of Dr. Maryanne Garry, and all the support of the Garry Lab. I wish to thank Dr. Andrew Evelo for his essential work on this research project. I could not have completed this project without the love and support of my fianceé, Mallory Dobner. Lastly I wish to thank all the people who participated in these experiments, without whom this research would not have been possible.

Table of Contents

Introduction

In everyday life, people encounter claims we might characterise as "partially-familiar," such that part of a claim is familiar, and part is unfamiliar. Take, for example, the true claim, "the Nile River flows northwards." If people subsequently encounter the false claim "the Nile River is the longest river in Asia," we could characterise this claim as partially-familiar. In this example, the phrase "the Nile River" is familiar because it has been repeated, whereas the phrase "the longest river in Asia" is completely new. How might people assess the truth of such claims? The familiar part of these claims may make lead people to classify the claim as a whole as true, even if it is false. Could people be lulled into believing a false claim if only a small part of that claim feels familiar? That is the question we[1] address in the three experiments we present here.

Evidence from the truth effect literature

At least three literatures suggest people will classify more partially-familiar claims as true than unfamiliar claims. First, the truth effect literature shows that people classify more repeated claims—such as trivia statements or news headlines—as true than new claims, even if those claims are false (Dechêne et al, 2010; Hasher et al, 1977; Henderson et al, 2022). In a typical "truth effect" experiment, people encounter several true and false claims across two consecutive sessions. Some claims appear only during the first or the second session, whereas others are repeated across both sessions. In the second session, people rate more repeated claims as true than new claims, even if those claims are false. The truth effect is robust— it occurs with only

[1] Here, my use of the word "we" reflects that, although the research in this thesis is my own, I conducted it under advice and direction of my supervisor, post-doctoral scholars in the lab, and with assistance from my fellow doctoral students. For this reason, I often use the words "our" and "we" in this thesis. Elsewhere in this thesis I use the word "we" for a different reason: to refer to what is generally known in the scientific community.

one repetition, across delays ranging from a few minutes to several months, and even for claims that contradict people's prior knowledge (for a review, see Henderson et al, 2022).

Why do people classify more repeated claims as true than new claims? One explanation for the truth effect is that repeated claims feel more familiar than new claims, and this increased familiarity makes the repeated claims easier—or more fluent—to process (Newman et al., 2019; Reber & Schwarz, 1999). Consistent with this explanation, people rate more claims as true when they process those claims fluently for reasons other than repetition. For example, subjects in one study rated claims as more true when those claims appeared in high-contrast colours than low-contrast colours, likely because the increased readability led people to process those claims more fluently (Reber & Schwarz, 1999; Unkelbach, 2006). Likewise, people in another study rated rhyming proverbs as more true than non-rhyming proverbs with the same meaning ("woes unite foes" vs. "woes unite enemies") likely because they processed the rhyming statements more fluently (McGlone & Tofighbakhsh, 2000). This account of the truth effect fits with broader literature about fluency, in which people interpret fluent processing along a range of positive attributes (Alter & Oppenheimer, 2009). We might think of these fluency attributions as "source monitoring errors"—mistakes people make in determining where their thoughts and feelings come from (Johnson et al, 1993; Lindsay, 2008). These errors are are thought to occur because people assume that the feelings they have in the moment are caused by the task at hand (Higgins, 1998; Schwarz, 2011). Therefore, when people decide whether something is true, they assume feelings of familiarity and fluency are diagnostic of truth (Alter & Oppenheimer, 2009).

Evidence from the educational psychology and memory literatures

Second, the educational psychology literature suggests people learn new information better when they are first reminded of how that information relates to what they already know (Ausubel, 1960; Ausubel, 1978; Corkill, 1992). For example, in one study American undergraduates remembered more information from an essay about Buddhism (a relatively unfamiliar religion) when they first read an introductory passage that compared Buddhism with Christianity (a relatively familiar religion; Ausubel & Fitzgerald, 1961). In this example, the familiar information about Christianity is akin to the familiar part of a partially-familiar claim, making the unfamiliar information about Buddhism easier to remember. In another study, people were better able to comprehend and remember an ambiguous paragraph when that paragraph came with a title, because the title provided context that helped people connect the paragraph to their existing knowledge (Bransford & Johnson, 1972). These findings suggest that familiar concepts help people to process and bring to mind unfamiliar information more easily. In the same manner, people may classify more partially-familiar claims as true than unfamiliar claims, because the familiar part of those claims may pave the way for people to accept them as true.

Third, the memory literature suggests that people can incorporate familiar details into their memories of events, even if those details were not part of the original event (Brown et al., 1977; Loftus, 1976). For example, in several studies mock eyewitnesses were more likely to identify an innocent suspect in a photo lineup when that they had seen that person's photo in an earlier lineup (for a review, see Deffenbacher et al., 2006). Why? One possibility is that seeing a repeated photo leads to a feeling of familiarity and easy processing. If people cannot recall the source of this fluent processing, they may misattribute the familiar face to their memory of the

culprit (Kirsten & Earles, 2017; Memon et al., 2002). The memory literature also suggests people are vulnerable to classifying new stimuli as "old" on recognition tests if those stimuli contain familiar details (Devitt et al, 2016; Reinitz et al, 1992; Verde, 2010). Take, for example, one study in which people read a list of words containing "birdsong" and "blackmail" and then 15 min later falsely recognised reading "blackbird" (Reinitz et al, 1992). One explanation for these errors is that if people cannot recall the source of the two constituent words "black" and "bird", they will falsely recognise the conjunction, "blackbird," as old (Jones & Jacoby, 2001; Leding & Lampinen, 2009; Marsh et al., 2002). In support of this possibility, manipulations that make it more difficult to recall the details of the source—but do not affect familiarity—such as dividing attention during encoding, increase the number of conjunction errors (Jones & Jacoby, 2001). People make similar errors, albeit less frequently, when only one of the constituent words is familiar such as "blacksmith" (Jones et al., 2006; Reinitz et al., 1992). This step-down pattern from "full conjunctions" to "half conjunction" suggests a relationship between the extent to which people would falsely recognise a target and the "dose" of familiarity in that target. Collectively, then, this work also suggests that to the extent people also assess the truth of partially-familiar claims by some "dose" of familiarity, we would expect people to classify more partially-familiar claims as true than unfamiliar claims, yet fewer than familiar claims.

A referential account of the truth effect

Considering the truth effect, education, and memory literatures together, there is parallel evidence to suggest that people will classify more partially-familiar claims as true than unfamiliar claims—but less than familiar claims. But, there is also reason to expect a different pattern of results. Another account of the truth effect argues that people assess the truth of a

claim based on consistency with their existing knowledge (for a review, see Brashier & Marsh, 2020). In this "referential theory," a claim feels true if the references to facts in memory are coherently linked with each other (Unkelbach et al, 2019; Unkelbach & Rom, 2017). Because people tend to believe incoming information by default, repeating a claim strengthens the links between the semantic references to that claim (Hassan & Barber, 2021; Gilbert, 1991). In support of this theory, people rate familiar claims as less true if they can remember those claims coming from an untrustworthy source, and as more true if they initially processed those claims more deeply (Unkelbach & Rom, 2017). On average, familiar claims tend to have more coherently-linked references to facts in memory than new claims, and hence when people encounter those claims, they classify more of them as true than unfamiliar claims (Unkelbach & Rom, 2017).

How might people assess the truth of partially-familiar claims, according to this referential theory? At first glance it might seem as though the familiar part of a partially-familiar claim would lead people to process the whole claim more fluently (Arkes et al., 1991; Begg et al., 1985). But a closer look suggests that although the repeated part of such a claim is more familiar, it is not more coherently linked to the other references in the claim (Unkelbach & Rom, 2017). To the extent that a feeling of truth arises from coherently-linked references in memory, rather than from fluency, we would not expect people to classify partially-familiar claims as true more than unfamiliar claims (Unkelbach et al., 2019). For example, if people first read the true claim "the Nile River flows northward", and subsequently read the false claim "the Nile River is the longest river in Asia", their semantic reference to "the Nile River" will be become more familiar, but it will be disconnected from the rest of the information in the claim (Unkelbach & Rom, 2017). In summary, both the fluency model and referential theory of the truth effect make

similar predictions for how people assess the truth of familiar claims, but they make different predictions for how people assess the truth of partially-familiar claims.

Other literatures make similar predictions about partially-familiar claims via a different route. Consider two experiences that may arise from an experience of partial familiarity: déjà vu and the "butcher-on-the-bus" phenomenon (Brown, 2003; Brown, 2020; Mandler, 1980). In new situations people sometimes experience déjà vu— a "gut feeling" of having been there before (Brown, 2004). If we recast the literature on déjà vu in terms of partial familiarity, we might say déjà vu can arise when a new situation feels partially-familiar—such as if your friend's new apartment has the same spatial layout as your childhood home (Cleary et al, 2012; Lindsay, 2008). Similarly, people are less likely to recognise a familiar face outside of its typical context, such as seeing one's local butcher on the bus instead of behind the meat counter (Brown, 2020; Mandler, 1980). One explanation for this "butcher-on-the-bus" phenomenon is that the context with which a face usually appears allows people to easily bring to mind identifying details, and those details are less accessible absent the familiar context (Gruppuso et al, 2007). Experiences such as déjà vu and butcher-on-the-bus may prompt people to think more carefully about why part of a situation feels familiar (Schwartz et al, 2016). But, if by thinking more carefully they can recall source-specifying details, then the feeling is resolved (Brown, 2020; Lindsay, 2008).

To the extent that people experience partially-familiar claims similarly to déjà vu and butcher-on-the-bus, we might expect them to think more carefully about whether those claims are true. Several studies suggest that thinking more carefully about whether claims are true reduces the truth effect (Brashier et al., 2020; Jalbert et al., 2020; Salovich et al., 2022).

Therefore, even if people cannot recall source-specifying details about the partially-familiar claims, they may not classify any more of those claims as true than unfamiliar claims.

Taking everything together, the literature suggests two possibilities. On the one hand, research from the truth effect, educational psychology, and memory literatures suggest people will classify more partially-familiar claims as true than unfamiliar claims, yet fewer than familiar claims. On the other hand, the referential theory of the truth effect, as well as research on déjà vu and butcher-on-the-bus suggest that people would classify the same number of partially-familiar claims as true as unfamiliar claims.

Experiment 1

We developed a set of materials and a counterbalancing scheme to use in Experiments 1 - 3. The materials are 280 trivia claims, half of which are true and half of which are false (See Appendices 1, 2 and 5). We ran a stimulus validation study to determine how plausible these claims were when people encountered them for the first time (See Stimulus Development Study in Appendix 1). On the basis of this study, we excluded claims from Experiments 1 - 3 that subjects would overwhelmingly classify as true or false. That is, we excluded claims that fewer than 15% or more than 85% of subjects rated true in the Stimulus Validation Study. By excluding these claims, we aimed to avoid ceiling and floor effects, and make it more likely for people to rely on a feeling of familiarity, rather than their knowledge. We selected a subset of 160 claims to use in Experiments 1- 3. To avoid confounding the effect of familiarity with reading order, we also varied which part of the partially-familiar claims was repeated and which part was new. For half of these claims the subject of the sentence was the familiar part, and for half of the claims the verb and object of the sentence was the familiar part. In Experiments 1 - 3. we refer to the

former as “partially-familiar-subject” claims and to the latter as “partially-familiar-verb-object” claims (See Appendix 5 for full list of materials). The data for Experiments 1 - 3 are available at https://researchbox.org/1340.

Method

Subjects

We recruited subjects on the crowdsourcing platform Amazon Mechanical Turk (MTurk, www.mturk.com) using Cloud Research (Litman et al, 2017). We used the “Power Analysis for ANOVA: Simulation” Shiny Web App (Lakens & Caldwell, 2021) to detect a mean difference of 0.04 between familiar, partially-familiar, and unfamiliar claims with 90% power, using variances observed in pilot testing. This process resulted in a target sample size of 400 subjects. Due to the way Cloud Research interacts with MTurk, we ultimately collected responses from 578 subjects. Subjects participated in exchange for 0.75 USD in Amazon credit. This experiment was not pre-registered. We excluded the responses from 173 subjects for failing to complete the experiment, from a further 84 subjects for failing to describe the task they had just completed, and from a further 29 subjects for searching the claims online. Thus the final sample size was 292 subjects (196 women, 94 men, 2 gender diverse). Subjects ranged in age from 19 – 89 years old (M = 43.85, Med = 41, SD = 13.86) and 282 subjects (97%) spoke English as a first language.

Design

Experiment 1 followed a within-subjects design with four conditions, (Familiarity: familiar, partially-familiar-subject, partially-familiar-verb-object, unfamiliar). That is, the independent variable was the familiarity of the claims, with the levels familiar, partially-familiar-subject, partially-familiar-verb-object, unfamiliar; and the dependent variable was the proportion

of those claims subjects classified as true. We counterbalanced Experiment 1 so that any effect of familiarity on people's assessments of truth would not be tied to any specific trivia claims. We assigned subjects to one of 16 counterbalances; subjects encountered a different subset of 40 trivia claims in the exposure and test phase in each counterbalance. As a result of this counterbalancing scheme, each claim served equally often as a familiar, partially-familiar or unfamiliar claim for subjects in different counterbalance groups (see Appendix 2).

Procedure

We told subjects the experiment was about visual and verbal learning, and that they would answer some questions about trivia claims and complete some visual puzzles. There were two phases: exposure and test. In the exposure phase, subjects read a list of 40 trivia claims and rated how interesting each one was on a 5- point Likert scale. We asked subjects to judge how interesting the claims were so they would engage with the claims without thinking about whether those claims were true. We presented the claims one at a time in a random order. Each claim appeared in the centre of the screen for 4 s before the interest scale appeared below the claim. The claims automatically advanced after subjects responded by selecting a point on the scale. Then, there was a 10 min delay to reduce rehearsal of the claims. During the delay, subjects completed a series of "spot-the-difference" and "Where's Waldo?" puzzles (Davis, 2020; South, 2011). Then in the test phase, subjects read another list of trivia claims and classified them as true or false. We told subjects to provide their "best guess" if they were unsure. Subjects saw 10 familiar claims, 20 partially-familiar claims and 10 unfamiliar claims. As in the exposure phase, each claim appeared in the centre of the screen for 4 s before a true/false scale appeared below the claim, and the claims automatically advanced after subjects selected a response. Subjects'

completion times for Experiment 1 ranged from 18 min 18 s - 58 min 35 s (M = 29 min 10 s, Med = 27 min 44 s, SD = 6 min 20 s).

