Behavior Research Methods
https://doi.org/10.3758/s13428-022-01904-6

Story stimuli for instantiating true and false beliefs about the world

Nikita A. Salovich[1] · Megan N. Imundo[2] · David N. Rapp[1,3]

Accepted: 7 June 2022

Abstract
We offer short story ("vignette") materials that have been developed and tested with the intention of influencing people's true and false beliefs about the world. First, we present norming data on the baseline rates at which participants from both U.S.-census matched and general U.S. online samples were correctly able to classify a selected set of accurate (e.g., aerobic exercise *strengthens* your heart and lungs) and inaccurate (e.g., aerobic exercise *weakens* your heart and lungs) assertions as "True" or "False." Next, we present data which validate that reading vignettes in which people discuss these accurate and inaccurate assertions influences participants' subsequent judgments of the validity of the asserted claims. These vignettes are brief, easy-to-read, allow for flexible and accountable online data collection, and reflect realistic accurate and inaccurate claims that people routinely encounter (e.g., preventative health behaviors, use of alternative medicines and therapies, etc.). As intended, vignettes containing inaccurate assertions increased participants' subsequent judgment errors, while vignettes containing accurate assertions decreased participants' subsequent judgment errors, both relative to participants' judgments after not reading related information. In an additional experiment, we used the vignette materials to replicate findings from Salovich et al. (2021), wherein participants reported lower confidence in correct judgments and higher confidence in incorrect judgments after having read inaccurate assertions. Overall, these materials are well suited for investigations on the consequences of exposures to accurate and inaccurate information, address limitations in currently available stimuli, and align with trends in research practice (e.g., online sampling) within psychological science.

Keywords Misinformation · Misconceptions · Fiction · Confidence judgments · Reading

There has been surging public, political, journalistic, and intellectual interest in the spread of inaccurate and misleading information (Lazer et al., 2018; Lewandowsky et al., 2017; Pennycook & Rand, 2021; Rapp & Salovich, 2018). In response, researchers spanning the social sciences have contributed to understandings of whether and how people respond to misinformation and disinformation (Rapp et al., 2020). This is clearly evidenced by the continually growing body of published work on the topic. Consider that a Google Scholar search for publications referencing the term "fake news" between 1900 and 2016 obtained approximately 5,000 hits, but since 2017 retrieves over 50,000 relevant publications (with an astonishing two-thirds of that number published since 2020). The results of these projects have informed theoretical understandings of people's experiences with, and the consequences of, exposures to inaccuracies. Studies on the belief and spread of inaccurate information suggest that information fact-checked as false travels faster and wider online than information fact-checked as true (Vosoughi et al., 2018), and that people are often unaware of their susceptibility to such false ideas (Lyons et al., 2021; Salovich & Rapp, 2021). People are not just persuaded by inaccuracies that are plausible or unfamiliar to them, but also when they should "know better," such as when they have the prior knowledge or resources available to determine that a given claim is false (Brashier et al., 2020; Donovan & Rapp, 2020; Fazio et al., 2019; Fazio, Brashier, et al., 2015a; Salovich et al., 2021; Salovich & Rapp, 2021).

Although researchers have leveraged a variety of methodologies and stimuli to interrogate the influence of false information, one recurring method of examining these effects

✉ Nikita A. Salovich
salovich@u.northwestern.edu

[1] Department of Psychology, Northwestern University, 2029 Sheridan Drive, Evanston, IL 60208, USA

[2] Department of Psychology, University of California, Los Angeles, 502 Portola Plaza, Los Angeles, CA 90095, USA

[3] School of Education and Social Policy, Northwestern University, 2120 Campus Drive, Evanston, IL 60208, USA

Published online: 05 July 2022

has involved embedding inaccurate statements in narrative materials including fiction (e.g., Gerrig & Prentice, 1991; Prentice et al., 1997; Marsh et al., 2003; Rapp, 2016). Inaccurate ideas routinely appear in popular fiction, as authors are not required to convey the truth in their written products, making these contexts viable and authentic as experimental materials. The inclusion of inaccurate ideas in fiction can be supported by or orthogonal to accompanying descriptions or dialogue, to allow for experimentally varying features associated with the presentation of misleading claims (Gerrig, 1993; Rapp, 2008; Salovich & Rapp, 2021). For these reasons, fictional narratives represent an externally valid set of materials that can be strategically deployed to answer both theoretical and applied questions about exposures to inaccuracies.

One well-replicated method using such materials asks participants to read stories that include a mixture of accurate and inaccurate ideas. For example, participants might read about characters discussing the false claim that tooth brushing leads to (rather than prevents) gum disease (Gerrig & Prentice, 1991), or that Oslo is the capital of Finland (when the capital is actually Helsinki; Marsh et al., 2003). After reading, participants are asked to judge the validity of those ideas or to answer questions related to the ideas. Findings have consistently demonstrated that reading false information in fiction leads to problematic consequences with respect to what people believe or at least report to be true. Participants are more likely to misjudge false ideas as true after having read inaccurate as compared to accurate assertions in stories (Appel & Richter, 2007; Donovan et al., 2018; Salovich & Rapp, 2021). They also reproduce story inaccuracies to answer related questions (e.g., answering "Oslo" to "What is the capital of Finland?") more so than they spontaneously produce those inaccuracies after having read accurate information or reading unrelated and unspecified information (Donovan & Rapp, 2020; Hinze et al., 2014; Marsh et al., 2003; Salovich et al., 2022). Even when participants successfully provide correct responses following exposures to inaccuracies, they can take longer to answer the questions (Gerrig & Prentice, 1991) and may be less sure of their answers (Appel & Richter, 2007; Salovich et al., 2021).

Although empirical projects on the effects of inaccuracies in stories have historically employed in-person data collection, recent advances and preferences have motivated online participant recruitment and experimentation techniques. Online platforms (e.g., Amazon Mechanical Turk) can provide greater diversity relative to undergraduate-only samples (Ipeirotis, 2010), and have become popular among researchers during the COVID-19 pandemic (Arechar & Rand, 2021). Unfortunately, online samples do not always produce responses comparable in quality to in-person samples, especially for studies that can take substantial time to complete (Goodman et al., 2013). This represents a critical concern for researchers interested in studying story-embedded inaccuracies as the materials are often quite lengthy (e.g., a 19-page single-page text as in Gerrig & Prentice, 1991; nine separate stories each 5 to 7 pages long as in Marsh, 2004). Despite such length, texts cannot include a large number of inaccurate statements, claims, or assertions for fear that participants may catch on to the purposes of the experiment or develop specific reading strategies. This is accompanied by considerations as to participants' waning attention when reading long segments of text, and the need for regular comprehension checks to ensure processing of the text content (which the aforementioned story stimuli do not always include).

In consideration of these concerns, and the increasing use of online samples and available platforms for data collection (e.g., Prolific, Amazon Mechanical Turk, CloudResearch, Prolific, SurveyMonkey, CrowdFlower), we have developed and validated a 39-item set of short texts, referred to here as vignettes, on topics related to real-world processes and phenomena. Each vignette describes a conversation that could occur in an everyday setting (e.g., talking to a neighbor), with parallel versions including either an accurate (e.g., brushing your teeth improves gum health) or inaccurate assertion (e.g., brushing your teeth can lead to gum disease) made by one of the characters. The use of these vignettes offers several advantages for researchers collecting data from online samples, and we highlight five: They are brief, simple, flexible, allow for participant accountability, and are realistic. We describe each feature in turn next.

First, the vignettes are *brief*, with each story 125–150 words long. Short texts are ideal for online participants, who may be sensitive to time-cost and compensation considerations and may be more easily recruited for short studies (Berinsky et al., 2012; Buhrmester et al., 2011). Online participants have also been shown to click away from longer texts at concerningly high rates (Imundo & Rapp, 2021) and may spend less time examining experimental materials than do in-person participants (Gibson et al., 2021). These vignettes may therefore be time- and cost-effective relative to longer texts.

Second, the vignettes are *simple* and easy-to-read. Each story falls within a grade-school level of reading difficulty as measured with Flesch–Kincaid scores. This is especially important given moves from in-person samples (which in psychological research often recruit from college undergraduate populations) to online recruitment. Online samples can include participants with more diverse educational backgrounds, though the average online survey worker still tends to be more highly educated than is the average U.S. adult (Huff & Tingley, 2015; Ogletree & Katz, 2021; Ross et al., 2010). Education demographics can be more varied if studies specifically recruit a U.S. representative sample, which is a convenient setting option on many online data

collection platforms. According to recent U.S. census data, 39.3% of participants in a representative sample would have attained no more than a high school diploma/GED (United States Census Bureau, 2017). Approximately 20% of U.S. adults' reading skills are limited to identifying a piece of information within a brief, simple text which uses only basic vocabulary (U.S. Department of Education, 2017), and only 20–30% of U.S. adults are scientifically literate, defined as being able to read and understand the Science section of *The New York Times* (Miller, 2004; U Mich News, 2011). Therefore, grade-school level materials are preferable to ensure comprehension of experimental texts by more diverse and representative educational samples of online participants.