Results and Discussion

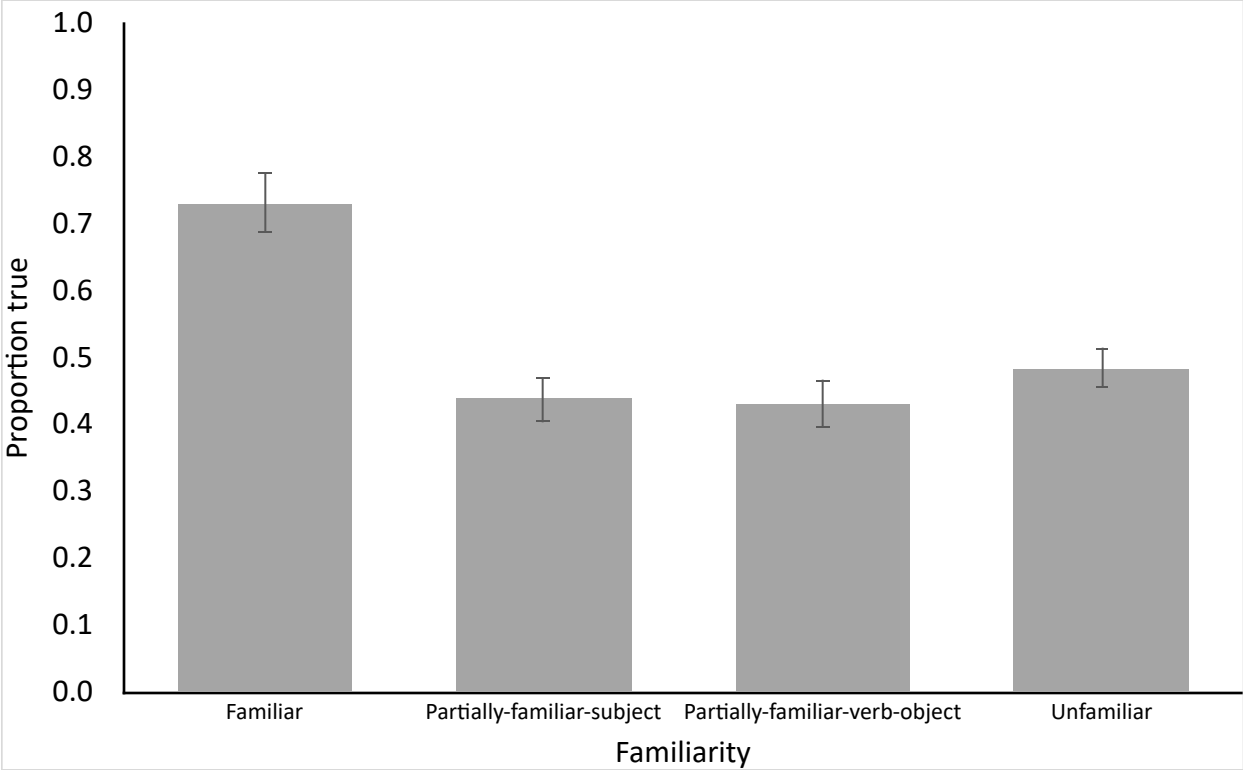
The primary question in this experiment was: how do people assess the truth of partially-familiar claims, relative to both familiar and unfamiliar claims? To address this research question, we tallied the proportion of familiar, partially-familiar and unfamiliar claims that each subject classified as true. These proportions, displayed in Figure 1, show two important findings.

First, subjects classified more familiar claims as true than they did unfamiliar claims. This finding is consistent with a large body of research on the truth effect—people rate claims as more true when those claims are repeated, possibly because the familiarity induced by repetition makes those claims feel easier to process. Second, there was no evidence that a "kernel of truth" mattered: that is, subjects classified roughly the same amount of partially-familiar claims as true as they did unfamiliar claims. In fact, the figure suggests that subjects responded to partially-familiar claims as though they were unfamiliar claims, regardless of whether the familiar part was the subject or verb and object of the sentence.

In other words, a within-subjects analysis of variance (ANOVA) showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 873) = 114.56$, $p < .001$, $\omega^2 = 0.21$. More specifically, we also conducted Bonferroni pairwise *t* tests on the differences between people's assessments of truth for each type of claim. Subjects classified significantly more familiar claims as true than unfamiliar claims, $M_{diff} = 0.25$, $t = 13.16$, $p < 0.001$, 95% CI [0.20, 0.29]. They also classified significantly more familiar claims as true than both partially-familiar-subject claims, $M_{diff} = 0.29$, $t = 15.63$, $p < 0.001$, 95% CI [0.24, 0.34], and partially-familiar-verb

Figure 1

Proportion of Claims Classified as True in Experiment 1



Note. Error bars are 95% confidence intervals around condition means. We adjusted the confidence intervals to make within-subjects comparisons using the Cousineau-Morey decorrelation method (Cousineau *et al*, 2021).

claims, $M_{diff} = 0.30$, $t = 15.98$, $p < 0.001$, 95% CI = [0.25, 0.35]. People also classified slightly fewer partially-familiar claims as true than unfamiliar claims, but this difference only reached significance for the partially-familiar-verb-object claims, $M_{diff} = 0.05$, $t = 2.28$, $p = .03$, 95% CI = [0.00, 0.09]. See Table S4 (Appendix 4) for all pairwise *t* tests.

As an unplanned analysis, we re-analysed the data considering each trivia claim as an item of analysis (see Appendix 4). Because we counterbalanced the materials so that each claim was familiar, partially-familiar or unfamiliar claim for subjects in different counterbalance

groups, we could compare subjects' assessments of truth for the same claim at these different levels of familiarity. This item analysis shows the same pattern of results as the main analysis. First, more people classified a claim as true when that claim was familiar than when it was unfamiliar. Second, roughly the same number of people classified a claim as true when that claim was partially-familiar as when it was unfamiliar. A within-subjects ANOVA again showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 477) = 171.82$, $p<.001$, $\omega^2 = 0.27$.

If we consider the pattern of results across all analyses, we found no evidence that partially-familiar claims act like a "kernel of truth." Instead, we found a significant truth effect for the familiar claims, but no evidence of a similar effect on people's assessments of truth for the partially-familiar claims. These results suggest that subjects were influenced by fluency when they classified the familiar claims as true or false, but were not when they classified the partially-familiar claims. This asymmetry is not easily explained by the literatures on processing fluency and the truth effect, which suggests that *anything* that makes a claim feel relatively easier to process should inflate people's assessments of truth (Reber & Schwarz, 1999). But the asymmetry can be explained by literature suggesting that a feeling of truth arises from coherently-linked references to a claim in memory. In this account, people have no more coherently-linked references to the partially-familiar claims than to the unfamiliar claims. This referential network account also fits with the idea that seeing something familiar in an unfamiliar context alerts people that something is wrong, and they may think more carefully when they classify these claims as true or false.

A critic might say that people failed to inflate their judgements of truth for the partially-familiar claims not because these claims have fewer coherent references but because they could recognise the "intact" claim from the exposure phase just ten minutes earlier. To understand why, let us consider how subjects would have evaluated each type of claim. When subjects encounter familiar claims, recent exposure to those claims makes them feel familiar and come to mind more easily. Even if subjects know they saw those claims just 10 min ago, they would still inflate their assessments of truth because people assume that repeated information is more likely to be true than new information (Arkes et al, 1991). By contrast, when subjects encounter partially-familiar claims, the intact claims from the exposure phase are highly accessible, making it easy to detect a discrepancy. Subjects may interpret this discrepancy as a signal that the partially-familiar claims are false. In other words, with a short delay, it might be easy for people to reject the partially-familiar claims by bringing to mind intact claims they saw in the exposure phase. This explanation is reminiscent of the discrepancy detection principle in memory research, where people can correctly reject misleading information when it is easy to notice discrepancies between that information and their original memories (Butler & Loftus, 2018; Tousignant et al, 1986). But, people rapidly forget source-specifying details—information about where they first learned something—as time increases, making it more difficult to detect such discrepancies across longer time delays (Johnson *et al*, 1993). To address this counter-explanation, in Experiment 2, we increased the delay between exposure and test to make it more difficult for people to retrieve source-specifying details.

Experiment 2

In Experiment 2 we followed the same method as in Experiment 1, but subjects completed the exposure and test phases on two separate sessions, separated by approximately 65 hr. If, in Experiment 1, subjects were rejecting the partially-familiar claims by comparing them to what they saw in the exposure phase, then we would expect subjects in Experiment 2 to classify more partially-familiar claims as true than unfamiliar claims, yet fewer than familiar claims. But, if the results of Experiment 1 were due to a coherent references account of the truth effect, then we would expect the same pattern of results in Experiment 2, because increasing the delay between exposure and test does not give partially-familiar claims more coherent references.

Method

Subjects

We recruited 400 subjects from MTurk. Subjects participated in exchange for $0.75 Amazon credit. We preregistered Experiment 2 on AsPredicted (https://aspredicted.org/; see Appendix 6 for a copy of the preregistration). Due to the aforementioned Cloud Research x Qualtrics interaction, we collected data from 445 subjects. We excluded responses from 34 subjects for failing to complete the exposure phase. Of the remaining 411 subjects who completed the exposure phase, 230 (56%) responded to our emails to begin the test. We excluded responses from 4 subjects for failing to complete the test phase, from a further 2 subjects for failing to describe the task, and from a further 16 subjects for searching the claims online, leaving a final sample size of 208 (156 women, 49 men and 3 gender diverse subjects). Subjects

ranged in age from 21 – 84 years old (M = 46.02, Med = 44, SD = 14.12) and 198 subjects (95%) spoke English as a first language.

Design

Experiment 2 followed the same design as Experiment 1.

Procedure

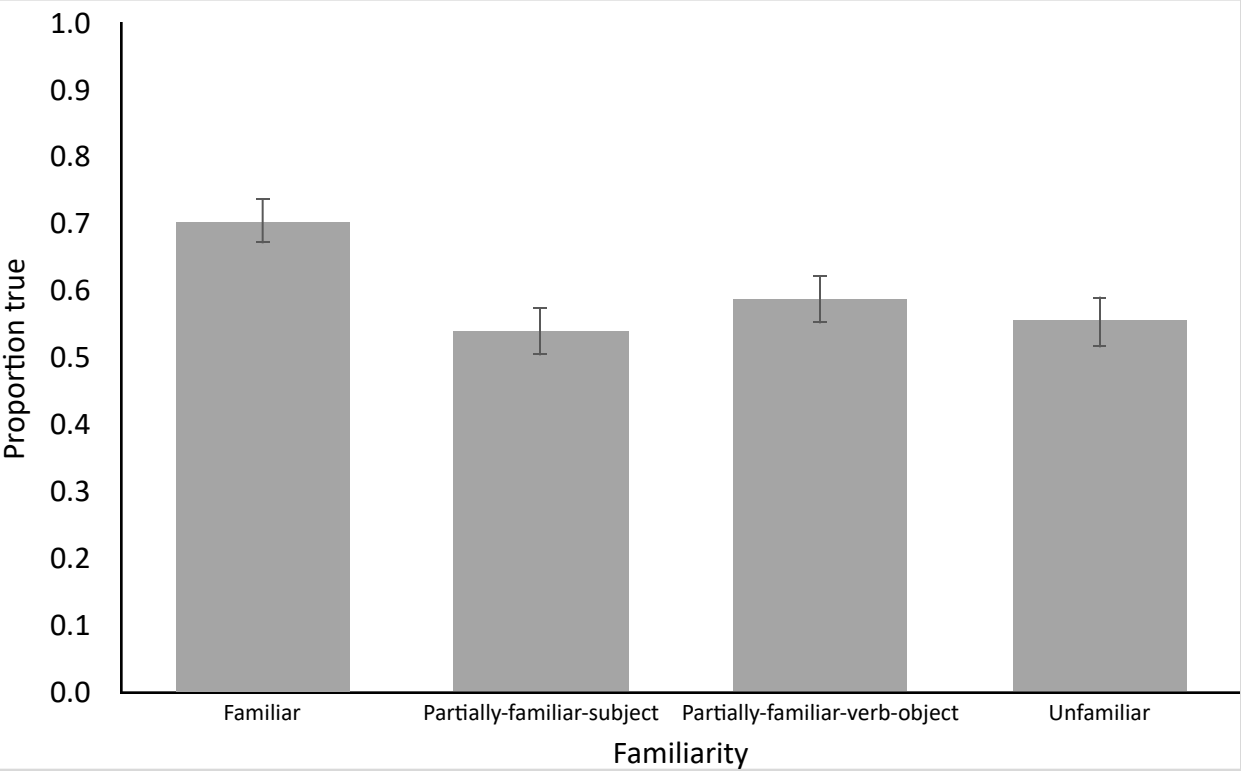
The procedure of Experiment 2 was identical to that of Experiment 1, except for the following variations. First, the exposure and tests phases were self-paced. That is, subjects were not forced to spend 4 s reading each claim before rating that claim for interest (exposure phase) or truth (test phase). Second, and most importantly, subjects completed the exposure and test phases on two consecutive sessions, separated by a 66 hr delay. We asked subjects not to search for any more information about the claims, or discuss the claims with other people, during the delay. We emailed subjects with a link to the test phase 2 days after they completed the exposure phase. After receiving this link, we asked subjects to complete the test phase within 24 hr. Thus, the delay between the exposure and test phases ranged from 44 hr 28 min 13 s – 143 hr 11 min 59 s (M = 67 hr 7 min 31 s, Med = 65 hr 4 min 46 s, SD = 15 hr 5 min 9 s). Subjects' completion times for the exposure phase ranged from 3 min 12 s - 36 min 35 s (M = 8 min 41 s, Med = 7 min 20 s, SD = 4 min 23 s) and their completion times for the test phase ranged from 2 min 53 s - 56 min 4 s (M = 8 min 54 s, Med = 7 min 16, SD = 6 min 3 s).

Results and Discussion

How did people assess the truth of partially-familiar claims, relative to both familiar and unfamiliar claims, across a longer time delay? Figure 2 shows the proportion of familiar, partially-familiar, and unfamiliar claims that each subject classified as true. As the figure shows,

Figure 2

Proportion of Claims Classified as True in Experiment 2



Note. Error bars are 95% confidence intervals around condition means. We adjusted the confidence intervals to make within-subjects comparisons using the Cousineau-Morey decorrelation method (Cousineau *et al*, 2021).

we replicated the findings from Experiment 1. That is, subjects classified more familiar claims as true than they did unfamiliar claims. In addition, subjects again classified roughly the same amount of partially-familiar claims as true as unfamiliar claims.

A within-subjects ANOVA showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 621) = 34.17$, $p < .001$, $\omega^2 = 0.09$. Bonferroni pairwise t-tests showed that subjects classified significantly more familiar claims as true than unfamiliar claims, $M_{diff} = 0.15$, $t = 8.31$, $p < .001$, 95% CI [0.10, 0.20]. They also classified significantly more familiar claims as true than both partially-familiar-subject claims, $M_{diff} = 0.17$, $t = 9.11$, $p < .001$, 95% CI

[0.12, 0.21], and partially-familiar-verb-object claims, $M_{diff} = 0.12$, $t = 6.51$, $p < .001$, 95% CI [0.07, 0.17]. The differences between partially-familiar and unfamiliar claims were not statistically significant.

As an unplanned analysis, we re-analysed the data considering each trivia claim as an item of analysis (see Appendix 4). Again, this item analysis shows the same pattern of results as the main analysis. First, more people classified a claim as true when that claim was familiar than when it was unfamiliar. Second, roughly the same number of people classified a claim as true when that claim was partially-familiar as when it was unfamiliar. A within-subjects ANOVA again showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 477) = 51.99$, $p<.001$, $\omega^2 = 0.08$.

In Experiment 2, we found no evidence that partially-familiar claims act like a "kernel of truth." Even across a delay of 66 hr, subjects responded to partially-familiar claims as though they were unfamiliar claims. Again, these results fit with a referential network account of the truth effect. People rapidly forget information about where they first learned something, so it is unlikely that after 66 hr subjects were rejecting the partially-familiar claims by recalling what they saw in the exposure phase. One possible counter-explanation comes from the fluency literature. People are more sensitive to *changes* in how easy it is to process information than any "absolute level" of fluency (Dechêne et al, 2009; Hansen *et* al, 2008). For example, familiar claims "feel true" because they are easier to process than the preceding information. In other words, the unfamiliar claims act like a "benchmark" for how easy the familiar claims are to process. If people do not have a clear benchmark for the familiar claims—such as in truth effect experiments in which subjects only encounter familiar claims—the size of truth effect tends to

decrease (Dechêne et al, 2009). But what is people's benchmark for the partially-familiar claims? By definition, these claims have two opposite benchmarks—familiar claims and unfamiliar claims. That is, when subjects encountered a partially-familiar claim in the preceding experiments, sometimes that claim was easier to process than the preceding information and sometimes it was more difficult to process. As a result, subjects may have failed to inflate their assessments of truth for the partially-familiar claims.

Taking the results of the preceding experiments together, subjects inflated their assessments of truth for familiar claims and they responded to partially-familiar claims as if those claims were unfamiliar. But, subjects may not have had a clear benchmark when they encountered partially-familiar claims. Therefore, in Experiment 3, we manipulated this benchmark by randomly assigning subjects to see partially-familiar claims alongside either familiar *or* unfamiliar claims. We expected that how people assess the truth of partially-familiar claims would depend on this benchmark. When subjects compare the partially-familiar claims to unfamiliar claims, we expected them to process these claims with relative ease. In this case, subjects should classify more partially-familiar claims as true than unfamiliar claims, yet fewer than familiar claims. When subjects compare the partially-familiar claims only to familiar claims, we expected them to process these claims with relative difficulty. In this case, subjects should classify roughly the same number of partially-familiar and unfamiliar claims as true (as in the preceding experiments). Support for these hypotheses would be consistent with the idea that fluency influences people's decision making more when there is a *consistent change* in how easy information is to process from one moment to the next. But, if Experiment 3 has the same pattern of results as the preceding experiments, that would fit better with a referential network account

of the truth effect—in which partially-familiar claims have no more coherently linked references to memory than do unfamiliar claims (Unkelbach & Rom, 2017).

Experiment 3

Method

Subjects

We recruited 400 subjects from MTurk. Subjects participated in exchange for $0.90 USD in Amazon credit ($0.30 after the first session and a bonus of $0.60 after the second session). We preregistered Experiment 3 on AsPredicted (https://aspredicted.org/; see Appendix 7 for a copy of the preregistration). Due to the aforementioned Cloud Research x Qualtrics interaction, we collected data from 409 subjects. We excluded responses from 4 subjects for failing to complete the exposure phase. Of the remaining 405 subjects who completed the exposure phase, 224 (55%) responded to our emails to begin the test phase. We excluded responses from 3 subjects for failing to complete the test phase, from a further 4 subjects for failing to describe the task, and from a further 9 subjects for searching the claims online, leaving a final sample size of 208 (46 women, 60 men and 2 gender diverse subjects). Subjects ranged in age from 22 – 84 years old (M = 45.89, Med = 45, SD = 14.12) and 199 subjects (96%) spoke English as a first language. The demographics of these subjects was similar to the subjects' demographics before we applied these inclusion criteria (see Appendix 3).

Design

Experiment 3 followed a 2 (Benchmark: familiar, unfamiliar) x 2 (Familiarity: partially-familiar, benchmark) mixed factorial design. We manipulated the benchmark between subjects and the claims within subjects. To simplify the design of Experiment 3, and because there were

no consistent differences between these claims in the preceding experiments, we collapsed across "partially-familiar-subject" and "partially-familiar-verb-object" claims. As in the preceding experiments, each trivia claim served equally often as a familiar, partially-familiar or unfamiliar claim for subjects in different counterbalance groups (see Appendix 2). We modified the counterbalancing scheme from the preceding experiments so as to assign subjects to one of two benchmarks at the same time as assigning them to a counterbalance group. There were 32 counterbalance groups. Counterbalance groups 1 - 16 formed the "familiar" benchmark, and groups 17 - 32 formed the unfamiliar benchmark.

Procedure

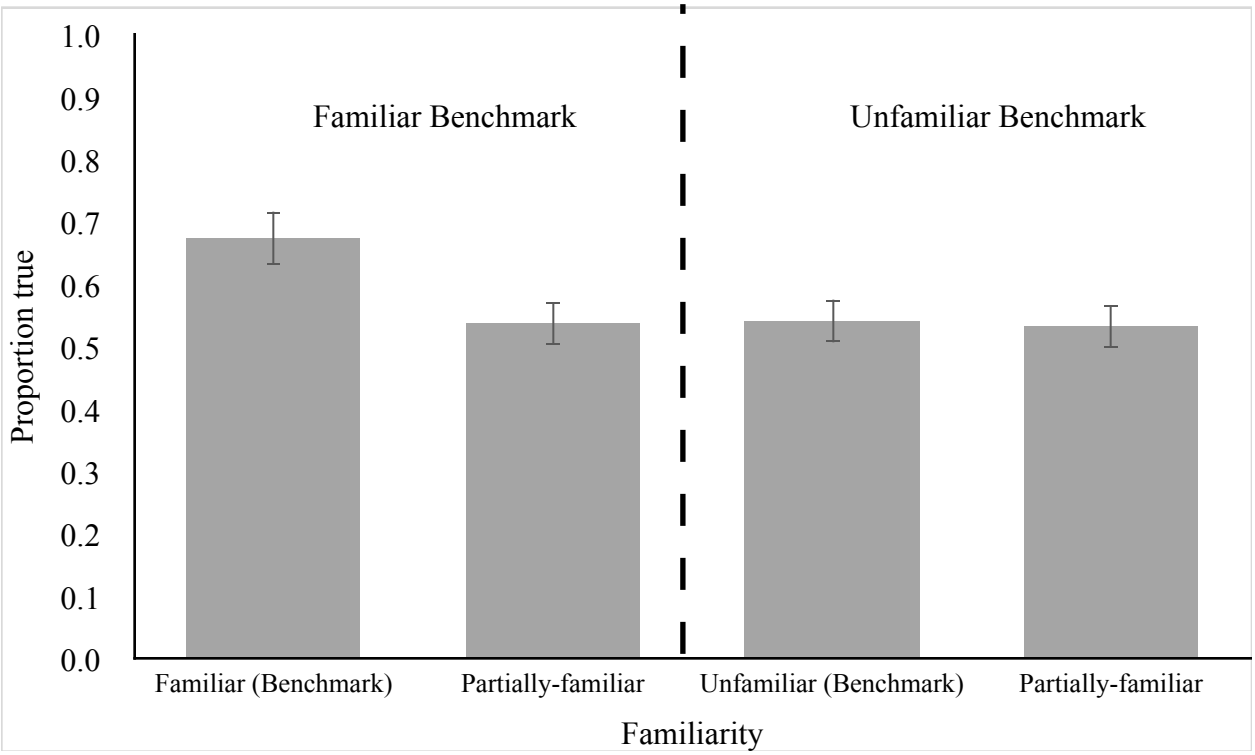
The procedure of Experiment 3 was identical to that of Experiment 2, except for the following variations. First, the delay between exposure and test phases differed due to variation in when participants completed the second session. The delay between the exposure and test phases ranged from 47 hr 51 min - 430 hr 6 min (M = 69 hr 7 min, Med = 60 hr 42 min, SD = 35 hr 38 min). Second, to ensure that subjects encountered the same amount of claims in the test phase as in the preceding experiments, and that the same proportion of those claims were partially-familiar, we doubled the number of "benchmark" claims that subjects encountered. Thus, in the test phase, subjects encountered 20 partially-familiar claims and either 20 familiar or 20 unfamiliar claims. As in the preceding experiments, the claims were presented one at a time in a random order. Subjects' completion times for the exposure phase ranged from 1 min 50 s - 52 min 47 s (M = 8 min 17 s, Med = 7 min, SD = 5 min 37 s) and their completion times for the test phase ranged from 3 min 50 s - 55 min 39 s (M = 8 min 5 s, Med = 7 min 1 s, SD = 4 min 41 s).

Results and Discussion

How do people assess the truth of partially-familiar claims, relative to either familiar or unfamiliar claims? Figure 3 shows the proportion of partially-familiar and benchmark claims that each subject classified as true.

Figure 3

Proportion of Claims Classified as True in Experiment 3



Note. Error bars are 95% confidence intervals around condition means. We adjusted the confidence intervals to make within-subjects comparisons using the Cousineau-Morey decorrelation method (Cousineau *et al*, 2021).

The figure shows three important findings. First, subjects' assessments of truth for the comparison claims depended on whether they were assigned to the "Familiar" or "Unfamiliar" benchmark. That is, "Familiar" subjects classified more benchmark claims as true than "Unfamiliar" subjects. This is the typical truth effect. Second, regardless of their benchmark,

subjects classified roughly the same number of partially-familiar claims as true. That is, we found no evidence that subjects' assessments of truth for the partially-familiar claims depends on their benchmark. Third, subjects classified roughly the same number of partially-familiar claims as true as unfamiliar claims—regardless of whether they actually encountered unfamiliar claims.

In other words, a 2x2 mixed factorial ANOVA showed a significant interaction between familiarity and benchmark condition, $F(1, 206) = 14.68$, $p < .001$, $\omega^2 = .03$. Followup Bonferroni pairwise t-tests show that subjects classified significantly more familiar claims as true than unfamiliar claims, $M_{diff} = 0.13$, $t = 5.22$, $p < .001$, 95% CI [0.07, 0.20]. "Familiar" subjects and "Unfamiliar" subjects also classified roughly the same number of partially-familiar claims as true, $M_{diff} = 0.004$, $t = 0.15$, $p = 1.00$, 95% CI [-0.06, 0.07]. As in the preceding experiments, subjects classified roughly the same number of partially-familiar and unfamiliar claims as true, in both the "Familiar," $M_{diff} = 0.01$, $t = 0.22$, $p = 1.00$, 95% CI [-0.06, 0.07], and "Unfamiliar" benchmarks, $M_{diff} = 0.01$, $t = 0.39$, $p = 1.00$, 95% CI [-0.05, 0.07].

We re-analysed the data considering each trivia claim as an item of analysis and found the same pattern of results (see Appendix 4). A within-subjects ANOVA again showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 477) = 32.559$, $p<.001$, $\omega^2 = 0.06$. Subjects classified significantly more familiar claims as true than both partially-familiar and unfamiliar claims.

This pattern of results replicates the findings from the preceding experiments. Again we found the typical truth effect—people classified more familiar claims as true than unfamiliar claims—even though we manipulated familiar and unfamiliar claims between subjects. We also

found no evidence that partially-familiar claims act like a "kernel of truth"—in both benchmark conditions subjects responded to partially-familiar claims as though they were unfamiliar claims.

Given these results, it is unlikely that subjects failed to inflate their assessments of truth for the partially-familiar claims due to an unclear benchmark. This pattern of results seems inconsistent with the idea that fluency is a relative judgement. If, as the literature suggests, familiar claims are only fluent when compared with unfamiliar claims, then "Familiar" subjects should have classified fewer partially-familiar claims as true than "Unfamiliar" subjects. But, when considering the referential network theory of the truth effect, it is not surprising that people classified the same number of partially-familiar and unfamiliar claims. In this theory, fluency comes from the number of coherently linked references from the concepts in the claim to facts in memory (Unkelbach & Rom, 2017). Familiar claims should have more coherently-linked references than both partially-familiar and unfamiliar claims, and partial-familiar claims should have roughly the same number of coherently linked references as unfamiliar claims.

General Discussion

We investigated how people assess the truth of partially-familiar claims claims, relative to both familiar and unfamiliar claims. In Experiment 1 we found no difference between subjects' assessments of truth for partially-familiar and unfamiliar claims. In Experiment 2 we addressed the counter-explanation that subjects failed to inflate their assessments of truth for the partially-familiar claims because they detected a discrepancy with the "intact" versions of those claims they saw earlier. We reasoned that increasing the time delay between the exposure and test phases would make it more difficult for subjects to detect such a discrepancy—if indeed that was what they were doing. In Experiment 2 we replicated the first experiment across a longer delay

(66 hr instead of 10 min), finding a similar pattern of results as before. In Experiment 3 we addressed the counter-explanation that subjects failed to inflate their assessments of truth for the partially-familiar claims because they lacked a clear benchmark for how easy those claims were to process. We reasoned that showing subjects partially-familiar claims alongside either familiar *or* unfamiliar claims (but not both) would provide such a benchmark. We found a similar pattern of results as in the preceding experiments, and observed that changing the benchmark made no impact on people's assessments of truth for the partially-familiar claims. Overall, there were two key findings of the experiments presented here. First, subjects classified more familiar claims as true than unfamiliar claims—the classic truth effect. Second, subjects classified roughly the same number of partially-familiar claims as true as they did unfamiliar claims.

These findings tentatively suggest that a "kernel of truth" does not lead people to inflate their assessments of truth for partially-familiar claims. The overall hypothesis, that people would process partially-familiar claims more fluently than unfamiliar claims, was not supported. The finding that subjects rated the truth of partially-familiar claims the same as unfamiliar claims fits better with a referential network account of the truth effect than with the standard fluency-based explanation. That is, we expected that the partially-familiar claims would feel more familiar and be easier to process than the unfamiliar claims. But, there is no reason to expect that these claims would have any more coherently-linked references to memory than the unfamiliar claims.

This pattern of results is reminiscent of a small body of research investigating illusions of truth for "contradicting" claims, compared with familiar and unfamiliar claims (Silva et al, 2017; Unkelbach & Rom, 2017). In these experiments subjects encounter a claim in the exposure phase (e.g., "Most accidents at work occur on Mondays"), then in the test phase they encounter a

version of this claim with one word changed so that it contradicts the meaning of the first claim (e.g, “Most accidents at work occur on Fridays”). How do people assess the truth of such “contradictory” claims? One study found that, with no delay between exposure and test, subjects rated the contradictory claims as less true than the unfamiliar claims (Unkelbach & Rom, 2017, Experiment 1). Another study randomly assigned subjects to complete the test phase either immediately or one week after the exposure phase (Silva et al, 2017). Subjects in the immediate test condition rated the contradictory claims as less true than the unfamiliar claims, whereas subjects in the delayed test condition rated the contradictory claims as more true than unfamiliar claims. A potential explanation for these results is that subjects in the immediate test condition rated the contradictory claims as false by detecting a discrepancy with what they saw earlier, whereas the source-specifying information required to detect such a discrepancy becomes less accessible over a longer time delay (Silva et al, 2017). By contrast, when we increased the delay between exposure and test, people classified roughly 10% more partially-familiar *and* unfamiliar claims as true, but the overall pattern of results did not change. The main difference between “contradictory” and partially-familiar claims is that the meaning of the partially-familiar claims does not necessarily contradict the “intact” version of those claims that subjects read in the exposure phase. For example, if subjects read the claim, “the Nile River flows northwards” in the exposure phase, and assessed the partially-familiar claim “the Nile River is the longest river in Asia” in the test phase, this partially-familiar claim does not contradict what they read earlier, even though it is false. People may be more likely to classify claims as false by detecting a discrepancy when the discrepant claim directly contradicts previous information, rather than merely differing from it. To further address this possibility, future researchers could test people's

assessments of truth for "contradictory" claims compared with partially-familiar claims across time delays of varying duration.

The data presented here are also reminiscent of those in a broader literature addressing how people can fail to update their beliefs to incorporate new knowledge. For example, several studies suggest that people draw on discredited information when making decisions (Greitemeyer, 2014; Johnson & Seifert, 1994; Walter & Tukachinsky, 2020). Subjects in these studies read some misinformation as part of a larger story (e.g., a story about a house fire that was caused by flammable material stored carelessly in a closet), and later on, they read a correction to the misinformation (e.g., the closet was actually empty). Although the misinformation only had a weak effect on subjects' beliefs after it had been corrected, it continued to influence their downstream decisions (e.g., saying that an insurance claim on the house could be refused due to negligence; Johnson & Seifert, 1994; Walter & Tukachinsky, 2020). Another study found that people who read misinformation followed by a correction believed the misinformation less than people who read the misinformation without a correction, but more than people who never read the misinformation in the first place—suggesting that people *try* to update their beliefs after reading a correction but do so insufficiently (Greitemeyer, 2014). In these studies, the corrections are analogous to partially-familiar claims, and the misinformation is analogous to the "intact" versions of those claims subjects read in the exposure phase. But, in the present experiments, the materials were counterbalanced so that the familiar claims subjects read in the test phases did not include any "intact" versions of partially-familiar claims. For example, subjects who read the claim "The Nile River flows northwards" in the exposure phase, would subsequently read *either* "The Nile River flows northwards" *or* "The Nile

River is the longest river in Asia" in the test phase (but not both). We counterbalanced the materials like this to make it more difficult for subjects to detect the manipulation of the experiments; but as a result we could not test whether subjects found the partially-familiar claims to be more believable than the "intact" versions of those claims. These studies on correcting misinformation suggest that—similar to the literature on discrepancy detection—people may classify partially-familiar claims as false as a consequence of failing to incorporate those claims with the "intact" versions they saw earlier.