Third, the independent nature of each vignette also allows them to be *flexibly* used by researchers depending on the aims of their study. Unlike longer narratives that may involve complex, interconnected storylines, researchers can selectively present a subset of vignettes to participants based on the constraints of their study. Editing or excluding portions of larger narrative texts often requires expending considerable time and effort towards rewrites to maintain the comprehensibility of the overall story. The more modest lengths of these vignettes allow researchers to vary the number of vignettes participants read, the type of vignette (i.e., whether it includes accurate or inaccurate information), the topic of the vignette (e.g., whether the referenced assertion is health-related), and the degree to which information provided in the vignette aligns with or contradicts people's prior knowledge.

Fourth, the materials can hold participants *accountable* by including story-specific true-or-false comprehension checks to monitor attention and comprehension. While related to the story events, these checks have been designed to be unrelated to the assertion of interest (e.g., True or False? Rachel and Tony are on their way to the airport) to avoid drawing attention to the relevant assertion or enhancing memory by encouraging additional retrieval of the assertion (e.g., Bjork, 1975; Roediger & Karpicke, 2006). Comprehension and attention checks are considered a feature of good experimental studies and are especially crucial for online samples (Aguinis et al., 2021; Mellis & Bickel, 2020). Online participants who cannot be directly observed may be prone to skimming experimental materials or randomly responding rather than closely reading in order to "speed" through surveys and receive the most compensation for the least time and effort (Smith et al., 2016). Although Amazon MTurk workers' attention can equal or even outperform that of undergraduate participants (Buhrmester et al., 2018), there is emerging evidence that online participants may fail attention checks at much higher rates than do in-person participants (Saravanos et al., 2021). This can be exacerbated as new users join online survey platforms, as happened during the COVID-19 pandemic (Arechar & Rand, 2021). Consequently, researchers could use vignette-specific comprehension checks to ensure participants are paying attention and can comprehend them. This allows for excluding participants who answer less than a prespecified percentage of comprehension checks correctly, or for excluding data from vignettes in which participants did not successfully answer the comprehension checks. Comprehension and attention checks can be administered in shorter stories without interrupting the described plot or events, which is an additional, recurring challenge when using longer texts.

Lastly, the texts contain information that is *realistic* with respect to the type of true and false ideas that people can encounter during their day-to-day experiences. The critical content is focused exclusively on assertions pertaining to real-world processes and phenomena. Unlike true and false declarative statements or "facts" used extensively in prior work (e.g., Marsh, 2004; Marsh et al., 2003; Marsh & Fazio, 2006), accurate and inaccurate assertions are uniquely characterized by a *preponderance of evidence* suggesting the idea is true or false. For example, whereas the statement "Dried grapes are called prunes" is factually incorrect, the assertion "Toothbrushing causes gum disease" is inaccurate based on the accumulation of research supporting that toothbrushing is beneficial for gum health, even if exceptions or counterexamples could be identified (e.g., that brushing your teeth too hard and too often can irritate your gums). As a result, assertions help represent complexity and nuance in the types of misleading claims and ideas people are regularly exposed to (e.g., popular press headlines, product advertisements, inauthentic reviews; Pennycook & Rand, 2021; Rapp et al., 2020), as well as topics that are of contemporary public concern (e.g., preventative health behaviors; Carrieri et al., 2019; Loomba et al., 2021; Pennycook et al., 2021b). Norming studies for declarative general knowledge statements are frequently updated in the research community (e.g., Coane & Umanath, 2021; Jalbert et al., 2019; Nelson & Narens, 1980; Tauber et al., 2013), while norming data for assertions are less frequently obtained, with some referenced materials dating back several decades (e.g., Gerrig & Prentice, 1991). With these considerations in mind, the current materials offer updated, normed, and validated accurate and inaccurate assertions for contemporary investigations.

Norming, validation, and replication for the newly designed vignettes

We report norming and validation of the vignette materials across both U.S.-census matched and a general online sample. In the norming studies, participants read a series of accurate assertions (e.g., "You can only catch warts from humans," "Getting the flu shot cannot give you the flu") or inaccurate assertions ("You can catch warts from toads,"

"Getting the flu shot can give you the flu"), and judged whether they were true or false based on what they knew or believed to be true about the world. Based on those baseline data, we conducted a validation experiment with the assertions appearing in vignettes, following procedures from previous studies examining the effects of exposure to true and false information (e.g., Gerrig & Prentice, 1991; Rapp et al., 2014; Salovich et al., 2021; Salovich & Rapp, 2021). Specifically, participants read multiple vignettes containing accurate and inaccurate assertions, and then judged the validity of a series of assertions, some of which related to information presented in the previously read vignettes. Following the successful validation, we tested whether the vignette materials could be used to replicate recent findings. Specifically, we attempted to replicate the results of Salovich et al. (2021), wherein participants reported lower confidence in correct judgments and higher confidence in incorrect judgments after having read inaccurate story content. The reported results across these norming, validation, and replication studies underscore the utility of these materials for future research.

Part 1: Norming

The purpose of norming studies 1 and 2 was to identify people's baseline beliefs about a set of real-world assertions. We were interested in which assertions participants were most and least likely to correctly identify as accurate or inaccurate based on their prior knowledge. This provides insight into which ideas target samples would identify as true prior to exposures to any information. Rather than specifically selecting for "easy" (well known) and "hard" (unknown) items (e.g., Marsh, 2004), we aimed to select assertions that varied in how much participants knew and believed about the topics. This would support our goal of creating materials that capture the variety of claims and ideas people can commonly encounter. In the norming studies, participants read a series of accurate or inaccurate assertions one-at-a-time and judged each assertion as to whether it was true or false based on what they knew to be true about the world. We calculated the proportion of people who correctly and incorrectly judged the accuracy of each assertion. Previous work has demonstrated that prior knowledge can affect the degree to which people are influenced by inaccurate information (e.g., Donovan & Rapp, 2020; Fazio, Brashier, et al., 2015a; Marsh et al., 2003; Salovich et al., 2022; Unkelbach & Speckmann, 2021). Therefore, knowing which assertions are more or less likely to be identified by participants as accurate is useful for any future work attempting to investigate the consequences of exposure to accurate and inaccurate assertions (e.g., Rapp, 2016). norming study 1 examined these baseline judgments with a representative U.S. sample, and norming study 2 used a general online sample.

Norming study 1: Assertion norming with a representative U.S. sample

Method

This study and all subsequent studies and experiments (STU00211662) received exempt approval by Northwestern University's IRB and were run using Qualtrics (https://www.qualtrics.com/).

Participants A representative sample of American adults, matched on the most recent (2019) U.S. census on age, gender, and race, was recruited using Prime Panels via CloudResearch (https://cloudresearch.com/). Based on the census-matched quota, participants were paid a pre-determined amount established by the study platform for their participation in the study, which was estimated to take between 10 and 15 minutes to complete. The platform continued to recruit participants until at least 100 participants successfully qualified and completed two comprehension checks[1], none of whom reported having looked up answers at any point during the study, with the sample matched to the target census demographic breakdown. This resulted in 122 participants in the final sample.

Materials Thirty-nine assertions (e.g., benefits of aerobic exercise; see Table 1 for full list) were developed for this study. These assertions were selected based on assertions used in prior studies of reliance on inaccurate information and from searching the internet for possible candidates.

Procedure After consenting to participate in the study, participants judged each assertion as true or false, presented one-at-a-time in a unique random order for every participant. For each participant, half of the assertions were false (the inaccurate version; e.g., aerobic exercise *weakens* your heart and lungs) and half were true (the accurate version; e.g., aerobic exercise *strengthens* your heart and lungs). After completing their judgments, participants were debriefed. Demographics were collected and verified separately by the study platform.

[1] Comprehension checks included two multiple-choice questions. The first asked people to select a word that described something that could be used during cooking, and the second asked people to select the answer of a simple arithmetic word problem.