The finding that subjects classified more familiar than unfamiliar claims as true fits with the truth effect literature. In fact, the mean differences between subjects' assessments of truth for familiar and unfamiliar claims (0.25 in Experiment 1, 0.15 in Experiment 2 and 0.13 in Experiment 3) are larger than the effect sizes typically found in the truth effect literature (see Dechêne et al, 2010; Fazio et al, 2019). A potential reason for these larger effect sizes is that we avoided warning subjects that they would encounter false information in the exposure phase. In a typical truth effect study, subject are told when they begin the experiment that half of the claims they are about to see are true and half are false, and they are asked to assess the truth of those claims in the exposure phase. By contrast, we did not tell subjects that any of the claims would be false, and asked them to assess how interesting those claims were in the exposure phase. The typical procedure may underestimate the size of the truth effect by leading subjects to think about accuracy when they first process the claims (Brashier et al; Jalbert et al, 2020).

One limitation of these experiments is that subjects made truth judgements only in the test phase, not in the exposure phase. As a consequence, we could not determine the extent to which repetition caused subjects to change their assessments of truth for any individual claims—

that is, whether subjects classified a claim as true in the test phase that they had previously classified as false in the exposure phase. Therefore, we cannot determine the extent to which familiarity leads people to classify the same claim as more true over time; only that on average, people classify more familiar claims as true than unfamiliar claims. Another limitation of the work in this thesis is that in all three experiments, half of the claims subjects encountered in the test phase were partially-familiar. But we know that reducing the proportion of fluent items on a recognition test enhances the effect of fluency on people's judgement—that is, the fewer fluent items there are, the more those items "pop out" from background information (Westerman, 2008). Therefore, to the extent that recognition and people's assessments of truth share mechanisms, we might expect people to classify more partially-familiar claims as true when there are relatively few of those claims. In the present research, we did not reduce the percentage of partially-familiar claims because we wanted to sample from a broad range of stimuli. But, future research should systematically vary the percentage of partially-familiar claims in the test phase—for example, from 10% to 90%.

Despite these limitations, the work presented here is a first step to addressing how people assess the truth of partially-familiar claims. This issue is important because in everyday life claims are often changed, updated or taken out of context when they are repeated—yet most of the truth effect literature tests only verbatim repetition of claims. For example, a recent literature review found that 82% of truth effect experiments only compared people's assessments of truth for exactly repeated claims with entirely new claims (Henderson et al, 2022). Our data are inconsistent with the idea that partially-familiar claims act like a "kernel of truth." Instead, these data fit better with the idea that people often fail to update their beliefs to incorporate new

knowledge, even when the updated information is more accurate (O'Brien et al, 2010; Walter & Tukachinsky, 2020). Taken together, these results show that when people are assessing truth, they respond to partially-familiar claims as if those claims were unfamiliar.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*(6), 576–605. https://doi.org/10.1016/0022-1031(91)90026-3

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, *51*(5), 267.

Ausubel, D. P. (1978). In Defense of Advance Organizers: A Reply to the Critics. In *Review of Educational Research* (Vol. 48, Issue 2, pp. 251–257).

Ausubel, D. P., & Fitzgerald, D. (1961). The role of discriminability in meaningful learning and retention. *Journal of Educational Psychology*, *52*(5), 266.

Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement, 17(3), 199. https://doi.org/10.1037/h0080140

Bransford, J. & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of verbal learning and verbal behavior. 11*(6), 717-726.

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*, 499–515. https://doi.org/10.1146/annurev-psych-010419-050807

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, *194*, 104054.

Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). *Types, sources, and claims of COVID-19 misinformation*. University of Oxford.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*(3), 311.

Brown, A. S. (2003). A review of the deja vu experience. *Psychological Bulletin*, *129*(3), 394.

Brown, A. S. (2004). The déjà vu illusion. *Current Directions in Psychological Science*, *13*(6), 256–259.

Brown, A. S. (2020). The butcher on the bus experience. *Memory Quirks*, 224–247.

Butler, B. J., & Loftus, E. F. (2018). Discrepancy detection in the retrieval-enhanced suggestibility paradigm. *Memory*, *26*(4), 483–492. https://doi.org/10.1080/09658211.2017.1371193

Cleary, A. M., Brown, A. S., Sawyer, B. D., Nomi, J. S., Ajoku, A. C., & Ryals, A. J. (2012). Familiarity from the configuration of objects in 3-dimensional space and its relation to déjà vu: A virtual reality investigation. *Consciousness and Cognition*, *21*(2), 969–975.

Corkill, A. J. (1992). Advance organizers: Facilitators of recall. *Educational Psychology Review*, *4*(1), 33–67.

Cousineau, D., Goulet, M.-A., & Harding, B. (2021). Summary plots with adjusted error bars: The superb framework with an implementation in R. *Advances in Methods and Practices in Psychological Science*, *4*(3), 25152459211035108.

Davis, C. (2020, November 20). Spot The Difference: How Many Can You Find? *The Delite*. https://www.thedelite.com/spot-the-difference-pictures-how-many-can-you-find

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology*, *45*(5), 1117–1122.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, *30*(3), 287–307.

Devitt, A. L., Monk-Fromont, E., Schacter, D. L., & Addis, D. R. (2016). Factors that influence the generation of autobiographical memory conjunction errors. *Memory*, *24*(2), 204–222. https://doi.org/10.1080/09658211.2014.998680

Fazio, L. K., 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

Frei, M. R., Jauchem, J. R., Dusch, S. J., Merritt, J. H., Berger, R. E., & Stedham, M. A. (1998). Chronic, low-level (1.0 W/kg) exposure of mice prone to mammary cancer to 2450 MHz microwaves. *Radiation Research*, *150*(5), 568–576.

Gilbert, D. T. (1991). How mental systems believe. *American psychologist*, *46*(2), 107.

Gilbert, E. S. (2009). Ionising radiation and cancer risks: What have we learned from epidemiology? *International Journal of Radiation Biology*, *85*(6), 467–482.

Greitemeyer, T. (2014). Article retracted, but the message lives on. *Psychonomic Bulletin & Review*, *21*(2), 557–561. https://doi.org/10.3758/s13423-013-0500-6

Gruppuso, V., Lindsay, D. S., & Masson, M. E. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review*, *14*(6), 1085–1089.

Hansen, J., Dechêne, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology*, *44*(3), 687–691. https://doi.org/10.1016/j.jesp.2007.04.005

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

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive Research: Principles and Implications*, *6*(1), 1-12.

Henderson, E. L., Westwood, S. J., & Simons, D. J. (2022). A reproducible systematic map of research on the truth effect. *Psychonomic Bulletin & Review*, *29*(3), 1065–1088.

Higgins, E. T. (1998). The aboutness principle: A pervasive influence on human inference. *Social Cognition*, *16*(1), 173–198. https://doi.org/10.1521/soco.1998.16.1.173

Jalbert, M., Newman, E., & Schwarz, N. (2019). *Trivia Claim Norming: Methods Report and Data*. https://doi.org/10.6084/m9.figshare.9975602

Jalbert, M., Newman, E., & Schwarz, N. (2020). Only half of what I'll tell you is true: Expecting to encounter falsehoods reduces illusory truth. *Journal of Applied Research in Memory and Cognition*, *9*(4), 602–613.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3. https://doi.org/10.1037/0033-2909.114.1.3

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language*, *45*(1), 82–102. https://doi.org/10.1016/j.jml.2006.01.002

Jones, T. C., Bartlett, J. C., & Wade, K. A. (2006). Nonverbal conjunction errors in recognition memory: Support for familiarity but not for feature bundling. *Journal of Memory and Language*, *55*(1), 138–155. https://doi.org/10.1016/j.jml.2006.01.002

Kelley, D. E., Noar, S. M., & Seidenberg, A. B. (2018). Understanding misinformation in the pro-tanning communication environment: A content analysis. *American Journal of Health Education*, *49*(4), 234–245. https://doi.org/10.1080/19325037.2018.1473181

Lakens, D., & Caldwell, A. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science, 4*(1).

Leding, J. K., & Lampinen, J. M. (2009). Memory conjunction errors: The effects of presentation duration and study repetition. *Memory*, *17*(5), 597–607.

Lindsay, S. (2008). Source Monitoring. In H. L. Roediger (Ed.), *Cognitive Psychology of Memory* (Vol. 2, pp. 325–348). Elsevier.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. Com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442.

Loftus, E. F. (1976). Unconscious transference in eyewitness identification. *Law & Psychol. Rev.*, *2*, 93.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585–589.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252.

Marsh, R. L., Hicks, J. L., & Davis, T. T. (2002). Source monitoring does not alleviate (and may exacerbate) the occurrence of memory conjunction errors. *Journal of Memory and Language*, *47*(2), 315–326.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*(5), 424–428.

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition*, *30*(8), 1219–1227.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, *19*(3), 338–368.

Newman, E., Jalbert, M. & Feigenson, N. (2019). Cognitive Fluency in the Courtroom. *The Routledge International Handbook of Legal and Investigative Psychology*. 102-115.

O'Brien, E. J., Cook, A. E., & Guéraud, S. (2010). Accessibility of outdated information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 979.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338–342.

Reinitz, M. T., Lammers, W. J., & Cochran. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, *20*(1), 1–11.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Salovich, N. A., Kirsch, A. M., & Rapp, D. N. (2022). Evaluative mindsets can protect against the influence of false information. *Cognition*, *225*, 105121.

Schwartz, B. L., Cleary, A. M., Dunlosky, J., & Tauber, S. (2016). Tip-of-the-tongue states, déjà vu experiences, and other odd metamemory experiences. *The Oxford Handbook of Metamemory*, 95.

Schwarz, N. (2011). Feelings-as-information theory. *Handbook of Theories of Social Psychology*, *1*, 289–308.

Silva, R. R., Garcia-Marques, T., & Reber, R. (2017). The informative value of type of repetition: Perceptual and conceptual fluency influences on judgments of truth. *Consciousness and Cognition*, *51*, 53–67. https://doi.org/10.1016/j.concog.2017.02.016

South, C. (2011, October 7). 10 Where's Waldo Puzzles Online. *Art-Sci*. https://art-sci.blogspot.com/2011/07/10-wheres-waldo-puzzles.html

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*(4), 1115–1143.

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, *14*(4), 329–338.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*(4), 339–345.

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126.

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247–253.

Walter, N., & Tukachinsky, R. (2020). 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.

Westerman, D. L. (2008). Relative fluency and illusions of recognition memory. *Psychonomic Bulletin & Review*, *15*(6), 1196–1200.

Appendix 1: Stimulus Development

As a starting point for stimulus development, we used a database of trivia questions that have been used in previous psychological experiments (Jalbert et al., 2019; Nelson & Narens, 1980; Tauber et al., 2013). We rephrased these questions into true trivia statements in the grammatical form: subject, verb, object (e.g., "The Nile River flows northwards"; "The Yangtze River is the longest river in Asia"). We created more true claims by searching Encyclopaedia Britannica online. We created a list of false claims by combining the subject of one true claim with the verb and object of a different true claim. For example, we generated the false claim, "The Yangtze River flows northwards" by combining the subject of a true claim "The Yangtze River is the longest river in Asia" with the verb and object of a different true claim "The Nile River flows northwards". For half of the partially-familiar claims we repeated the subject of the sentence from the exposure phase, and for the other half we repeated the verb and object, to make sure any effect of partial familiarity on people's assessments of truth did not depend on which part of the claim was repeated.

Stimulus Development Study

The purpose of this study was to establish a "baseline" of plausibility for the claims. We did this by asking people to read a subset of claims in one session and classify them as true or false. We used these baseline assessments of truth to select which claims to use in the main experiments. In the main experiments we wanted claims that people would not overwhelming classify as true or false when seeing them for the first time. We wanted to avoid claims for which large numbers of people would draw on their prior knowledge to decide that it was true or false.

There was no manipulation in this study—subjects simply classified a subset of 70 out of 280 claims as true or false.

Method

Subjects. We recruited subjects from the United States and Canada on MTurk. Subjects participated in exchange for $0.30 Amazon credit. This study was not pre-registered. We set a target sample size of 300 subjects, to make reasonable inferences about how often subjects classify each of the claims as true. Due to the way MTurk interacts with Qualtrics, we collected responses from 407 subjects. We then excluded 106 responses for failing to complete the study, 52 responses for searching the claims online, and a further 48 responses failing our attention check. This process yielded a final sample of 201 subjects (104 women). Subjects ranged in age from 20 – 76 years old (M = 41.6, SD = 12.5).

Design. There was no manipulation in this study. Subjects simply classified a subset of 70 out of 280 trivia claims.

Procedure. We told subjects the study was about visual and verbal learning, and they would see some trivia claims and decide whether they were true or false. We asked subjects to give their "general impressions" of the claims and not to search the claims online. Everyone saw the trivia claims one at a time, in random order. To make sure subjects had enough time to read each claim and form an assessment of truth, each claim appeared in the centre of the screen for 4 s before a true/false scale appears below the claim.

Results and Discussion

Recall the primary purpose of this study was to determine how many people would classify the claims as true when seeing them for the first time, so that we could select which

claims to use in Experiments 1 - 3. To address this question, we first tallied the proportion of subjects who judged each trivia claim true, then further subdivided those data by whether the claims were true or false. People classified the claims as true more often when those claims were actually true compared with when they were false, and this difference was statistically significant, $t(278) = 6.84$, $p < 0.001$. We also aimed to avoid claims in which large numbers of subjects might draw on their prior knowledge to decide a claim was true or false. To address this aim, we excluded claims where the proportion of subjects who classified it as true was less than 0.15 and greater than 0.85. This process yielded 176 claims. To decide which of these claims to use in Experiment 1 - 3 we excluded the next 16 claims with the highest mean truth scores. Thus, we used 160 claims in Experiments 1- 3

Appendix 2: Counterbalancing

Counterbalancing Scheme for Experiments 1 and 2

We counterbalanced the subset of claims that served as familiar, partially-familiar, and unfamiliar in Experiments 1 and 2. In this counterbalancing scheme, subjects were randomly assigned to one of 16 groups, where for each group a different subset of claims served as familiar, partially-familiar. When we calculated proportion truth judgements in the results of Experiments 1 and 2, we summed across all 16 counterbalances. Thus, we aimed to separate any effect of familiarity on people's judgements of truth from the effects of individual items because each individual claim served equally often as familiar, partially-familiar and unfamiliar for subjects in different counterbalances. In counterbalances 1 – 8, subjects saw 40 true claims in the exposure phase; in conditions 9 – 16, subjects saw 40 false claims at exposure. All subjects saw 20 true claims and 20 false claims at exposure. Because we combined two true claims to create each pair of false claims, the partially-familiar claims were either all true or all false depending on which counterbalance subjects were assigned to. In counterbalances 1 - 8 all the partially-familiar claims were true, and in counterbalances 9 – 16 all the partially-familiar claims were false. Table S1 shows the protocol we used to set up this counterbalancing scheme.

Table S1: Protocol for counterbalancing scheme in Experiment 1 and 2

Example claim	Claim type	First letter	Subject	Second letter	Verb and object
The Nile River flows northwards	AB	A	The Nile River	B	flows northwards
The Yangtze River is the longest river in Asia	CD	C	The Yangtze River	D	is the longest river in Asia
The Nile River is the longest River in Asia	AD	A	The Nile River	D	is the longest river in Asia
The Yangtze River flows northwards	CB	C	The Yangtze River	B	flows northwards

There are 40 AB claims, 40 CD claims, 40 AD claims and 40 CB claims, and we split the claims into 4 groups of 10. We combined the subject, verb and object of the true AB and CD to create an equal number of false AD and CB claims. Table S2 shows the full counterbalancing scheme.

Table S2: Counterbalancing scheme for Experiments 1 and 2

Counterbalance	Exposure				Test			
					Familiar	Partially-familiar-subject	Partially-familiar-verb-object	Unfamiliar
1	AB1	AB2	AB3	AB4	AB1	AD3	CB4	CD2
2	AB1	AB2	AB3	AB4	AB4	AD2	CB3	CD1
3	AB1	AB2	AB3	AB4	AB3	AD1	CB2	CD4
4	AB1	AB2	AB3	AB4	AB2	AD4	CB1	CD3
5	CD1	CD2	CD3	CD4	CD1	CB3	AD4	AB2

6	CD1	CD2	CD3	CD4	CD4	CB2	AD3	AB1
7	CD1	CD2	CD3	CD4	CD3	CB1	AD2	AB4
8	CD1	CD2	CD3	CD4	CD2	CB4	AB1	AB3
9	AD1	AD2	AD3	AD4	AD1	AB3	CD4	CB2
10	AD1	AD2	AD3	AD4	AD4	AB2	CD3	CB1
11	AD1	AD2	AD3	AD4	AD3	AB1	CD2	CB4
12	AD1	AD2	AD3	AD4	AD2	AB4	CD1	CB3
13	CB1	CB2	CB3	CB4	CB1	CD3	AB4	AD2
14	CB1	CB2	CB3	CB4	CB4	CD2	AB3	AD2
15	CB1	CB2	CB3	CB4	CB3	CD1	AB2	AD4
16	CB1	CB2	CB3	CB4	CB2	CD4	AB1	AD3

Counterbalancing Scheme for Experiment 3

In Experiment 3 we modified the counterbalancing scheme from the preceding experiments so as to assign subjects to one of two comparison groups at the same time. There were 32 counterbalance groups. Counterbalance groups 1 - 16 formed the “familiar” comparison condition, and groups 17 - 32 formed the unfamiliar condition. We also collapsed across the partially-familiar-subject and partially-familiar-verb-object claims. Table S3 shows the full counterbalancing scheme.

Table S3: Counterbalancing scheme for Experiment 3

				Test			
Counterbalance	Benchmark condition	Exposure		Familiar (benchmark)	Partially-familiar (familiar benchmark)	Partially-familiar (unfamiliar benchmark)	Unfamiliar (benchmark)
1	Familiar	AB1	AB2	AB1	AD2		
2	Familiar	AB1	AB2	AB1	CB2		
3	Familiar	AB1	AB2	AB2	AD1		
4	Familiar	AB1	AB2	AB2	CB1		

5	Familiar	CD1	CD2	CD1	CB2		
6	Familiar	CD1	CD2	CD1	AD2		
7	Familiar	CD1	CD2	CD2	CB1		
8	Familiar	CD1	CD2	CD2	AD1		
9	Familiar	AD1	AD2	AD1	AB2		
10	Familiar	AD1	AD2	AD1	AB2		
11	Familiar	AD1	AD2	AD2	AB1		
12	Familiar	AD1	AD2	AD2	CD1		
13	Familiar	CB1	CB2	CB1	AB2		
14	Familiar	CB1	CB2	CB1	CD2		
15	Familiar	CB1	CB2	CB2	AB1		
16	Familiar	CB1	CB2	CB2	CD1		
17	Unfamiliar	AB1	AB2			AD2	CD1
18	Unfamiliar	AB1	AB2			CB2	CD1
19	Unfamiliar	AB1	AB2			AD1	CD2
20	Unfamiliar	AB1	AB2			CB1	CD2
21	Unfamiliar	CD1	CD2			CB2	AB1
22	Unfamiliar	CD1	CD2			AD2	AB1
23	Unfamiliar	CD1	CD2			CB1	AB2
24	Unfamiliar	CD1	CD2			AD1	AB2
25	Unfamiliar	AD1	AD2			AB2	CB1
26	Unfamiliar	AD1	AD2			CD2	CB1
27	Unfamiliar	AD1	AD2			AB1	CB2
28	Unfamiliar	AD1	AD2			CD1	CB2
29	Unfamiliar	CB1	CB2			AB2	AD1
30	Unfamiliar	CB1	CB2			CD2	AD1
31	Unfamiliar	CB1	CB2			AB1	AD2
32	Unfamiliar	CB1	CB2			CD1	AD2

Appendix 3: Supplemental Demographic Information

Due to a high rate of exclusions, we compared subjects' demographic information before and after applying the exclusion criteria. We found no meaningful differences between the two; therefore, it is unlikely that the attrition rat influenced the pattern of results.

Demographic information for all subjects in Experiment 1

In Experiment 1 there were 578 subjects (252 women, 155 men, 2 gender diverse subjects). Subjects ranged in age from 19 – 89 years old (M = 41.61, Med = 39, SD = 13.18) and 396 subjects (69%) spoke English as a first language. 169 subjects withdrew from the experiment before giving their demographic information.

Demographic Information for Retained Subjects in Experiment 1

In Experiment 1 we retained data from 292 subjects (196 women, 94 men, 2 gender diverse subjects). Subjects ranged in age from 19 – 89 years old (M = 43.85, Med = 41, SD = 13.86) and 282 subjects (97%) spoke English as a first language.

Demographic Information for all Subjects in Experiment 2

In Experiment 2 there were 445 subjects (295 women, 121 men, 4 gender diverse subjects). Subjects ranged in age from 18 – 84 years old (M = 43.18, Med = 41, SD = 13.98) and 400 subjects (90%) spoke English as a first language. 25 subjects withdrew from the experiment before giving their demographic information.

Demographic Information for Retained Subjects in Experiment 2

In Experiment 2 we retained data from 208 subjects (156 women, 49 men and 3 gender diverse). Subjects ranged in age from 21 – 84 years old (M = 46.02, Med = 44, SD = 14.12) and 198 subjects (95%) spoke English as a first language.

Demographic Information for all Subjects in Experiment 3

In Experiment 3 there were 409 subjects (269 women, 138 men, 2 gender diverse subjects). Subjects ranged in age from 18 – 84 years old (M = 42.89, Med = 41, SD = 14.31) and 394 subjects (96%) spoke English as a first language.

Demographic Information for Retained Subjects in Experiment 3

In Experiment 3 we retained data from 208 subjects (146 women, 60 men and 2 gender diverse subjects). Subjects ranged in age from 22 – 84 years old (M = 45.89, SD = 14.12, Median = 45) and 199 subjects (96%) spoke English as a first language.

Appendix 4: Supplemental Results

Supplemental Results for Experiment 1

We conducted Bonferroni pairwise *t* tests on the differences between people's assessments of truth for each type of claim in Experiment 1, as shown in Table S4.

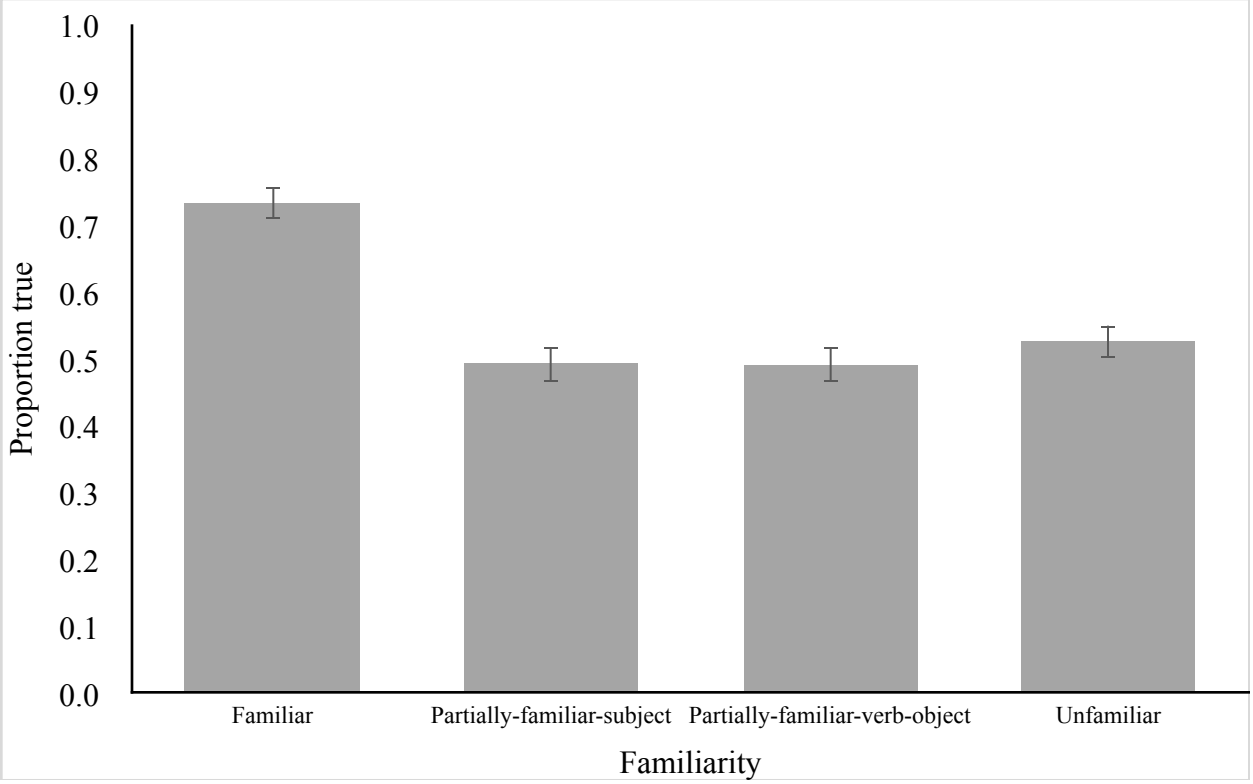
Table S4: Bonferroni pairwise t test for Experiment 1

Comparison	$M_{difference}$	95% CI lower limit	95% CI upper limit	SE	*t*	*p*
Familiar — partially-familiar-subject	0.29	0.24	0.34	0.02	15.63	<0.001
Familiar — partially-familiar-verb-object	0.30	0.25	0.35	0.02	15.98	<0.001
Familiar — unfamiliar	0.25	0.20	0.29	0.02	13.16	<0.001
Partially-familiar-subject — unfamiliar	0.05	-0.10	0.03	0.02	2.48	0.08
Partially-familiar-verb -object— unfamiliar	0.05	-0.09	0.00	0.02	2.82	0.03
Partially-familiar-subject — partially-familiar-verb-object	0.00	-0.04	0.06	0.02	0.35	1.00

Item Analysis for Experiment 1

We calculated the proportion of subjects in Experiment 1 who classified each claim true when that claim was familiar, partially-familiar-subject, partially-familiar-verb-object and unfamiliar. Those proportions are shown in Figure S1.

Figure S1: Proportion of subjects who classified the claims as true in Experiment 1



Using the claims as items of analysis, a within-items ANOVA showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 477) = 171.82$, $p<.001$, $\omega^2 = 0.27$. We conducted Bonferroni pairwise *t* tests, as shown in Table S5.

Table S5: Bonferroni pairwise t tests for the item analysis for Experiment 1

Comparison	$M_{difference}$	95% CI lower limit	95% CI upper limit	SE	*t*	*p*
Familiar — partially-familiar-subject	0.24	0.21	0.27	0.01	19.23	<0.001
Familiar — partially-familiar-verb-object	0.24	0.21	0.28	0.01	19.34	<0.001
Familiar — unfamiliar	0.21	0.17	0.24	0.01	16.48	<0.001
Partially-familiar-subject — unfamiliar	0.04	0.00	0.07	0.01	2.76	0.04
Partially-familiar-verb -object— unfamiliar	0.04	0.00	0.07	0.01	2.86	0.03
Partially-familiar-subject — partially-familiar-verb-object	0.00	-0.03	0.04	0.01	0.11	1.00

Supplemental Results for Experiment 2

We conducted Bonferroni pairwise *t* tests on the differences between people's assessments of truth for each type of claim in Experiment 2, as shown in Table S6.

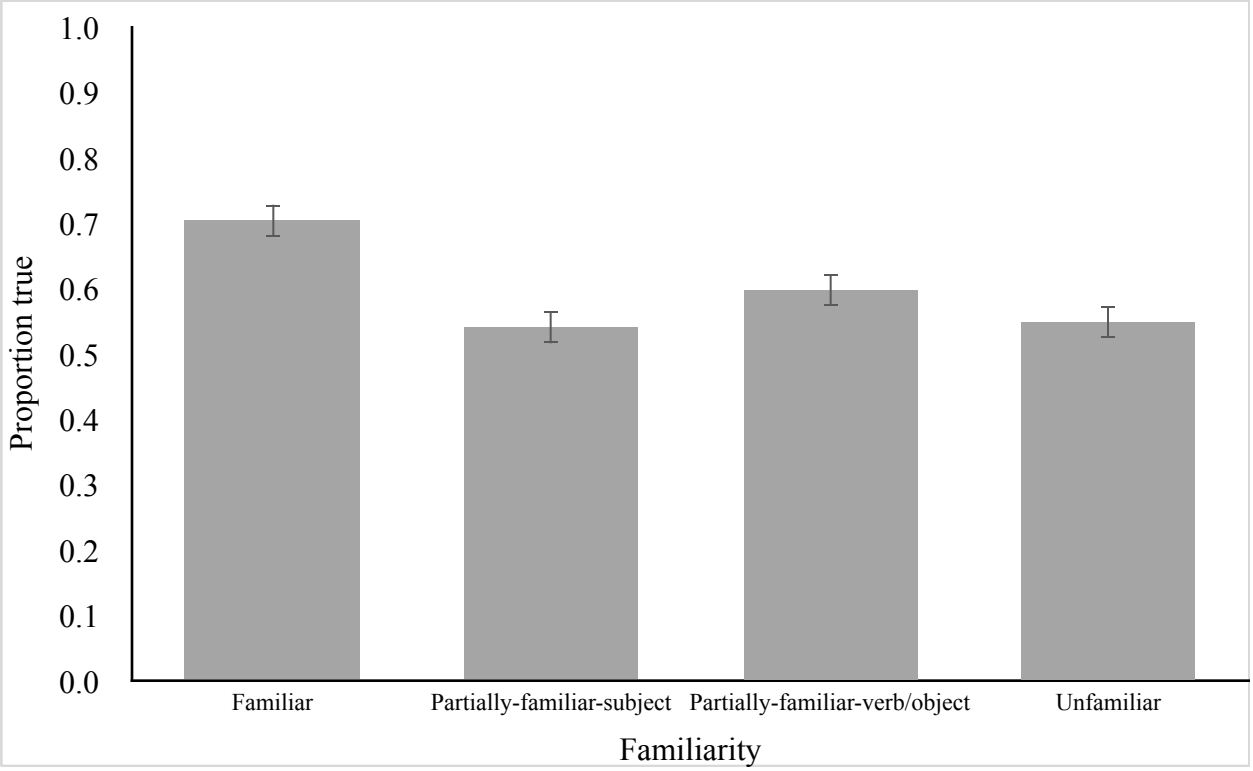
Table S6:Bonferroni pairwise t tests for Experiment 2

Paired comparison	$M_{difference}$	95% CI lower limit	95% CI upper limit	SE	*t*	*p*
Familiar – partially-familiar-subject	0.17	0.12	0.21	0.02	9.11	<0.001
Familiar– partially-familiar-verb-object	0.12	0.07	0.17	0.02	6.51	<0.001
Familiar – unfamiliar	0.15	0.10	0.20	0.02	8.31	<0.001
Partially-familiar-subject– unfamiliar	-0.01	-0.06	0.03	0.02	0.80	1.00
Partially-familiar-verb – unfamiliar	0.03	-0.02	0.08	0.02	1.81	0.43
Partially-familiar-subject — partially-familiar-verb-object	-0.05	-0.10	0.00	0.02	2.60	0.06

Item Analysis for Experiment 2

We calculated the proportion of subjects in Experiment 2 who rated each claim true when that claim was familiar, partially-familiar-subject, partially-familiar-verb-object and unfamiliar. Those proportions are shown in Figure S2.