Table 1 Assertion and story norms

Assertion		Word count	Flesch–Kincaid grade level		Experiment 3 confidence norms [c]			Correct prior knowledge norms [a,b]								
								General sample					Representative sample			
Accurate form	Inaccurate form		Accurate version	Inaccurate version	*n*	*M*	*SD*	*SE*	*n*	*M*	*SD*	*SE*	*n*	*M*	*SD*	*SE*
A penny dropped from the Empire State Building would cause minimal harm to someone.	A penny dropped from the Empire State Building could kill someone.	150	5.0	4.5	44	3.86	1.27	0.19	53	.48	.50	.07	133	.44	.50	.04
A person's likelihood of developing depression is influenced by many genetic and environmental factors.	A person's likelihood of developing depression is controlled by a single gene.	145	6.6	6.2	46	4.09	1.05	0.15	53	.84	.37	.05	133	.77	.42	.04
Aerobic exercise strengthens your heart and lungs.	Aerobic exercise weakens your heart and lungs.	136	6.3	6.3	44	4.55	0.98	0.15	53	.89	.31	.04	133	.91	.29	.02
Brushing your teeth improves gum health.	Brushing your teeth can lead to gum disease.	136	5.3	4.6	43	4.40	1.00	0.15	53	.91	.29	.04	133	.94	.24	.02
College graduates often earn more than those without a college degree.	College graduates often earn less than those without a college degree.	142	7.8	7.0	46	4.30	0.87	0.13	53	.79	.41	.06	133	.67	.47	.04
Cracking your knuckles does not play a role in developing arthritis.	Cracking your knuckles causes arthritis.	128	6.4	5.6	46	3.89	1.10	0.16	53	.77	.42	.06	133	.58	.49	.04
Drinking alcohol decreases your core body temperature.	Drinking alcohol increases your core body temperature.	125	5.2	5.0	44	3.36	1.22	0.18	53	.45	.50	.07	133	.46	.50	.04
Eating within an hour of swimming does not affect your chances of experiencing muscle cramps.	Eating within an hour of swimming increases your chances of experiencing muscle cramps.	137	5.6	5.6	43	3.84	1.17	0.18	53	.63	.48	.07	133	.46	.50	.04
Epilepsy is not contagious.	Epilepsy is contagious.	130	7.9	8.5	46	4.70	0.84	0.12	53	.88	.32	.04	133	.86	.35	.03
Fingernails grow back at their usual rate after they have been cut.	Fingernails grow back faster after they have been cut.	150	3.4	3.4	46	4.00	1.19	0.16	53	.82	.38	.05	133	.63	.48	.04
Getting the flu shot cannot give you the flu.	Getting the flu shot can give you the flu.	145	5.2	3.8	43	4.12	1.12	0.17	53	.55	.50	.07	133	.56	.50	.04
It is easier to float in salt water.	It is harder to float in salt water.	145	5.4	5.4	46	3.96	1.30	0.19	53	.75	.43	.06	133	.69	.46	.04
Lightning can strike the same place twice.	Lightning never strikes the same place twice.	125	4.8	4.5	43	4.02	0.96	0.15	53	.82	.38	.05	133	.80	.40	.03
Microwave radiation does not cause cancer.	Microwave radiation causes cancer.	137	6.7	6.4	43	3.86	0.94	0.14	53	.57	.50	.07	133	.38	.49	.04
Muscle soreness is caused by microscopic muscle tears.	Muscle soreness is caused by lactic acid buildup.	130	7.8	6.5	43	3.74	1.05	0.16	53	.48	.50	.07	133	.47	.50	.04
No emotional changes are associated with moon phases.	A full moon can cause an increase in anxiety.	144	5.5	6.1	44	3.43	1.44	0.22	53	.54	.50	.07	133	.44	.50	.04
People with dark skin can get sunburned.	People with dark skin cannot get sunburned.	147	5.2	5.7	44	4.52	0.79	0.12	53	.77	.42	.06	133	.83	.38	.03

Table 1 (continued)

Assertion		Word count	Flesch–Kincaid grade level		Experiment 3 confidence norms [c]		Correct prior knowledge norms [a,b]								
							General sample					Representative sample			
Plants take carbon dioxide out of the atmosphere.	Plants do not take carbon dioxide out of the atmosphere.	148	5.7	7.2	43	4.26 0.95	0.15	53	.70	.46	.06	133	.76	.43	.04
Playing on computers or smartphones can keep you up at night.	Playing on computers or smartphones late at night can help you fall asleep.	137	5.3	6.1	44	4.50 1.02	0.15	53	.88	.32	.04	133	.80	.40	.03
Seasons are caused by the tilt of the Earth.	Seasons are caused by the distance of the Earth to the Sun.	140	5.3	5.7	44	4.09 0.98	0.15	53	.52	.50	.07	133	.50	.50	.04
Sleeping will help you consolidate knowledge and may actually help you to remember things.	Sleeping won't help you consolidate knowledge and may actually cause you to forget things.	127	5.3	4.6	43	3.91 1.27	0.19	53	.82	.38	.05	133	.74	.44	.04
Smokers become addicted to cigarettes from nicotine.	Smokers become addicted to cigarettes from smoking habits.	133	7.4	7.4	44	4.39 0.84	0.13	53	.64	.48	.07	133	.56	.50	.04
Some of a person's body heat is lost through the head.	Most of a person's body heat is lost through the head.	144	6.0	4.1	43	4.09 1.11	0.17	53	.59	.49	.07	133	.59	.49	.04
Students do not have an optimal learning style.	Students have an optimal learning style.	143	7.8	9.7	46	3.63 1.08	0.16	53	.46	.50	.07	133	.29	.45	.04
Sugar does not cause hyperactivity in children.	Sugar causes hyperactivity in children.	142	8.3	8.9	46	3.74 1.29	0.19	53	.32	.47	.06	133	.29	.45	.04
Taking a foreign language broadens your mind.	Taking a foreign language is a waste of time.	128	4.4	4.5	43	4.14 0.99	0.15	53	.91	.29	.04	133	.87	.34	.03
Talking to plants does not affect their health.	Talking to plants makes them healthier.	128	5.5	3.7	46	3.72 1.20	0.18	53	.43	.50	.07	133	.41	.49	.04
The Amazon holds the majority of the Earth's biodiversity.	The Amazon does not hold the majority of the Earth's biodiversity.	143	8.3	7.1	44	3.14 1.32	0.20	53	.75	.43	.06	133	.61	.49	.04
The moon does not produce its own glow.	The moon produces its own glow.	142	3.8	5.5	44	4.32 1.14	0.17	53	.75	.43	.06	133	.61	.49	.04
The moon's phases are caused by its position relative to Earth and the sun.	The moon's phases are caused by clouds.	149	3.8	3.3	43	3.95 1.29	0.20	53	.80	.40	.05	133	.88	.32	.03
The T-Rex mostly ate meat.	The T-Rex mostly ate plants.	140	4.3	4.8	43	3.81 1.28	0.19	53	.66	.47	.07	133	.68	.47	.04
There is no evidence linking deodorant to breast cancer.	Deodorant causes breast cancer.	129	8.1	6.3	44	3.89 1.22	0.18	53	.64	.48	.07	133	.70	.46	.04
Tides are controlled by the moon's gravitational pull.	Tides are unrelated to the moon's gravitational pull.	139	7.1	6.2	44	4.41 0.90	0.14	53	.82	.38	.05	133	.78	.41	.04
Twins cannot communicate telepathically.	Twins can communicate telepathically.	140	6.2	6.7	46	4.35 0.90	0.13	53	.75	.43	.06	133	.50	.50	.04
Vitamin C is an ineffective treatment for a cold.	Vitamin C is an effective treatment for a cold.	125	4.3	5.4	46	3.67 1.16	0.17	53	.41	.49	.07	133	.26	.44	.04
Waking sleepwalkers may help keep them safe.	Waking sleepwalkers disorients and harms them.	147	4.2	5.3	46	3.48 1.26	0.19	53	.46	.50	.07	133	.35	.48	.04

Table 1 (continued)

Assertion		Word count	Flesch–Kincaid grade level		Experiment 3 confidence norms [c]				Correct prior knowledge norms [a,b] General sample				Representative sample			
Wearing a seatbelt can increase your chances of living through an accident.	Wearing a seatbelt can reduce your chances of living through an accident.	146	5.2	3.9	44	4.77	0.48	0.07	53	.82	.38	.05	133	.82	.38	.03
You can only catch warts from humans.	You can catch warts from toads.	134	2.3	2.9	46	3.67	1.30	0.19	53	.50	.50	.07	133	.48	.50	.04

[a]Correct prior knowledge norms refer to the proportion of the sample in the assertion norming studies, who made their validity judgments without reading the vignettes, which responded correctly to the given item. For example, responding "true" to an accurate statement or responding "false" to an inaccurate statement.

[b]Descriptive norms reported here are pooled across accurate and inaccurate versions of each assertion. An expanded Table 1 reporting separate norms for accurate and inaccurate versions of each assertion can be found on OSF (osf.io/yzuq8/).

[c]Experiment 32 confidence norms refers to participants' average confidence in their response on the validity judgment task for control items (i.e., those items for which participants did not read one of the corresponding vignettes).

Results

These norming data appear in Table 1 with accuracy distributions presented in Fig. 1. Participants correctly identified the accurate versions of the assertions as accurate (e.g., brushing your teeth improves gum health) more often ($M = .67$, $SD = .42$) than they correctly identified the inaccurate versions of the assertions as inaccurate (e.g., brushing your teeth can lead to gum diseases; $M = .55$, $SD = .44$). Overall, participants were more likely to rate accurate assertions as true than they were to rate inaccurate assertions as false. This suggests that respondents were slightly biased to agree versus disagree with the presented assertions. Participants' accuracy in their validity judgments also varied considerably across the assertions. The sample correctly responded to the assertions 61.0% ($SD = 44.6\%$) of the time, with the most correctly responded to topic indicating an association between brushing your teeth and gum health ($M = .94$, $SD = .24$), and the least correctly responded to topic discussing vitamin C as an (in)effective treatment for a cold ($M = .26$, $SD = .44$).