Figure S2: Proportion of subjects who classified the claims as true in Experiment 2



Using the claims as items of analysis, a within-items ANOVA showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 477) = 51.985$, $p<.001$, $\omega^2 = 0.08$. We conducted Bonferroni pairwise *t* tests, as shown in Table S7.

Table S7: Bonferroni pairwise t tests for the item analysis for Experiment 2

Comparison	$M_{difference}$	95% CI lower limit	95% CI upper limit	SE	*t*	*p*
Familiar — partially-familiar-subject	0.16	0.12	0.20	0.02	11.11	<0.001
Familiar — partially-familiar-verb-object	0.11	0.07	0.15	0.02	7.30	<0.001
Familiar — unfamiliar	0.15	0.11	0.19	0.02	10.49	<0.001
Partially-familiar-subject — unfamiliar	-0.01	-0.05	0.03	0.02	0.62	0.54
Partially-familiar-verb -object— unfamiliar	0.05	0.01	0.09	0.02	3.19	0.003
Partially-familiar-subject — partially-familiar-verb-object	0.06	0.02	0.09	0.02	3.81	<0.001

Supplemental Results for Experiment 3

We conducted Bonferroni pairwise *t* tests on the differences between people's assessments of truth for each type of claim in Experiment 3, as shown in Table S8.

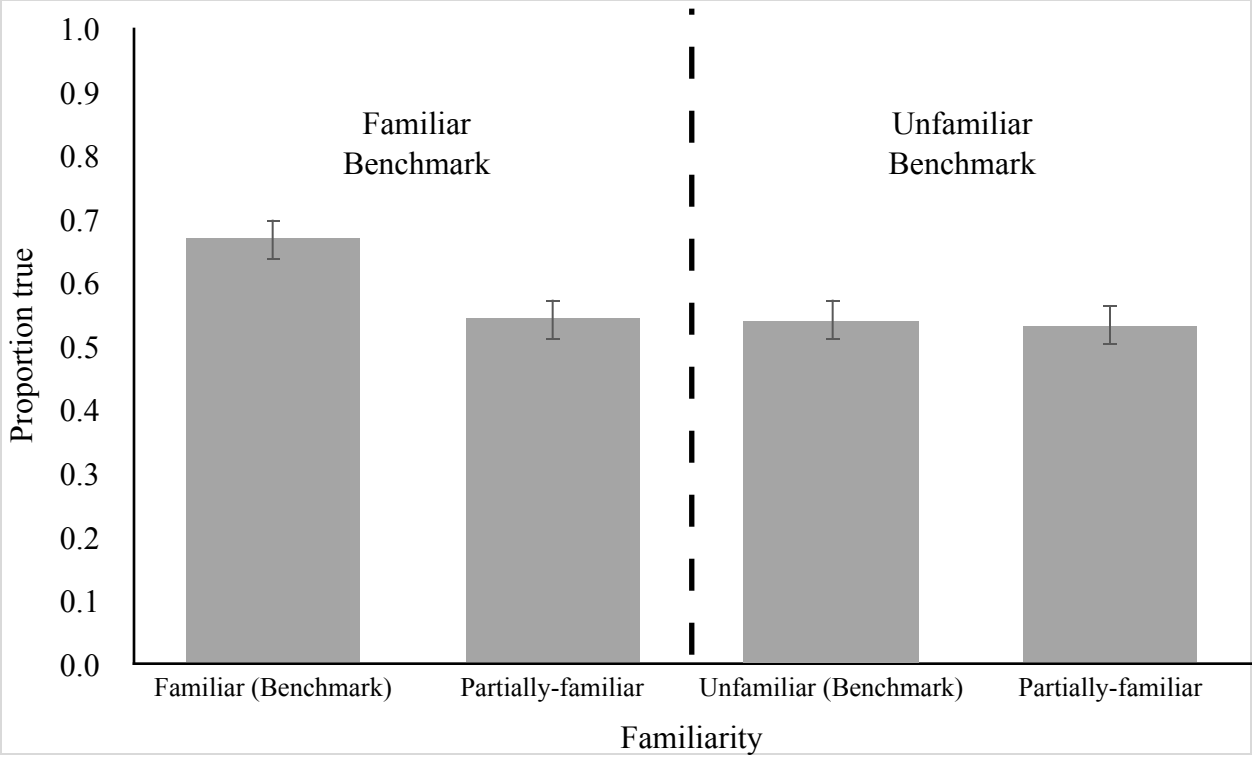
Table S8 *Bonferroni pairwise t tests for Experiment 3*

Paired comparison	$M_{difference}$	95% CI lower limit	95% CI upper limit	SE	*t*	*p*
Familiar – partially-familiar (Familiar Benchmark)	0.14	0.07	0.20	0.02	5.90	<0.001
Partially-familiar (Unfamiliar Benchmark) – Unfamiliar	0.01	-0.05	0.07	0.02	0.39	1.00
Partially-familiar (Familiar Benchmark) – Partially-familiar (Unfamiliar Benchmark)	0.004	-0.06	0.07	0.03	0.15	1.00
Familiar — Unfamiliar	0.13	0.07	0.20	0.03	5.22	<0.001
Partially-familiar (Familiar Benchmark) — Unfamiliar	0.01	-0.06	0.07	0.03	0.22	1.00
Partially-familiar (Unfamiliar Benchmark) — Familiar	0.14	0.07	0.21	0.03	5.59	<0.001

Item Analysis for Experiment 3

We calculated the proportion of subjects in Experiment 3 who rated each claim true when that claim was familiar, partially-familiar and unfamiliar. Those proportions are shown in Figure S3.

Figure S3: Proportion of subjects who rated the claims as true in Experiment 3



Using the claims as items of analysis, within-items ANOVA showed a significant effect of familiarity on peoples' assessments of truth, $F(3, 477) = 32.559$, $p<.001$, $\omega^2 = 0.06$. We conducted Bonferroni pairwise *t* tests, as shown in Table S9.

Table S9: Bonferroni pairwise t tests for the item analysis for Experiment 3

Comparison	$M_{difference}$	95% CI lower limit	95% CI upper limit	SE	*t*	*p*
Familiar – partially-familiar (Familiar Benchmark)	0.13	0.08	0.17	0.02	7.83	<0.001
Partially-familiar (Unfamiliar Benchmark) – Unfamiliar	-0.01	-0.05	0.03	0.02	0.53	1.00
Partially-familiar (Familiar Benchmark) – Partially-familiar (Unfamiliar Benchmark)	0.01	-0.03	0.05	0.02	0.60	1.00
Familiar — Unfamiliar	0.13	0.08	0.17	0.02	7.90	<0.001
Partially-familiar (Familiar Benchmark) — Unfamiliar	0.00	-0.04	0.04	0.02	0.07	1.00
Partially-familiar (Unfamiliar Benchmark) — Familiar	0.13	0.09	0.18	0.02	8.43	<0.001

Appendix 5: Materials

Full List of Trivia Claims

Each claim is presented here with data from the Stimulus Validation Study: number of judgements (N), proportion true (P) and standard deviation (SD).

Table S10: Full list of trivia claims

Group	Claim	Type	T/F	N	P	SD
1	The Nile River flows northwards	AB	T	40	0.75	0.44
1	The Nile River is the longest River in Asia	AD	F	48	0.46	0.50
1	The Yangtze River flows northwards	CB	F	42	0.45	0.50
1	The Yangtze River is the longest river in Asia	CD	T	48	0.65	0.48
2	Taboga Island is in Panama	AB	T	40	0.70	0.46
2	Taboga Island is part of Indonesia	AD	F	48	0.69	0.47
2	The Cocos Islands are in Panama	CB	F	42	0.60	0.50
2	The Cocos Islands are part of Indonesia	CD	F	48	0.67	0.48
3	Whale sharks subsist on plankton	AB	T	40	0.70	0.46
3	Whale sharks are carnivorous	AD	F	48	0.56	0.50
3	Most sea turtles subsist on plankton	CB	F	42	0.62	0.49
3	Most sea turtles are carnivorous	CD	T	48	0.54	0.50
4	Tomatoes are a good source of the phytochemical lycopene	AB	T	40	0.83	0.38
4	Tomatoes contain toxins	AD	F	48	0.54	0.50
4	Rhubarb leaves are a good source of the phytochemical lycopene	CB	F	42	0.62	0.49
4	Rhubarb leaves contain toxins	CD	T	48	0.71	0.46
5	Polaris is the name of the north star	AB	T	40	0.95	0.22

5	Polaris is the brightest star in the sky excluding the sun	AD	F	48	0.75	0.44
5	Sirius is the name of the north star	CB	F	42	0.74	0.45
5	Sirius is the brightest star in the sky excluding the sun	CD	T	48	0.58	0.50
6	Kava is a beverage made from the root of the pepper plant	AB	T	40	0.45	0.50
6	Kava is an alcoholic beverage fermented from honey	AD	F	48	0.58	0.50
6	Mead is a beverage made from the root of the pepper plant	CB	F	42	0.31	0.47
6	Mead is an alcoholic beverage fermented from honey	CD	T	48	0.75	0.44
7	The banana plant is a large herb	AB	T	40	0.38	0.49
7	The banana plant is in the mustard family	AD	F	48	0.21	0.41
7	Cabbages are a large herb	CB	F	42	0.38	0.49
7	Cabbages are in the mustard family	CD	T	48	0.46	0.50
8	Starfish don’t have brains	AB	T	40	0.68	0.47
8	Starfish have small, circular gills	AD	F	48	0.75	0.44
8	Eels don't have brains	CB	F	42	0.55	0.50
8	Eels have small, circular gills	CD	T	48	0.69	0.47
9	Plato is the name of Socrates' most famous student	AB	T	40	0.78	0.42
9	Plato is known as the "Father of Geometry"	AD	F	48	0.50	0.51
9	Euclid is the name of Socrates' most famous student	CB	F	42	0.57	0.50
9	Euclid is known as the "Father of Geometry"	CD	T	48	0.69	0.47
10	Snowboarding is believed to have originated in the United States	AB	T	40	0.78	0.42

10	Snowboarding has been traditionally credited to a Belgian	AD	F	48	0.60	0.49
10	The invention of roller skates is believed to have originated in the United States	CB	F	42	0.81	0.40
10	The invention of roller skates has been traditionally credited to a Belgian	CD	T	48	0.69	0.47
11	A panda's rear paws point inwards	AB	T	40	0.70	0.46
11	Pandas use their rear paws primarily for mating	AD	F	48	0.54	0.50
11	A walrus's tusks point inwards	CB	F	42	0.57	0.50
11	Walruses use their tusks primarily for mating	CD	T	48	0.50	0.51
12	The otter belongs to the weasel family	AB	T	40	0.83	0.38
12	The otter belongs to the cavy family	AD	F	48	0.44	0.50
12	Guinea pigs belong to the weasel family	CB	F	42	0.45	0.50
12	Guinea pigs belong to the cavy family	CD	T	48	0.52	0.50
13	The stationary ball in lawn bowls is called a jack	AB	T	40	0.60	0.50
13	The stationary ball in lawn bowls is concave on the bottom	AD	F	48	0.50	0.51
13	The stones used in curling are called jacks	CB	F	42	0.50	0.51
13	The stones used in curling are concave on the bottom	CD	T	48	0.58	0.50
14	The plastic things on the ends of shoelaces are called aglets	AB	T	40	0.80	0.41
14	The plastic things on the ends of shoelaces are called tines	AD	F	48	0.63	0.49
14	The metal spikes on forks are called aglets	CB	F	42	0.52	0.51

14	The metal spikes on forks are called tines	CD	T	48	0.83	0.38
15	Giraffes have excellent eyesight	AB	T	40	0.68	0.47
15	Giraffes can walk and run in all directions	AD	F	48	0.56	0.50
15	Crabs have excellent eyesight	CB	F	42	0.57	0.50
15	Crabs can walk and run in all directions	CD	T	48	0.73	0.45
16	Halvah is a confection made of sesame seeds	AB	T	40	0.73	0.45
16	Halvah is a dish from Africa	AD	F	48	0.42	0.50
16	Couscous is a confection made of sesame seeds	CB	F	42	0.33	0.48
16	Couscous is a dish from Africa	CD	T	48	0.67	0.48
17	Sticky toffee pudding is a classic British dessert	AB	T	40	0.90	0.30
17	Sticky toffee pudding has more sugar than ice cream	AD	F	48	0.88	0.33
17	Sherbert is a classic British dessert	CB	F	42	0.69	0.47
17	Sherbert has more sugar than ice cream	CD	T	48	0.63	0.49
18	Moose are a type of ruminant animal	AB	T	40	0.63	0.49
18	Moose can weigh up to 350kg/770lb	AD	F	48	0.75	0.44
18	Pigs are a type of ruminant animal	CB	F	42	0.57	0.50
18	Pigs can weigh up to 350kg/770lb	CD	T	48	0.79	0.41
19	Mercury is the liquid metal inside a thermometer	AB	T	40	0.88	0.33
19	Mercury is the fourth most common element on Earth	AD	F	48	0.65	0.48
19	Magnesium is the liquid metal inside a thermometer	CB	F	42	0.24	0.43
19	Magnesium is the fourth most common element on Earth	CD	T	48	0.65	0.48

20	Lake Baikal is the world's largest lake by volume	AB	T	40	0.60	0.50
20	Lake Baikal holds 10% of the world's surface fresh water	AD	F	48	0.58	0.50
20	Lake Superior is the world's largest lake by volume	CB	F	42	0.69	0.47
20	Lake Superior holds 10% of the world's surface fresh water	CD	T	48	0.71	0.46
21	Cats are the only pet not mentioned in the Bible	AB	T	40	0.48	0.51
21	Cats are the most common pet in the world	AD	F	48	0.67	0.48
21	Dogs are the only pet not mentioned in the Bible	CB	F	42	0.40	0.50
21	Dogs are the most common pet in the world	CD	T	48	0.85	0.36
22	Mount Vesuvius is an active volcano in Italy	AB	T	40	0.90	0.30
22	The Vesuvius Mountains form a high wall between France and Spain	AD	F	48	0.63	0.49
22	Mount Pyrenees is an active volcano in Italy	CB	F	42	0.62	0.49
22	The Pyrenees Mountains form a high wall between France and Spain	CD	T	48	0.56	0.50
23	Sisal is commonly used to make dart boards	AB	T	40	0.60	0.50
23	Sisal is the lightest known wood	AD	F	48	0.65	0.48
23	Balsa is commonly used to make dart boards	CB	F	42	0.60	0.50
23	Balsa is the lightest known wood	CD	T	48	0.75	0.44
24	The highest waterfall in the world is in Venezuela	AB	T	40	0.73	0.45

24	The highest waterfall in the world is in Argentina	AD	F	48	0.69	0.47
24	The second largest country in South America is Venezuela	CB	F	42	0.48	0.51
24	The second largest country in South America is Argentina	CD	T	48	0.83	0.38
25	Volleyball was originally called mintonette	AB	T	40	0.53	0.51
25	Volleyball has been traced back to the baths of Rome	AD	F	48	0.58	0.50
25	Handball was originally called mintonette	CB	F	42	0.52	0.51
25	Handball has been traced back to the baths of Rome	CD	T	48	0.58	0.50
26	Cross country skiing has been included in the Winter Olympic Games since 1924	AB	T	40	0.78	0.42
26	Cross country skiing is the navigation of large bumps on the ski slope	AD	F	48	0.56	0.50
26	Mogul skiing has been included in the Winter Olympic Games since 1924	CB	F	42	0.69	0.47
26	Mogul skiing is the navigation of large bumps on the ski slope	CD	T	48	0.65	0.48
27	Greenland is a part of the Kingdom of Denmark	AB	T	40	0.58	0.50
27	Greenland is the most sparsely populated country in Europe	AD	F	48	0.88	0.33
27	Iceland is a part of the Kingdom of Denmark	CB	F	42	0.55	0.50
27	Iceland is the most sparsely populated country in Europe	CD	T	48	0.67	0.48
28	Both sexes of lion are polygamous	AB	T	40	0.48	0.51
28	Both sexes of lion hunt in packs	AD	F	48	0.60	0.49