Norming study 2: Assertion norming with a general online sample

Method

Participants A general sample ($N = 56$) was recruited from Amazon Mechanical Turk via CloudResearch. Participants were on average 39.58 years old; 46.43% were male, 50% were female, and 3.57% identified as third gender or non-binary. When asked to select any race that applied, 83.93% identified as white/Caucasian, 3.57% as Black/African American, 7.41% as Asian/Pacific Islander, and 5.36% as Latino. Participants received $1.50 for their participation in the study and the average completion time was 6.65 min. Three participants were removed for failing one of the comprehension checks presented at the beginning of the study or for reporting they looked up answers during the study, leaving 53 participants in the final sample.

Materials and procedure The same materials and procedure as in norming study 1 were used here.

Results

These norming data also appear in Table 1 with accuracy distributions presented in Fig. 1. Replicating norming study 1, participants correctly identified the accurate versions of the assertions as true more often ($M = .70$, $SD = .40$) than they identified the inaccurate versions of the assertions as false ($M = .63$, $SD = .43$). Again, participants were more likely to agree than disagree with the accuracy of presented

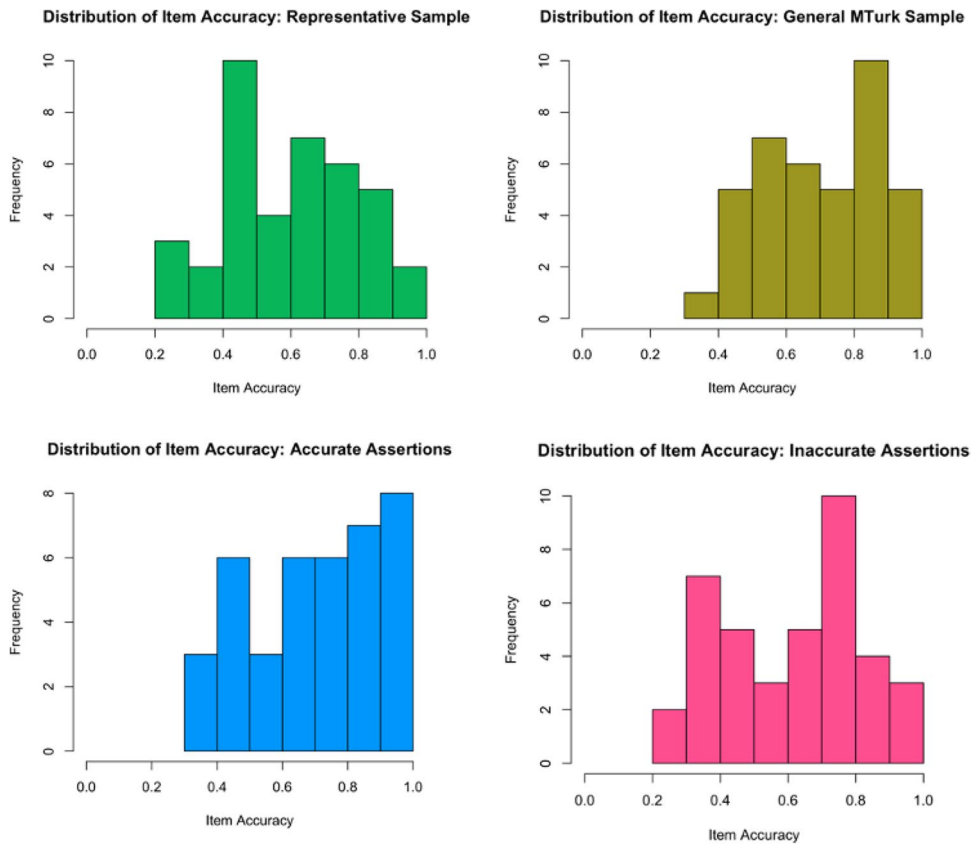

Fig. 1 Distribution of item accuracy within representative and general MTurk samples, and across accurate and inaccurate assertions. Distributions within representative (Experiment 16; *top left quadrant*) and general MTurk (Experiment 24; *top right quadrant*) samples use accuracy calculations collapsed across accurate and inaccurate assertions. Distributions across accurate and inaccurate assertions use accuracy calculations collapsed across representative and general MTurk samples

assertions. Judgments again reflected variation in prior knowledge. On average, the general MTurk sample correctly endorsed assertions 66.8% (SD = 43.6%) of the time. The most correctly responded to topics were brushing your teeth and gum health (M = .91, SD = .29), and the impact of taking a foreign language on your mind (M = .91, SD = .29); the least correctly responded to topic was the relationship between sugar intake and child hyperactivity (M = .32, SD = .47).

Discussion

These results provided norming data for the selected general assertions. Participants were better able to identify accurate versions of the assertions as true than they were to identify inaccurate versions as false, which may in part be due to a general tendency to agree with the presented information. The data also suggest that the different assertions varied considerably in terms of participants' prior knowledge, ranging across items from around one-quarter to almost all participants in the samples holding correct knowledge. While the samples were roughly split on some assertion topics (e.g., seasons are caused by the tilt of the Earth), the norming data for many of the assertion topics suggest that participants held systematic prior knowledge that was either aligned with the preponderance of evidence or potentially contradicted that evidence (see Table 1). For example, for some assertions (e.g., vitamin C is an effective treatment for a cold), participants' responses suggest they may, on average, actually hold misconceptions, as they systematically rated the accurate version as false and inaccurate version as true.

Examination of the results of the representative and general U.S. samples indicated similar patterns of judgments, and thus suggests similar prior knowledge. Though there was some minor variation in the accuracy of the validity judgments between the two samples, assertion topics that were correctly endorsed by participants in the representative sample were also correctly endorsed by participants in the

general sample, and topics that were incorrectly endorsed in the representative sample were judged similarly in the general sample, $r = .90$, 95% CI [.82, .95], $p < .001$. That the general sample overall had greater prior knowledge than the representative sample may reflect the tendency for general MTurk samples to have had more educational experiences as compared to representative U.S. samples (Huff & Tingley, 2015). That said, the goal was not to compare the two samples, but rather to report both samples for comprehensiveness, as the norming data could be used to inform future research conducted with these materials involving either population.

In sum, norming studies 1 and 2 provided the baseline rates at which participants in both representative and general online samples held correct or incorrect beliefs about various assertions. As intended, participants' validity judgments demonstrated that there was substantial variability in the background knowledge they possessed about the selected assertions: These samples generally knew some assertions to be true or false, less consistently knew the validity of other assertions, and even held misconceptions about some of them. In Experiments 16 and 24, we validated the vignette materials that included these assertions with both U.S. census-matched and general online samples.

Part 2: Materials validation

Experiments 16 (with a representative U.S. sample) and 24 (with a general U.S. sample) investigated whether participants would exhibit an influence of having read the assertions when included in fictional narratives. Brief vignettes, each centered on one assertion, were developed and presented to participants. We anticipated, in line with previous work, that participants would be influenced by the accurate and inaccurate information presented in the texts, as measured by the accuracy of their post-reading validity judgments.

Experiment 1: Story validation with a representative U.S. sample

Method

Materials Thirty-nine vignettes, each containing one of the normed assertions from norming studies 1 and 2, were created for this study. Each vignette began with an opening statement (e.g., Elizabeth was getting her mail when she saw Sarah returning home) followed by a brief excerpt of conversation between two characters about the story topic. There were two versions of each vignette as a function of whether the assertion included in the vignette was accurate or inaccurate. The vignettes were 125–150 words long, ranging in Flesch–Kincaid grade level (https://readabilityformulas.com/flesch-grade-level-readability-formula.php) from 2.3–9.7 ($M = 5.69$), suggesting they should be easy-to-understand for the typical participant aged 16+. For each vignette there was an accompanying comprehension check question unrelated to the assertion (e.g., True or False? Elizabeth was getting her mail). Nineteen of the correct answers to the comprehension check questions were true, and 20 of the correct answers were false.

Design Experiment 16 used a 3 (assertion accuracy: true, false, or control/not included) x 2 (test statement accuracy: true or false) within-participants design.

Participants A representative U.S. sample matched on the latest (2019) U.S. census on age, gender, and race, was recruited using Prime Panels via CloudResearch. Based on the census-matched quota, participants were paid a predetermined amount established by the study platform for their participation in the study, which was estimated to take between 20 and 25 min to complete. As in norming study 1, the platform continued to recruit participants until at least 50 participants successfully qualified and completed two comprehension checks, and the sample matched the target census demographics. Participants who scored less than 70% correct on the comprehension questions or reported looking up answers at any point during the study were excluded from analyses, resulting in 62 participants in the final sample.

Procedure Assertion accuracy and test statement accuracy were counterbalanced across six possible conditions. Participants read 26 of the 39 vignettes in a random order (the other 13 unread vignettes served as the control topics). After reading each vignette, participants answered a comprehension question. Participants then completed what they were told was a general knowledge assessment, and were asked to answer based on their prior knowledge. For this task, participants judged 39 statements as either true or false, one-at-a-time, presented in a random order to each participant. Twenty-six of the statements were related to the assertions appearing in the short stories, and 13 were new statements. Each statement was presented in either its accurate form (e.g., aerobic exercise weakens your heart and lungs) or inaccurate form (e.g., aerobic exercise strengthens your heart and lungs).

Analysis All analyses were conducted using generalized linear mixed-effects models (GLMM) with the R packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017), with assertion accuracy as a fixed effect (control set as referent condition), and participants and vignette as random intercepts. This mixed-effect analysis simultaneously accounts for variance due to random selection of

participants and random selection of items (Richter, 2006). By accounting for these random effects, we can ascertain the effect of assertion validity on subsequent validity judgments. Doing so also allows us to better generalize to other instances beyond the sampled subjects and items included in this analysis (e.g., other subjects or other stimuli; Baayen et al., 2008). Models were fit with a binomial distribution, given that each response was binary (errors were coded as 1). All data and R scripts are publicly available on OSF (https://osf.io/5bvyx/).

Results

Descriptive statistics for error rates can be found in Table 2. Participants produced more judgment errors after reading inaccurate assertions ($M = .50$, $SD = .20$), $b = 0.78$, $z = 7.15$, $p < .001$, and fewer judgment errors after reading accurate assertions ($M = .29$, $SD = .19$), $b = -0.32$, $z = -2.84$, $p = .005$, as compared to after reading stories that did not reference the assertions (control; $M = .34$, $SD = .15$). See Fig. 2 for illustrated effects.

Experiment 2: Story validation with a general online sample

Method

Design, materials, and procedure Experiment 24 used the same design, materials, and procedure as Experiment 16.

Participants A general sample ($N = 50$) was recruited from Amazon Mechanical Turk. Participants were on average 39.58 years old; 64.00% were male, 34.00% were female, and 2.00% identified as third gender or non-binary. When asked to select any race that applied, 66.00% identified as White/Caucasian, 4.00% as Black/African American, 2.00% as Asian/Pacific Islander, 8.00% as Latino, 4.00% as Native American, 2.00% as Other, and 2.00% preferred not to say. Participants received $3 and took on average 24.50 min to complete the study. Participants who scored less than 70% correct on the comprehension questions or reported looking up answers were excluded from analyses, resulting in a final sample of 42 participants.

Analysis Results were analyzed using the same model specifications as in Experiment 16.

Results

The results of Experiment 24 replicated with a general online sample (see Table 2). Participants produced more judgment errors after reading inaccurate assertions ($M = .38$, $SD = .22$), $b = 0.51$, $z = 3.48$, $p < .001$, and fewer errors after reading accurate assertions ($M = .23$, $SD = .17$), $b = -0.44$, $z = -2.78$, $p = .005$, as compared to after not reading any information related to the assertions (control; $M = .29$, $SD = .14$). See Fig. 2 for illustrated effects.

Table 2 Mean error rates in Experiments 16, 24, and 32

Study	*M*	*SD*	*SE*
Experiment 16			
Accurate	.29	.19	.02
Inaccurate	.50	.20	.02
Control	.34	.15	.02
Experiment 24			
Accurate	.23	.17	.03
Inaccurate	.38	.22	.03
Control	.29	.14	.02
Experiment 32			
Accurate	.36	.18	.02
Inaccurate	.44	.17	.01
Control	.38	.18	.02

Discussion

In line with our expectations, the results of Experiments 16 and 9 indicated that participants' judgments were influenced by the assertions in the vignettes. Participants were more likely to make correct judgments after reading accurate information and were more likely to make incorrect judgments after reading inaccurate information. These results align with prior work to demonstrate that people rely on the falsehoods they have read even when they should know better (e.g., Rapp, 2016). The findings are also consistent with the view that such exposures may lead people to rely less on their own knowledge (Rapp & Salovich, 2018; Salovich et al., 2022), instead offering responses informed by the assertions recently encountered in the narrative texts.

Part 3: Replication and extension with confidence ratings

Experiment 3: Replication of Salovich et al. (2021)

The purpose of Experiment 32 was two-fold. First, we aimed to replicate the key effects obtained in Experiments 16 and 24, wherein participants made more judgment errors after reading vignettes containing inaccurate assertions, as compared to having read vignettes containing accurate assertions or no related information. Second, we aimed to further validate the materials by replicating the findings of a recent study which investigated whether exposure to accurate and inaccurate information in a narrative influenced participants'

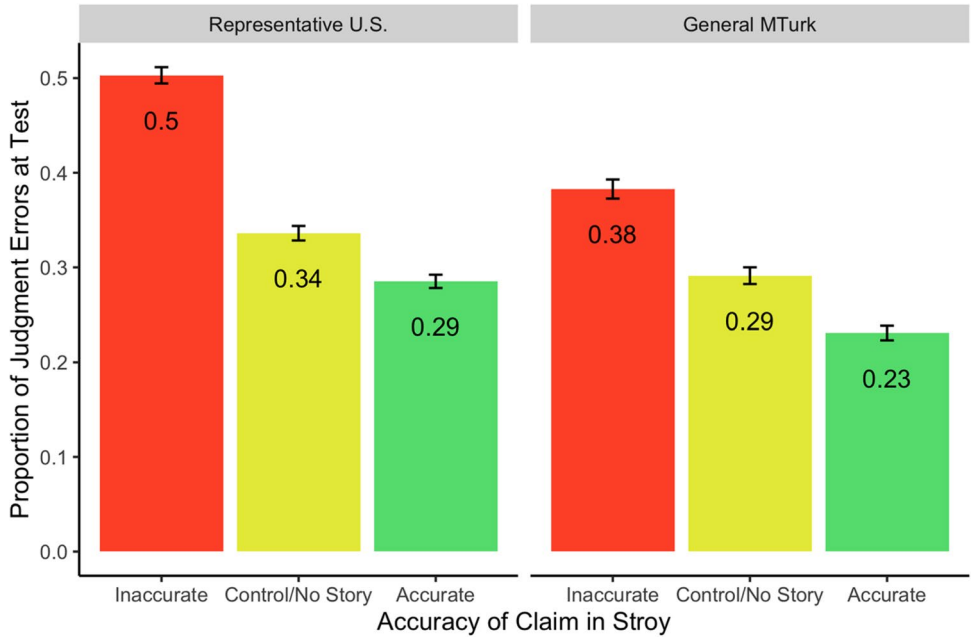

Fig. 2 Proportion of judgment errors after reading inaccurate, control/no story, and accurate stories. Responses from the census-matched, representative U.S. sample are depicted on the left (Experiment 16) and from the general MTurk sample are depicted on the right (Experiment 24). *Error bars* represent standard error

confidence in judging the validity of related claims (Salovich et al., 2021 Experiment 16). This issue is critical to the idea that reading inaccurate information may lead people to rely less on their knowledge, potentially by influencing their confidence in what they have read as compared to what they know (Rapp & Salovich, 2018).

In Salovich et al. (2021), participants read a 19-page fictional story entitled "The Kidnapping" (Gerrig & Prentice, 1991), which was designed to examine the consequences of exposure to accurate and inaccurate assertions. The story follows college students interacting over the course of a day and contains 16 critical assertions presented through conversations between characters. Half of the assertions appeared in their accurate form (e.g., "Frequent tooth brushing prevents gum disease") and half in their inaccurate form (e.g., "Frequent tooth brushing leads to gum disease"). After reading the text, participants judged the validity of single statements related to the critical assertions and provided confidence ratings for each judgment. While participants made more judgment errors after having read inaccurate assertions than after having read accurate assertions or stories without assertions, they were overall less confident in their incorrect as compared to correct judgments. However, following exposures to inaccurate story content, this confidence-response accuracy relationship was attenuated. In the current experiment, we aimed to provide a conceptual replication of these effects using the current materials, assessing whether vignettes containing inaccurate assertions could also lead to increased confidence in errors and reduced confidence in accurate knowledge. Evidence suggesting that similar effects emerge with the updated corpus of vignettes would further validate their intended function on a topic of growing interest (i.e., relations between confidence and reliance on inaccurate information).

Method

Participants A general sample ($N = 150$) was recruited from Amazon Mechanical Turk. Participants were on average 39.78 years old; 47.33% were male, 51.33% were female. When asked to select any race that applied, 81.33% identified as white/Caucasian, 8.00% as Black/African American, 9.33% as Asian/Pacific Islander, 0.67% as Latino, 2.00% as Native American, 2.00% as Other, and 0.67% preferred not to say. Participants received $3 and took on average 27.27 minutes to complete the study. As in Experiments 16 and 24, participants who scored less than 70% correct on the comprehension questions or reported looking up answers were excluded from analyses, resulting in a final sample of 133 participants.