28	Hyenas are polygamous	CB	F	42	0.74	0.45
28	Hyenas hunt in packs	CD	T	48	0.83	0.38
29	Mount Everest is located in the Himalayas mountain range	AB	T	40	0.90	0.30
29	Mount Everest is the largest non-volcanic mountain in base area	AD	F	48	0.83	0.38
29	Mount Logan is located in the Himalayas mountain range	CB	F	42	0.50	0.51
29	Mount Logan is the largest non-volcanic mountain in base area	CD	T	48	0.52	0.50
30	Glial cells function primarily to support neurons	AB	T	40	0.58	0.50
30	Glial cells work to fight off infections	AD	F	48	0.54	0.50
30	Lymphocyte cells function primarily to support neurons	CB	F	42	0.64	0.48
30	Lymphocyte cells work to fight off infections	CD		48	0.88	0.33
31	Asia has the highest average elevation of the continents	AB	T	40	0.80	0.41
31	South America has the highest average elevation of the continents	AD	F	48	0.71	0.46
31	Asia contains the world's longest mountain range	CB	F	42	0.79	0.42
31	South America contains the world's longest mountain range	CD	T	48	0.65	0.48
32	Domesticated goats are descended from the pasang	AB	T	40	0.40	0.50
32	Domesticated goats have a divided upper lip	AD	F	48	0.46	0.50
32	Sheep are descended from the pasang	CB	F	42	0.40	0.50
32	Sheep have a divided upper lip	CD	T	48	0.67	0.48

33	California produces most of the world's almonds	AB	T	40	0.78	0.42
33	California is the world's third largest producer of olives	AD	F	48	0.79	0.41
33	Morocco produces most of the world's almonds	CB	F	42	0.62	0.49
33	Morocco is the world's third largest producer of olives	CD	T	48	0.65	0.48
34	Venice is the city in Italy known for its canals	AB	T	40	0.95	0.22
34	Venice is the location of Michelangelo's statue of David	AD	F	48	0.67	0.48
34	Florence is the city in Italy known for its canals	CB	F	42	0.67	0.48
34	Florence is the location of Michelangelo's statue of David	CD	T	48	0.71	0.46
35	Bagels are made from boiled dough	AB	T	40	0.73	0.45
35	Bagels originated in 13th Century Austria	AD	F	48	0.60	0.49
35	Croissants are made from boiled dough	CB	F	42	0.43	0.50
35	Croissants originated in 13th Century Austria	CD	T	48	0.56	0.50
36	Sea urchins have their mouths on their underside	AB	T	40	0.88	0.33
36	Sea urchins produce a poison called holothurin	AD	F	48	0.83	0.38
36	Sea cucumbers have their mouths on their underside	CB	F	42	0.64	0.48
36	Sea cucumbers produce a poison called holothurin	CD	T	48	0.60	0.49
37	Vitamin A deficiency can cause night blindness	AB	T	40	0.80	0.41

37	Vitamin A deficiency can lead to a fear of water	AD	F	48	0.23	0.42
37	Rabies can cause night blindness	CB	F	42	0.60	0.50
37	Rabies can lead to a fear of water	CD	T	48	0.54	0.50
38	Amino acids are the essential building blocks of all proteins - [Field-1]	AB	T	40	0.90	0.30
38	Amino acids generate chemical energy in cells	AD	F	48	0.77	0.42
38	Mitochondria are the essential building blocks of all proteins	CB	F	42	0.69	0.47
38	Mitochondria generate chemical energy in cells	CD	T	48	0.85	0.36
39	A broch is the glow around the moon	AB	T	40	0.40	0.50
39	A broch is the space between your eyebrows	AD	F	48	0.50	0.51
39	A glabella is the glow around the moon	CB	F	42	0.48	0.51
39	A glabella is the space between your eyebrows	CD	T	48	0.50	0.51
40	Mount Rushmore cost less than $1 million	AB	T	40	0.38	0.49
40	Mount Rushmore was built in the 19th Century	AD	F	48	0.69	0.47
40	The Eiffel Tower cost less than $1 million	CB	F	42	0.50	0.51
40	The Eiffel Tower was built in the 19th Century	CD	T	48	0.63	0.49
41	Verkhoyansk is a town in northeastern Russia	AB	T	40	0.85	0.36
41	Verkhoyansk is a dormant volcano in Tanzania	AD	F	48	0.63	0.49
41	Kilimanjaro is a town in northeastern Russia	CB	F	42	0.33	0.48

41	Kilimanjaro is a dormant volcano in Tanzania	CD	T	48	0.73	0.45
42	Phosphenes are the stars and flashes of light you see when you rub your eyes	AB	T	40	0.68	0.47
42	Phosphene is the water-transporting tissues in plants	AD	F	48	0.67	0.48
42	Xylems are the stars and flashes of light you see when you rub your eyes	CB	F	42	0.52	0.51
42	Xylem is the water-transporting tissues in plants	CD	T	48	0.63	0.49
43	Plasma is the noncellular portion of blood	AB	T	40	0.83	0.38
43	Plasma is a hormone secreted by the pancreas	AD	F	48	0.35	0.48
43	Glucagon is the noncellular portion of blood	CB	F	42	0.67	0.48
43	Glucagon is a hormone secreted by the pancreas	CD	T	48	0.56	0.50
44	Finland has more saunas than cars	AB	T	40	0.78	0.42
44	Finland has more sheep than people	AD	F	48	0.69	0.47
44	New Zealand has more saunas than cars	CB	F	42	0.48	0.51
44	New Zealand has more sheep than people	CD	T	48	0.73	0.45
45	The Kentucky Derby is the oldest sporting event in the United States	AB	T	40	0.80	0.41
45	The Kentucky Derby first occurred in 1903	AD	F	48	0.73	0.45
45	The World Series is the oldest sporting event in the United States	CB	F	42	0.55	0.50
45	The World Series first occurred in 1903	CD	T	48	0.75	0.44
46	The steeplechase is a footrace over an obstacle course	AB	T	40	0.60	0.50

46	The steeplechase is a downhill sledding event over ice	AD	F	48	0.44	0.50
46	The luge is a footrace over an obstacle course	CB	F	42	0.26	0.45
46	The luge is a downhill sledding event over ice	CD		48	0.77	0.42
47	Carrots come in four different colors	AB	T	40	0.65	0.48
47	Carrots are a member of the sunflower family	AD	F	48	0.13	0.33
47	Lettuce comes in four different colors	CB	F	42	0.60	0.50
47	Lettuce is a member of the sunflower family	CD	T	48	0.27	0.45
48	Vanilla flavoring comes from the pod of an orchid	AB	T	40	0.55	0.50
48	Vanilla flavoring comes from the bark of a willow tree	AD	F	48	0.40	0.49
48	Aspirin comes from the pod of an orchid	CB	F	42	0.19	0.40
48	Aspirin comes from the bark of a willow tree	CD	T	48	0.48	0.50
49	Hippopotamuses have the largest mouths of any animal on land	AB	T	40	0.90	0.30
49	Hippopotamuses are the only mammals that can't jump	AD	F	48	0.48	0.50
49	Elephants have the largest mouths of any animal on land	CB	F	42	0.36	0.48
49	Elephants are the only mammals that can't jump	CD	T	48	0.50	0.51
50	Puffins use twigs to scratch their bodies	AB	T	40	0.58	0.50
50	Puffins swallow air to become buoyant when they swim	AD	F	48	0.75	0.44
50	Armadillos use twigs to scratch their bodies	CB	F	42	0.48	0.51

50	Armadillos swallow air to become buoyant when they swim	CD	T	48	0.65	0.48
51	The platypus is one of the only mammals to lay eggs	AB	T	40	0.73	0.45
51	The platypus sleeps more than any other mammal on Earth	AD	F	48	0.38	0.49
51	Little brown bats are one of the only mammals to lay eggs	CB	F	42	0.36	0.48
51	Little brown bats sleep more than any other mammal on earth	CD	T	48	0.69	0.47
52	Sperm whales are the loudest animal on earth	AB	T	40	0.60	0.50
52	Sperm whales have transparent eyelids so they can see underwater	AD	F	48	0.77	0.42
52	Beavers are the loudest animal on earth	CB	F	42	0.24	0.43
52	Beavers have transparent eyelids so they can see underwater	CD	T	48	0.65	0.48
53	A snake's eyelids are immovable	AB	T	40	0.58	0.50
53	A snake's eyelids are twice as long as its body	AD	F	48	0.25	0.44
53	A chameleon's tongue is immovable	CB	F	42	0.24	0.43
53	A chameleon's tongue is twice as long as its body	CD	T	48	0.71	0.46
54	Beef is the most popular meat in the United States	AB	T	40	0.83	0.38
54	Ham is the most popular meat in the United States	AD	F	48	0.42	0.50
54	Beef is the rear leg of a hog	CB	F	42	0.12	0.33
54	Ham is the rear leg of a hog	CD	T	48	0.65	0.48
55	Potatoes were first cultivated in Peru	AB	T	40	0.73	0.45
55	Corn was first cultivated in Peru	AD	F	48	0.67	0.48

55	Potatoes are the most widely grown crop in the United States	CB	F	42	0.52	0.51
55	Corn is the most widely grown crop in the United States	CD	T	48	0.67	0.48
56	The anaconda is the largest snake in the world	AB	T	40	0.95	0.22
56	The anaconda usually lives for 20 years	AD	F	48	0.75	0.44
56	The cobra is the largest snake in the world	CB	F	42	0.38	0.49
56	The cobra usually lives for 20 years	CD	T	48	0.60	0.49
57	Mayonnaise is usually made with raw egg yolks	AB	T	40	0.85	0.36
57	Mayonnaise was first served in the United States in 1892	AD	F	48	0.71	0.46
57	Chocolate mousse is usually made with raw egg yolks	CB	F	42	0.55	0.50
57	Chocolate mousse was first served in the United States in 1892	CD	T	48	0.77	0.42
58	Bandy is a game similar to ice hockey	AB	T	40	0.45	0.50
58	Sipà is a game similar to ice hockey	AD	F	48	0.50	0.51
58	Bandy is a traditional sport in the Philippines	CB	F	42	0.62	0.49
58	Sipà is a traditional sport in the Philippines	CD	T	48	0.60	0.49
59	Pluto is part of the Kuiper belt	AB	T	40	0.70	0.46
59	Black holes are part of the Kuiper belt	AD	F	48	0.44	0.50
59	Pluto formed when a dying star collapsed in on itself	CB	F	42	0.57	0.50
59	Black holes form when a dying star collapses in on itself	CD	T	48	0.83	0.38
60	Algeria is a predominantly Muslim country	AB	T	40	0.83	0.38

60	Madagascar is a predominantly Muslim country	AD	F	48	0.50	0.51
60	Algeria has French as an official language	CB	F	42	0.64	0.48
60	Madagascar has French as an official language	CD	T	48	0.73	0.45
61	Deer do not have gall bladders	AB	T	40	0.53	0.51
61	Salmon do not have gall bladders	AD	F	48	0.69	0.47
61	Deer use the Earth’s magnetic field to navigate	CB	F	42	0.55	0.50
61	Salmon use the Earth’s magnetic field to navigate	CD	T	48	0.71	0.46
62	The Arctic Ocean is centered approximately on the North Pole	AB	T	40	0.80	0.41
62	Taylor Glacier is centered approximately on the North Pole	AD	F	48	0.69	0.47
62	The Arctic Ocean is a breeding habitat for penguins	CB	F	42	0.81	0.40
62	Taylor Glacier is a breeding habitat for penguins	CD	T	48	0.69	0.47
63	Gnu is another name for wildebeest	AB	T	40	0.58	0.50
63	Gnu are a type of flying insect	AD	F	48	0.38	0.49
63	Tachinid is another name for wildebeest	CB	F	42	0.33	0.48
63	Tachinids are a type of flying insect	CD	T	48	0.58	0.50
64	Many of the genes in baker's yeast are also present in humans	AB	T	40	0.58	0.50
64	Many of the genes in baker's yeast continue growing for hundreds of years	AD	F	48	0.75	0.44
64	Some kinds of plant cells are also present in humans	CB	F	42	0.45	0.50

64	Some kinds of plant cells continue growing for hundreds of years	CD	T	48	0.77	0.42
65	Canada is the second largest country in the world in area	AB	T	40	0.63	0.49
65	Canada has been inhabited by humans since 6,000 BC	AD	F	48	0.69	0.47
65	Brazil is the second largest country in the world in area	CB	F	42	0.38	0.49
65	Brazil has been inhabited by humans since 6,000 BC	CD	T	48	0.75	0.44
66	Oysters have been fished in Colchester, England, since before the Roman Empire	AB	T	40	0.80	0.41
66	Oysters are the biggest food export in the Bahamas	AD	F	48	0.54	0.50
66	Lobsters have been fished in Colchester, England, since before the Roman Empire	CB	F	42	0.74	0.45
66	Lobsters are the biggest food export in the Bahamas	CD	T	48	0.50	0.51
67	Figs grow without blossoms on their branches	AB	T	40	0.60	0.50
67	Figs have more acid than lemons	AD	F	48	0.38	0.49
67	Most limes grow without blossoms on their branches	CB	F	42	0.60	0.50
67	Most limes have more acid than lemons	CD	T	48	0.73	0.45
68	Tigers use their stripes as camouflage	AB	T	40	0.80	0.41
68	Clownfish use their stripes as camouflage	AD	F	48	0.85	0.36
68	Tigers use their stripes to recognize each other	CB	F	42	0.69	0.47
68	Clownfish use their stripes to recognize each other	CD	T	48	0.69	0.47
69	The Dead Sea is the lowest body of water on the surface of Earth	AB	T	40	0.90	0.30

69	The Dead Sea sits on the border of three tectonic plates	AD	F	48	0.65	0.48
69	The Banda Sea is the lowest body of water on the surface of Earth	CB	F	42	0.62	0.49
69	The Banda Sea sits on the border of three tectonic plates	CD	T	48	0.56	0.50
70	Badminton is usually played indoors	AB	T	40	0.50	0.51
70	Badminton originated in Medieval France	AD	F	48	0.79	0.41
70	Tennis is usually played indoors	CB	F	42	0.26	0.45
70	Tennis originated in Medieval France	CD	T	48	0.69	0.47

Trivia Claims Used in Experiments 1 - 3

Each claim is presented here with data from the Stimulus Validation Study: number of judgements (N), proportion true (P) and standard deviation (SD).