Design, materials, and procedure Experiment 32 used the same materials and design as Experiments 16 and 24. The only change in the procedure occurred during the judgment task. After identifying each assertion as true or false, participants were asked "How confident are you in your response?" on a scale from 1 (random guess) to 5 (sure of response), as in Salovich et al. (2021).

Analysis Error rates were analyzed using the same model specifications as Experiments 16 and 24. To test the relationship between response accuracy and confidence across conditions, response accuracy (1 for an error, 0 for a correct response) was added to the model as a fixed effect, along with interaction terms with assertion accuracy. Also, rather than a binomial distribution, confidence ratings were fit by a linear mixed effect model (LMM). We chose these particular model specifications for consistency and comparison with Salovich et al. (2021). Similar to those data, confidence ratings were negatively skewed (skewness = – 1.06). Therefore, we transformed confidence ratings by squaring them, which reduced the skewness of the distribution prior to analysis (skewness = – .59). We also removed responses in which participants indicated that their selection was a random guess (4.16% of responses), as such responses are less likely to represent an intentional accuracy judgment made with low confidence, and rather an arbitrary choice between the two options (Salovich et al., 2021).

Results

Error rates Replicating Experiments 16 and 24, participants were more likely to make judgment errors after reading vignettes containing inaccurate assertions ($M = .44$, $SD = .17$) as compared to not having read related information on the assertion topics (control; $M = .38$, $SD = .18$), $b = 0.27$, $z = 3.56$, $p < .001$. While participants were numerically less likely to produce judgment errors after reading vignettes containing accurate assertions ($M = .36$, $SD = .18$) versus no information, this difference did not reach statistical significance, $b = -0.11$, $z = -1.37$, $p = .17$.

Relationship between confidence and accuracy As predicted, participants' confidence was higher after having read vignettes containing accurate assertions ($M = 4.28$, $SD = .93$) than after not having read relevant assertion information ($M = 4.15$, $SD = .98$), $b = 0.99$, $t = 3.73$, $p < .001$. However, there was no difference in confidence after reading inaccurate assertions ($M = 4.15$, $SD = .97$) and after not reading relevant information, $p = .45$. There was also a main effect of response accuracy, $b = -1.74$, $t = -5.49$, $p < .001$, indicating that confidence was lower for errors than for correct responses. This provides evidence of a significant positive relationship between confidence and response accuracy.

We next considered whether the confidence-response accuracy relationship was affected by the accuracy of information contained in the earlier-read vignettes. The assertion accuracy x response accuracy interaction investigating the consequences of exposure to inaccurate assertions was marginally significant, $b = .78$, $t = 1.80$, $p = .07$. This effect was examined further to determine whether reading inaccurate assertions affected the confidence-response accuracy relationship in the predicted direction. To do so, simple contrasts were calculated using the emmeans R package (Lenth, 2019). As seen in Fig. 3, there was a strong relationship between confidence and accuracy when participants had read accurate assertions in the story or no relevant information, with confidence higher for correct responses than for errors, $b = 1.76$, $z = 5.49$, $p < .001$. This relationship, while still positive, was attenuated after reading inaccurate assertions in the stories, $b = .98$, $z = 3.15$, $p = .002$. Participants were overall more confident in their correct as compared to incorrect judgments but became less confident in correct judgments and more confident in incorrect judgments after having read inaccurate assertions. There was no difference in the confidence–accuracy relationship after reading accurate assertions as compared to not having read related content in the vignettes, $p > .05$.

We next implemented a graphical technique to further illustrate the relationship between confidence and response accuracy, using confidence accuracy characteristic curves (see Mickes, 2015; Salovich et al., 2021) for each vignette-assertion condition (see Fig. 3). This technique simply plots confidence ratings on the *x*-axis against mean accuracy rates on the *y*-axis. If participants' metacognitive judgments are reliable, the data should demonstrate a positive relationship between confidence and response accuracy, with the slope of the line indicating the strength of the relationship. Consistent with the pattern described above, the confidence-response accuracy relationship was stronger after participants had read accurate assertions or no information as compared to after having read inaccurate assertions in the stories.

Discussion

In line with Experiments 16 and 24, participants produced more errors after reading vignettes containing inaccurate assertions as compared to not having been exposed to any related information. Although participants produced numerically fewer errors after reading vignettes containing accurate assertions as compared to not having read related information, unlike in Experiments 16 and 24, this difference did not reach statistical significance. This difference could potentially be driven by the added considerations involved in making confidence judgments at test. That said, it is clear across experiments that people's judgments were influenced after having read inaccurate assertions.

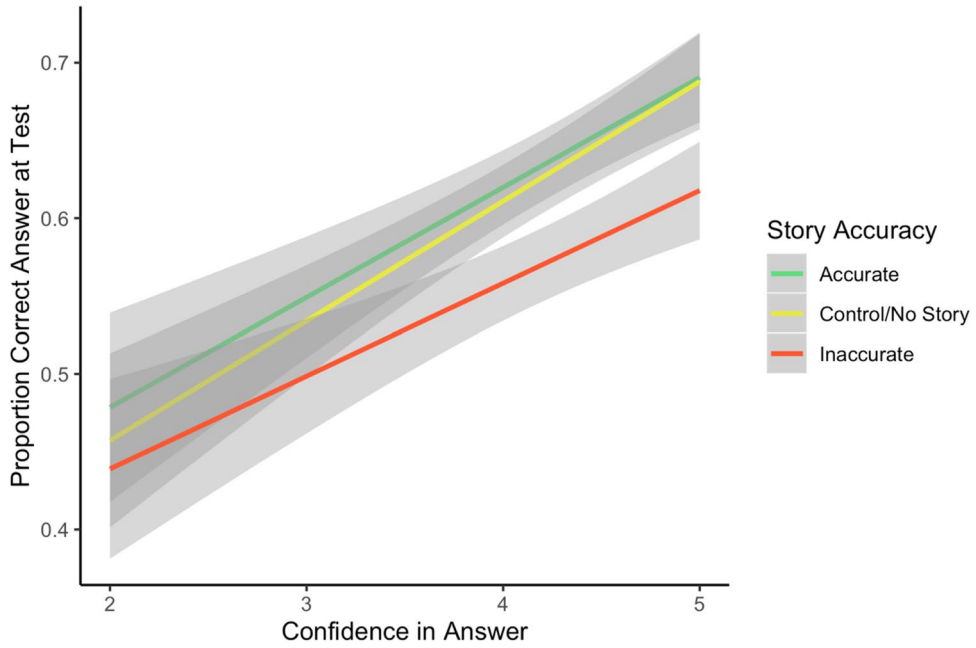

Fig. 3 Confidence–accuracy resolution after reading accurate, control/no story, and inaccurate stories. On the *x*-axis, "2" represents low confidence, and "5" represents sure of answer. Answers that were reported to be random guesses were dropped. *Shaded regions* represent 95% confidence interval

Additionally, Experiment 3 was successful in replicating the key patterns of effects obtained by Salovich et al. (2021) using the vignette materials. First, participants provided higher confidence ratings for correct as compared to incorrect responses after reading inaccurate assertions. The positive correlation between accuracy and confidence, referred to as *confidence–accuracy resolution* (Horry et al., 2014), has been consistently identified in various research projects, including (but not limited to) reading information presented in fiction (e.g., Bulevich & Thomas, 2012; Dehon & Brédart, 2004; Higham et al., 2011; Salovich et al., 2021; Weinstein et al., 2010). Importantly, as also demonstrated in Salovich et al. (2021), this confidence–accuracy relationship was attenuated after reading vignettes containing inaccurate assertions, despite the critical interaction only reaching marginal statistical significance. Exposure to this information problematically led people to become more confident in incorrect judgments and less confident in correct judgments, as compared to after exposure to accurate assertions or not reading related information. These results replicate patterns of responses found in previous work now using updated materials that were normed and validated as part of the current project.

General discussion

This study presents updated short story ("vignette") materials that are effective at influencing people's judgments about assertions. In Part 1, we normed 39 assertions on real-world topics to obtain baseline rates at which participants from both U.S.-census matched and general U.S. online samples would classify them as true or false. In Part 2, we embedded the normed assertions in newly created vignette stimuli and validated that these materials influenced participants' post-reading judgments. Across two experiments drawing from U.S. census-matched and general online populations, exposure to the accurate and inaccurate assertions presented in the vignettes affected participants' subsequent validity judgments. Reading texts containing inaccurate assertions increased judgment errors, and reading texts containing accurate assertions decreased judgment errors, relative to responses made after not reading any related information. In Part 3, we replicated and extended these findings by testing whether reading the vignettes affected participants' confidence in their post-reading validity judgments. As in Salovich et al. (2021), participants overall were more confident in their correct versus incorrect validity judgments; however, this confidence–accuracy resolution was attenuated after reading vignettes containing inaccurate assertions. These results corroborate the idea that exposures to false information can affect people's confidence in what they believe to be true (e.g., Rapp & Salovich, 2018), and importantly for the current project, demonstrate that the new materials function as intended by replicating an emerging finding in the literature.

Despite increasing interest in how exposure to inaccurate or misleading information in fiction can affect people's beliefs, researchers have often relied on outdated norms and lengthy materials to examine these effects. The updated

vignettes validated in this paper offer the opportunity to both efficiently and flexibly explore the consequences of reading true and false ideas in stories, as regularly occurs in experiences with fiction (Gerrig, 1993). The vignettes are short, easy-to-read, independent of one another, and each accompanied with a comprehension check, making them well suited for the increasingly popular use of remote and/or online data collection. Besides the practicality of materials, the accurate and inaccurate assertions included within the vignettes also help represent the complexity and nuance of the types of misleading claims and ideas people are routinely exposed to every day (e.g., via popular press headlines, social media threads, and blog posts), and use topics that are of contemporary public concern (e.g., preventative health behavior, fad diets, and use of alternative medicines and therapies). Unlike most prior norming research, which has focused on identifying known versus unknown general knowledge trivia items (e.g., participants are likely to know that Jupiter is the largest planet in the solar system, but less likely to know that Michelangelo's statue of David is located in the city of Florence; e.g., Marsh, 2004; Tauber et al., 2013), the current project focused on assertions that potentially reflect commonly held correct beliefs as well as problematic misconceptions relevant to people's lives. Some assertions were associated with consistently correct validity judgments (indicating commonly known information), some with more mixed responses (suggesting uncertainty or potential variability in beliefs within the sample), and even some with consistently incorrect judgments (indicating commonly held misconceptions). This uniquely allows for engaging in research that assesses the influence of exposures to inaccurate information in fiction when participants, on average, possess correct understandings (i.e., evidence-aligned prior knowledge, such as correctly believing that toothbrushing prevents gum disease) or misconceptions (i.e., evidence-misaligned prior knowledge, such as incorrectly believing vitamin C is an effective cure for colds). In other words, both the norming and vignettes identify novel topics fruitful for testing the influence of misinformation as well as its correction.

In addition to projects considering how story content may interact with and/or influence people's prior understandings, the assertion-based vignettes afford the ability to conduct other investigations that would be difficult, if not impossible, using true and false declarative facts. For example, researchers could manipulate the degree to which explanations are provided in support of accurate or inaccurate claims, such as varying the amount of justification for why toothbrushing might be beneficial or harmful, or why seatbelts may or may not be an effective safety precaution (Lassonde et al., 2016; Rapp et al., 2020). Other projects could consider varying the actual contents of the justifications, perhaps using logic or fact-based arguments, moral-based arguments, anecdotal evidence, and so on, to explore the potential differential influence of these types of reasonings given recent interest in their persuasive power (Kubin et al., 2021; Wolsko et al., 2016).

A focus on assertions included in short stories also affords possible manipulations related to the characters and/or context of the narrative. For instance, researchers could vary the trustworthiness and/or credibility of the source of the information communicated within the vignette, either manipulated explicitly (e.g., through expectations based on past behaviors of the character; Andrews & Rapp, 2014; Rapp & Gerrig, 2006; Sparks & Rapp, 2011; Wertgen et al., 2021) or implicitly (e.g., through demographic characteristics like race, gender, political affiliation, etc.; Groggel et al., 2019; Mena et al., 2020; Rapp et al., 2019; Swire et al., 2017). Researchers could also modify presentations of the vignette content to offer them in different modalities or information environments (Corneille et al., 2020; Fazio, Dolan, & Marsh, 2015b). For example, with growing interest in people's uptake of information online, the vignettes could be adapted as short-form videos (e.g., Butler et al., 2009), as posts on social media (e.g., Pennycook et al., 2021a), or stripped of a narrative component completely (e.g., Fazio, Dolan, & Marsh, 2015b; Salovich et al., 2022). These kinds of manipulations are possible with the current materials as they are intentionally brief, simple, flexible, allow for participant accountability, and are realistic. They are also open to modification, which could be useful for making them even more authentic so as to fit in a variety of information-rich contexts.

The assertions and vignettes normed and validated in this paper could be implemented in many possible examinations at the intersection of discourse processing and epistemic cognition, and should prove informative for studies investigating people's experiences with inaccurate information (including misinformation, disinformation, and fake news contexts). Multiple-item materials can be resource intensive to both construct and test, particularly considering the added creative component demanded by writing texts. We hope that others will find these materials useful for investigating the theoretical and practical implications of exposures to accurate and inaccurate information.

Data availability All materials, data, and R code are publicly available on OSF (https://osf.io/5bvyx/).

Declarations

Conflicts of interest The authors declare that there were no conflicts of interest with respect to the authorship or publication of this article.

References

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk Research: Review and Recommendations. *Journal of Management, 47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Andrews, J. J., & Rapp, D. N. (2014). Partner characteristics and social contagion: Does group composition matter? *Applied Cognitive Psychology, 28*(4), 505–517. https://doi.org/10.1002/acp.3024

Appel, M., & Richter, T. (2007). Persuasive Effects of Fictional Narratives Increase Over Time. *Media Psychology, 10*(1), 113–134. https://doi.org/10.1080/15213260701301194

Arechar, A. A., & Rand, D. G. (2021). Turking in the time of COVID. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01588-4

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis, 20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123–144). Lawrence Erlbaum.

Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition, 194*, 104054. https://doi.org/10.1016/j.cognition.2019.104054

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use. *Perspectives on Psychological Science, 13*(2), 149–154. https://doi.org/10.1177/1745691617706516

Bulevich, J. B., & Thomas, A. K. (2012). Retrieval effort improves memory and metamemory in the face of misinformation. *Journal of Memory and Language, 67*(1), 45–58. https://doi.org/10.1016/j.jml.2011.12.012

United States Census Bureau (2017). *Educational Attainment in the United States: 2017*. https://www.census.gov/data/tables/2017/demo/education-attainment/cps-detailed-tables.html

Butler, A. C., Zaromb, F. M., Lyle, K. B., & Roediger, H. L. (2009). Using Popular Films to Enhance Classroom Learning: The Good, the Bad, and the Interesting. *Psychological Science, 20*(9), 1161–1168. https://doi.org/10.1111/j.1467-9280.2009.02410.x

Carrieri, V., Madio, L., & Principe, F. (2019). Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health Economics, 28*(11), 1377–1382. https://doi.org/10.1002/hec.3937

Coane, J. H., & Umanath, S. (2021). A database of general knowledge question performance in older adults. *Behavior Research Methods, 53*(1), 415–429. https://doi.org/10.3758/s13428-020-01493-2

Corneille, O., Mierop, A., & Unkelbach, C. (2020). Repetition increases both the perceived truth and fakeness of information: An ecological account. *Cognition, 205*, 104470. https://doi.org/10.1016/j.cognition.2020.104470

Dehon, H., & Brédart, S. (2004). False Memories: Young and Older Adults Think of Semantic Associates at the Same Rate, but Young Adults Are More Successful at Source Monitoring. *Psychology and Aging, 19*(1), 191–197. https://doi.org/10.1037/0882-7974.19.1.191

Donovan, A. M., & Rapp, D. N. (2020). Look it up: Online search reduces the problematic effects of exposures to inaccuracies. *Memory & Cognition, 48*(7), 1128–1145. https://doi.org/10.3758/s13421-020-01047-z

Donovan, A. M., Theodosis, E., & Rapp, D. N. (2018). Reader, interrupted: Do disruptions during encoding influence the use of inaccurate information? *Applied Cognitive Psychology, 32*(6), 775–786. https://doi.org/10.1002/acp.3464

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015a). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Fazio, L. K., Dolan, P. O., & Marsh, E. J. (2015b). Learning misinformation from fictional sources: Understanding the contributions of transportation and item-specific processing. *Memory, 23*(2), 167–177. https://doi.org/10.1080/09658211.2013.877146

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

Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading* (pp. xi, 273). Yale University Press.