Table S11: List of trivia claims used in Experiments 1 - 3

Group	Claim	Type	T/F	N	P	SD
1	The Nile River flows northwards	AB1	T	40	0.75	0.44
1	The Nile River is the longest River in Asia	AD1	F	48	0.46	0.50
1	The Yangtze River flows northwards	CB1	F	42	0.45	0.50
1	The Yangtze River is the longest river in Asia	CD1	T	48	0.65	0.48
3	Whale sharks subsist on plankton	AB1	T	40	0.70	0.46
3	Whale sharks are carnivorous	AD1	F	48	0.56	0.50
3	Most sea turtles subsist on plankton	CB1	F	42	0.62	0.49
3	Most sea turtles are carnivorous	CD1	T	48	0.54	0.50
6	Kava is a beverage made from the root of the pepper plant	AB1	T	40	0.45	0.50
6	Kava is an alcoholic beverage fermented from honey	AD1	F	48	0.58	0.50

6	Mead is a beverage made from the root of the pepper plant	CB1	F	42	0.31	0.47
6	Mead is an alcoholic beverage fermented from honey	CD1	T	48	0.75	0.44
7	The banana plant is a large herb	AB1	T	40	0.38	0.49
7	The banana plant is in the mustard family	AD1	F	48	0.21	0.41
7	Cabbages are a large herb	CB1	F	42	0.38	0.49
7	Cabbages are in the mustard family	CD1	T	48	0.46	0.50
8	Starfish don’t have brains	AB1	T	40	0.68	0.47
8	Starfish have small, circular gills	AD1	F	48	0.75	0.44
8	Eels don't have brains	CB1	F	42	0.55	0.50
8	Eels have small, circular gills	CD1	T	48	0.69	0.47
9	Plato is the name of Socrates' most famous student	AB1	T	40	0.78	0.42
9	Plato is known as the "Father of Geometry"	AD1	F	48	0.50	0.51
9	Euclid is the name of Socrates' most famous student	CB1	F	42	0.57	0.50
9	Euclid is known as the "Father of Geometry"	CD1	T	48	0.69	0.47
11	A panda's rear paws point inwards	AB1	T	40	0.70	0.46
11	Pandas use their rear paws primarily for mating	AD1	F	48	0.54	0.50
11	A walrus's tusks point inwards	CB1	F	42	0.57	0.50
11	Walruses use their tusks primarily for mating	CD1	T	48	0.50	0.51
13	The stationary ball in lawn bowls is called a jack	AB1	T	40	0.60	0.50
13	The stationary ball in lawn bowls is concave on the bottom	AD1	F	48	0.50	0.51

13	The stones used in curling are called jacks	CB1	F	42	0.50	0.51
13	The stones used in curling are concave on the bottom	CD1	T	48	0.58	0.50
15	Giraffes have excellent eyesight	AB1	T	40	0.68	0.47
15	Giraffes can walk and run in all directions	AD1	F	48	0.56	0.50
15	Crabs have excellent eyesight	CB1	F	42	0.57	0.50
15	Crabs can walk and run in all directions	CD1	T	48	0.73	0.45
16	Halvah is a confection made of sesame seeds	AB1	T	40	0.73	0.45
16	Halvah is a dish from Africa	AD1	F	48	0.42	0.50
16	Couscous is a confection made of sesame seeds	CB1	F	42	0.33	0.48
16	Couscous is a dish from Africa	CD1	T	48	0.67	0.48
18	Moose are a type of ruminant animal	AB2	T	40	0.63	0.49
18	Moose can weigh up to 350kg/770lb	AD2	F	48	0.75	0.44
18	Pigs are a type of ruminant animal	CB2	F	42	0.57	0.50
18	Pigs can weigh up to 350kg/770lb	CD2	T	48	0.79	0.41
20	Lake Baikal is the world's largest lake by volume	AB2	T	40	0.60	0.50
20	Lake Baikal holds 10% of the world's surface fresh water	AD2	F	48	0.58	0.50
20	Lake Superior is the world's largest lake by volume	CB2	F	42	0.69	0.47
20	Lake Superior holds 10% of the world's surface fresh water	CD2	T	48	0.71	0.46
23	Sisal is commonly used to make dart boards	AB2	T	40	0.60	0.50
23	Sisal is the lightest known wood	AD2	F	48	0.65	0.48
23	Balsa is commonly used to make dart boards	CB2	F	42	0.60	0.50
23	Balsa is the lightest known wood	CD2	T	48	0.75	0.44
25	Volleyball was originally called mintonette	AB2	T	40	0.53	0.51

25	Volleyball has been traced back to the baths of Rome	AD2	F	48	0.58	0.50
25	Handball was originally called mintonette	CB2	F	42	0.52	0.51
25	Handball has been traced back to the baths of Rome	CD2	T	48	0.58	0.50
26	Cross country skiing has been included in the Winter Olympic Games since 1924	AB2	T	40	0.78	0.42
26	Cross country skiing is the navigation of large bumps on the ski slope	AD2	F	48	0.56	0.50
26	Mogul skiing has been included in the Winter Olympic Games since 1924	CB2	F	42	0.69	0.47
26	Mogul skiing is the navigation of large bumps on the ski slope	CD2	T	48	0.65	0.48
31	Asia has the highest average elevation of the continents	AB2	T	40	0.80	0.41
31	South America has the highest average elevation of the continents	AD2	F	48	0.71	0.46
31	Asia contains the world's longest mountain range	CB2	F	42	0.79	0.42
31	South America contains the world's longest mountain range	CD2	T	48	0.65	0.48
32	Domesticated goats are descended from the pasang	AB2	T	40	0.40	0.50
32	Domesticated goats have a divided upper lip	AD2	F	48	0.46	0.50
32	Sheep are descended from the pasang	CB2	F	42	0.40	0.50
32	Sheep have a divided upper lip	CD2	T	48	0.67	0.48
33	California produces most of the world's almonds	AB2	T	40	0.78	0.42
33	California is the world's third largest producer of olives	AD2	F	48	0.79	0.41
33	Morocco produces most of the world's almonds	CB2	F	42	0.62	0.49
33	Morocco is the world's third largest producer of olives	CD2	T	48	0.65	0.48

35	Bagels are made from boiled dough	AB2	T	40	0.73	0.45
35	Bagels originated in 13th Century Austria	AD2	F	48	0.60	0.49
35	Croissants are made from boiled dough	CB2	F	42	0.43	0.50
35	Croissants originated in 13th Century Austria	CD2	T	48	0.56	0.50
37	Vitamin A deficiency can cause night blindness	AB2	T	40	0.80	0.41
37	Vitamin A deficiency can lead to a fear of water	AD2	F	48	0.23	0.42
37	Rabies can cause night blindness	CB2	F	42	0.60	0.50
37	Rabies can lead to a fear of water	CD2	T	48	0.54	0.50
39	A broch is the glow around the moon	AB3	T	40	0.40	0.50
39	A broch is the space between your eyebrows	AD3	F	48	0.50	0.51
39	A glabella is the glow around the moon	CB3	F	42	0.48	0.51
39	A glabella is the space between your eyebrows	CD3	T	48	0.50	0.51
40	Mount Rushmore cost less than $1 million	AB3	T	40	0.38	0.49
40	Mount Rushmore was built in the 19th Century	AD3	F	48	0.69	0.47
40	The Eiffel Tower cost less than $1 million	CB3	F	42	0.50	0.51
40	The Eiffel Tower was built in the 19th Century	CD3	T	48	0.63	0.49
42	Phosphenes are the stars and flashes of light you see when you rub your eyes	AB3	T	40	0.68	0.47
42	Phosphene is the water-transporting tissues in plants	AD3	F	48	0.67	0.48
42	Xylems are the stars and flashes of light you see when you rub your eyes	CB3	F	42	0.52	0.51

42	Xylem is the water-transporting tissues in plants	CD3	T	48	0.63	0.49
44	Finland has more saunas than cars	AB3	T	40	0.78	0.42
44	Finland has more sheep than people	AD3	F	48	0.69	0.47
44	New Zealand has more saunas than cars	CB3	F	42	0.48	0.51
44	New Zealand has more sheep than people	CD3	T	48	0.73	0.45
45	The Kentucky Derby is the oldest sporting event in the United States	AB3	T	40	0.80	0.41
45	The Kentucky Derby first occurred in 1903	AD3	F	48	0.73	0.45
45	The World Series is the oldest sporting event in the United States	CB3	F	42	0.55	0.50
45	The World Series first occurred in 1903	CD3	T	48	0.75	0.44
46	The steeplechase is a footrace over an obstacle course	AB3	T	40	0.60	0.50
46	The steeplechase is a downhill sledding event over ice	AD3	F	48	0.44	0.50
46	The luge is a footrace over an obstacle course	CB3	F	42	0.26	0.45
46	The luge is a downhill sledding event over ice	CD3	T	48	0.77	0.42
48	Vanilla flavoring comes from the pod of an orchid	AB3	T	40	0.55	0.50
48	Vanilla flavoring comes from the bark of a willow tree	AD3	F	48	0.40	0.49
48	Aspirin comes from the pod of an orchid	CB3	F	42	0.19	0.40
48	Aspirin comes from the bark of a willow tree	CD3	T	48	0.48	0.50
50	Puffins use twigs to scratch their bodies	AB3	T	40	0.58	0.50
50	Puffins swallow air to become buoyant when they swim	AD3	F	48	0.75	0.44
50	Armadillos use twigs to scratch their bodies	CB3	F	42	0.48	0.51

50	Armadillos swallow air to become buoyant when they swim	CD3	T	48	0.65	0.48
51	The platypus is one of the only mammals to lay eggs	AB3	T	40	0.73	0.45
51	The platypus sleeps more than any other mammal on Earth	AD3	F	48	0.38	0.49
51	Little brown bats are one of the only mammals to lay eggs	CB3	F	42	0.36	0.48
51	Little brown bats sleep more than any other mammal on earth	CD3	T	48	0.69	0.47
52	Sperm whales are the loudest animal on earth	AB3	T	40	0.60	0.50
52	Sperm whales have transparent eyelids so they can see underwater	AD3	F	48	0.77	0.42
52	Beavers are the loudest animal on earth	CB3	F	42	0.24	0.43
52	Beavers have transparent eyelids so they can see underwater	CD3	T	48	0.65	0.48
53	A snake's eyelids are immovable	AB4	T	40	0.58	0.50
53	A snake's eyelids are twice as long as its body	AD4	F	48	0.25	0.44
53	A chameleon's tongue is immovable	CB4	F	42	0.24	0.43
53	A chameleon's tongue is twice as long as its body	CD4	T	48	0.71	0.46
55	Potatoes were first cultivated in Peru	AB4	T	40	0.73	0.45
55	Corn was first cultivated in Peru	AD4	F	48	0.67	0.48
55	Potatoes are the most widely grown crop in the United States	CB4	F	42	0.52	0.51
55	Corn is the most widely grown crop in the United States	CD4	T	48	0.67	0.48
58	Bandy is a game similar to ice hockey	AB4	T	40	0.45	0.50
58	Sipà is a game similar to ice hockey	AD4	F	48	0.50	0.51
58	Bandy is a traditional sport in the Philippines	CB4	F	42	0.62	0.49
58	Sipà is a traditional sport in the Philippines	CD4	T	48	0.60	0.49

61	Deer do not have gall bladders	AB4	T	40	0.53	0.51
61	Salmon do not have gall bladders	AD4	F	48	0.69	0.47
61	Deer use the Earth's magnetic field to navigate	CB4	F	42	0.55	0.50
61	Salmon use the Earth's magnetic field to navigate	CD4	T	48	0.71	0.46
63	Gnu is another name for wildebeest	AB4	T	40	0.58	0.50
63	Gnu are a type of flying insect	AD4	F	48	0.38	0.49
63	Tachinid is another name for wildebeest	CB4	F	42	0.33	0.48
63	Tachinids are a type of flying insect	CD4	T	48	0.58	0.50
64	Many of the genes in baker's yeast are also present in humans	AB4	T	40	0.58	0.50
64	Many of the genes in baker's yeast continue growing for hundreds of years	AD4	F	48	0.75	0.44
64	Some kinds of plant cells are also present in humans	CB4	F	42	0.45	0.50
64	Some kinds of plant cells continue growing for hundreds of years	CD4	T	48	0.77	0.42
65	Canada is the second largest country in the world in area	AB4	T	40	0.63	0.49
65	Canada has been inhabited by humans since 6,000 BC	AD4	F	48	0.69	0.47
65	Brazil is the second largest country in the world in area	CB4	F	42	0.38	0.49
65	Brazil has been inhabited by humans since 6,000 BC	CD4	T	48	0.75	0.44
66	Oysters have been fished in Colchester, England, since before the Roman Empire	AB4	T	40	0.80	0.41
66	Oysters are the biggest food export in the Bahamas	AD4	F	48	0.54	0.50

66	Lobsters have been fished in Colchester, England, since before the Roman Empire	CB4	F	42	0.74	0.45
66	Lobsters are the biggest food export in the Bahamas	CD4	T	48	0.50	0.51
67	Figs grow without blossoms on their branches	AB4	T	40	0.60	0.50
67	Figs have more acid than lemons	AD4	F	48	0.38	0.49
67	Most limes grow without blossoms on their branches	CB4	F	42	0.60	0.50
67	Most limes have more acid than lemons	CD4	T	48	0.73	0.45
70	Badminton is usually played indoors	AB4	T	40	0.50	0.51
70	Badminton originated in Medieval France	AD4	F	48	0.79	0.41
70	Tennis is usually played indoors	CB4	F	42	0.26	0.45
70	Tennis originated in Medieval France	CD4	T	48	0.69	0.47

Appendix 6: Preregistration for Experiment 2

'*Illusions of truth for partially-familiar claims (48h delay)*'
(AsPredicted #78909)

Created: 11/03/2021 04:23 PM (PT)

Author(s)
Daniel McLennan (The University of Waikato) - dm226@students.waikato.ac.nz
Maryanne Garry (The University of waikato) - maryanne.garry@waikato.ac.nz

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Question: how do people come to assess the truth of partially-familiar claims, relative to both familiar and unfamiliar claims?

3) Describe the key dependent variable(s) specifying how they will be measured.
The dependent variable is the proportion of "true" responses in the second session. In each condition we will tally the number of claims, out of 10, that each person classified as true.

4) How many and which conditions will participants be assigned to?
We will use a within-subjects design with four conditions: familiar, partially-familiar (subject), partially-familiar (verb + object), unfamiliar.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will run a within-subjects ANOVA on mean proportion of true response in each condition from (4) above.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude participants' data from our analyses if they fail to complete the study, and if their responses indicate they were not paying attention. At the beginning of the experiment we will ask participants to complete the experiment in a single session without stopping, pausing, talking to other people, or searching the claims online. We will exclude data from participants who admit to doing any of those things at the end of the study. As an attention check, we will ask participants to briefly describe the task at the end of the experiment and exclude data from those who gave nonsensical descriptions (for example, "good survey"). We will also exclude data from participants who gave the same response to every claim.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
We will collect data from 400 participants

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Appendix 7: Preregistration for Experiment 3

'*Illusions of truth for partially-familiar claims (changing context)*'
(AsPredicted #83451)

Created: 12/16/2021 04:18 PM (PT)

Author(s)
Daniel McLennan (The University of Waikato) - dm226@students.waikato.ac.nz
Andrew Evelo (University of Waikato) - andrew.evelo@waikato.ac.nz
Maryanne Garry (The University of Waikato) - maryanne.garry@waikato.ac.nz

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Question: how do people come to assess the truth of partially-familiar claims, relative to either familiar or unfamiliar claims?

3) Describe the key dependent variable(s) specifying how they will be measured.
The dependent variable is the proportion of "true" responses in the second session. In each condition we will tally the number of claims, out of 20, that each person classified as true.

4) How many and which conditions will participants be assigned to?
There is one between-subjects factor with 2 levels (Context: familiar, unfamiliar) and one within-subjects factor with 2 levels (Comparison: partially-familiar, context). All subjects will make true/false ratings for 20 partially-familiar claims, along with either 20 familiar claims or 20 unfamiliar claims.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will conduct a 2 (Context: familiar, unfamiliar) x 2 (Comparison: partially-familiar, context) mixed factorial ANOVA on the proportion of claims subjects classify as true.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude a subject's data from our analyses if they fail to complete the study, and if their responses indicate they were not paying attention. At the beginning of the experiment we will ask subjects to complete the experiment in a single session without stopping, pausing, talking to other people, or searching the claims online. We will exclude data from subjects who admit to doing any of those things at the end of the study. As an attention check, we will ask subjects to briefly describe the task at the end of the experiment and exclude data from those who give nonsensical descriptions (for example, "good survey"). We will also exclude data from subjects who give the same response to every claim.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
We will collect data from 400 subjects.

8) Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
N/A