Gerrig, R. J., & Prentice, D. A. (1991). The Representation of Fictional Information. *Psychological Science, 2*(5), 336–340. https://doi.org/10.1111/j.1467-9280.1991.tb00162.x

Gibson, A., Alarcon, G., Lee, M., Hamdan, I. A. (2021). 'Could you please pay attention?' Comparing in-person and MTurk responses on a computer code review task. *Proceedings of the 54th Hawaii International Conference on System Sciences*. http://hdl.handle.net/10125/71121

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making, 26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Groggel, A., Nilizadeh, S., Ahn, Y.-Y., Kapadia, A., & Rojas, F. (2019). Race and the beauty premium: Mechanical Turk workers' evaluations of Twitter accounts. *Information, Communication & Society, 22*(5), 709–716. https://doi.org/10.1080/1369118X.2018.1543443

Higham, P. A., Luna, K., & Bloomfield, J. (2011). Trace-strength and source-monitoring accounts of accuracy and metacognitive resolution in the misinformation paradigm. *Applied Cognitive Psychology, 25*(2), 324–335. https://doi.org/10.1002/acp.1694

Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: Plausibility effects on memory for misinformation. *Memory & Cognition, 42*(2), 305–324. https://doi.org/10.3758/s13421-013-0359-9

Horry, R., Colton, L.-M., & Williamson, P. (2014). Confidence–accuracy resolution in the misinformation paradigm is influenced by the availability of source cues. *Acta Psychologica, 151*, 164–173. https://doi.org/10.1016/j.actpsy.2014.06.006

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics, 2*(3), 2053168015604648. https://doi.org/10.1177/2053168015604648

Imundo, M. N., & Rapp, D. N. (2021). When fairness is flawed: Effects of false balance reporting and weight-of-evidence statements on beliefs and perceptions of climate change. *Journal of Applied Research in Memory and Cognition.* https://doi.org/10.1016/j.jarmac.2021.10.002

Ipeirotis, P. G. (2010). *Demographics of Mechanical Turk* (SSRN Scholarly Paper ID 1585030). Social Science Research Network. https://papers.ssrn.com/abstract=1585030

Jalbert, M., Newman, E., & Schwarz, N. (2019). *Trivia claim norming: Methods report and data*. ResearchGate. https://doi.org/10.6084/m9.figshare.9975602

Kubin, E., Puryear, C., Schein, C., & Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proceedings of the National Academy of Sciences, 118*(6). https://doi.org/10.1073/pnas.2008389118

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(1), 1–26.

Lassonde, K. A., Kendeou, P., & O'Brien, E. J. (2016). Refutation texts: Overcoming psychology misconceptions that are resistant to change. *Scholarship of Teaching and Learning in Psychology, 2*(1), 62–74. https://doi.org/10.1037/stl0000054

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.3.01. https://CRAN.R-project.org/package=emmeans

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

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour, 5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences, 118*(23). https://doi.org/10.1073/pnas.2019527118

Marsh, E. J. (2004). Story stimuli for creating false beliefs about the world. *Behavior Research Methods, Instruments, & Computers, 36*(4), 650–655. https://doi.org/10.3758/BF03206546

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition, 34*(5), 1140–1149. https://doi.org/10.3758/BF03193260

Marsh, E. J., Meade, M. L., & Roediger III, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language, 49*(4), 519–536. https://doi.org/10.1016/S0749-596X(03)00092-5

Mellis, A. M., & Bickel, W. K. (2020). Mechanical Turk data collection in addiction research: Utility, concerns and best practices. *Addiction, 115*(10), 1960–1968. https://doi.org/10.1111/add.15032

Mena, P., Barbe, D., & Chan-Olmsted, S. (2020). Misinformation on Instagram: The impact of trust endorsements on message credibility. *Social Media + Society*, 1–9. https://doi.org/10.1177/2056305120935102

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4*(2), 93–102. https://doi.org/10.1016/j.jarmac.2015.01.003

Miller, J. D. (2004). Public Understanding of, and Attitudes toward, Scientific Research: What We Know and What We Need to Know. *Public Understanding of Science, 13*(3), 273–294. https://doi.org/10.1177/0963662504044908

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 & Verbal Behavior, 19*, 338–368. https://doi.org/10.1016/S0022-5371(80)90266-2

Ogletree, A. M., & Katz, B. (2021). How do older adults recruited using MTurk differ from those in a national probability sample? *The International Journal of Aging and Human Development, 93*(2), 700–721. https://doi.org/10.1177/0091415020940197

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. *Trends in Cognitive Sciences, 25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021a). A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology, 7*(1), 25293. https://doi.org/10.1525/collabra.25293

Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2021b). Beliefs About COVID-19 in Canada, the United Kingdom, and the United States: A Novel Test of Political Polarization and Motivated Reasoning. *Personality and Social Psychology Bulletin*, 01461672211023652. https://doi.org/10.1177/01461672211023652

Prentice, D. A., Gerrig, R. J., Bailis, D.S. (1997). What readers bring to the processing of fictional texts. *Psychonomic Bulletin Review, 4*(3), 416–420. https://doi.org/10.3758/BF03210803

Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition, 36*(3), 688–701. https://doi.org/10.3758/MC.36.3.688

Rapp, D. N. (2016). The Consequences of Reading Inaccurate Information. *Current Directions in Psychological Science, 25*(4), 281–285. https://doi.org/10.1177/0963721416649347

Rapp, D. N., & Gerrig, R. J. (2006). Predilections for narrative outcomes: The impact of story contexts and reader preferences. *Journal of Memory and Language, 54*(1), 54–67. https://doi.org/10.1016/j.jml.2005.04.003

Rapp, D. N., & Salovich, N. A. (2018). Can't We Just Disregard Fake News? The Consequences of Exposure to Inaccurate Information. *Policy Insights from the Behavioral and Brain Sciences, 5*(2), 232–239. https://doi.org/10.1177/2372732218785193

Rapp, D. N., Hinze, S. R., Kohlhepp, K., & Ryskin, R. A. (2014). Reducing reliance on inaccurate information. *Memory & Cognition, 42*(1), 11–26. https://doi.org/10.3758/s13421-013-0339-0

Rapp, D. N., Imundo, M. N., & Adler, R. M. (2019). Do individual differences in conspiratorial and political leanings influence the use of inaccurate information? In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and fake news in education* (pp. 103–120). Information Age Publishing.

Rapp, D. N., Donovan, A. M., & Salovich, N. A. (2020). Assessing and Modifying Knowledge: Facts vs. Constellations. In: *Handbook of Learning from Multiple Representations and Perspectives*. Routledge.

Richter, T. (2006). What Is Wrong With ANOVA and Multiple Regression? Analyzing Sentence Reading Times With Hierarchical Linear Models. *Discourse Processes, 41*(3), 221–250. https://doi.org/10.1207/s15326950dp4103_1

Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Mechanical Turk. *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 2863–2872. https://doi.org/10.1145/1753846.1753873

Salovich, N. A., & Rapp, D. N. (2021). Misinformed and unaware? Metacognition and the influence of inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(4), 608–624. https://doi.org/10.1037/xlm0000977

Salovich, N. A., Donovan, A. M., Hinze, S. R., & Rapp, D. N. (2021). Can confidence help account for and redress the effects of reading inaccurate information? *Memory & Cognition, 49*(2), 293–310. https://doi.org/10.3758/s13421-020-01096-4

Salovich, N. A., Kirsch, A. M, & Rapp, D. N. (2022). Evaluative mindsets can protect against the influence of false information. *Cognition*. Advance online publication. https://doi.org/10.1016/j.cognition.2022.105121

Saravanos, A., Zervoudakis, S., Zheng, D., Stott, N., Hawryluk, B., & Delfino, D. (2021). The Hidden Cost of Using Amazon Mechanical Turk for Research. In C. Stephanidis, M. M. Soares, E. Rosenzweig, A. Marcus, S. Yamamoto, H. Mori, P.-L. P. Rau, G. Meiselwitz, X. Fang, & A. Moallem (Eds.), *HCI International 2021—Late Breaking Papers: Design and User Experience* (pp. 147–164). Springer International Publishing. https://doi.org/10.1007/978-3-030-90238-4_12

Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research, 69*(8), 3139–3148. https://doi.org/10.1016/j.jbusres.2015.12.002

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 230–247. https://doi.org/10.1037/a0021331

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science, 4*, 160802. https://doi.org/10.1098/rsos.160802

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. https://doi.org/10.3758/s13428-012-0307-9

United States Census Bureau (2017). Educational Attainment in the United States: 2017. https://www.census.gov/data/tables/2017/demo/education-attainment/cps-detailed-tables.html

University of Michigan News. (2011). U.S. public's knowledge of science: Getting better but a long way to go. https://news.umich.edu/u-s-public-s-knowledge-of-science-getting-better-but-a-long-way-to-go/

U.S. Department of Education, National Center for Education Statistics, Program for the International Assessment of Adult Competencies (PIAAC). (2017). W*hat are the literacy levels of adults, and how does the United States compare to other countries?* https://nces.ed.gov/fastfacts/display.asp?id=69

University of Michigan News. (2011). *U.S. public's knowledge of science: Getting better but a long way to go.* https://news.umich.edu/u-s-public-s-knowledge-of-science-getting-better-but-a-long-way-to-go/

Unkelbach, C., & Speckmann, F. (2021). Mere Repetition Increases Belief in Factually True COVID-19-Related Information. *Journal of Applied Research in Memory and Cognition, 10*(2), 241–247. https://doi.org/10.1016/j.jarmac.2021.02.001

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Weinstein, Y., McDermott, K. B., & Chan, J. C. K. (2010). True and false memories in the DRM paradigm on a forced choice test. *Memory (Hove, England), 18*(4), 375–384. https://doi.org/10.1080/09658211003685533

Wertgen, A. G., Richter, T., & Rouet, J.-F. (2021). The Role of Source Credibility in the Validation of Information Depends on the Degree of (Im-)Plausibility. *Discourse Processes, 58*(5–6), 513–528. https://doi.org/10.1080/0163853X.2021.1881342

Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology, 65*, 7–19. https://doi.org/10.1016/j.jesp.2016.02.005