

1 You [Don't] Have to Tell a Story! A Registered Report Testing the Effectiveness of
2 Narrative versus Non-narrative Misinformation Corrections

3 Ullrich K. H. Ecker¹, Lucy H. Butler¹ & Anne Hamby²

4 ¹ School of Psychological Science, University of Western Australia, 35 Stirling Hwy, Perth
5 6009, Australia; ullrich.ecker@uwa.edu.au (UE), 21984487@student.uwa.edu.au (LB)

6 ² College of Business and Economics, Boise State University, 1910 University Drive, Boise,
7 ID, 83725, USA; annehamby@boisestate.edu

8 Word count: 5,909 (main text only, including table captions)

9 Address correspondence to: Ullrich Ecker, School of Psychological Science (M304),
10 University of Western Australia, 35 Stirling Hwy, Perth 6009, Australia. Telephone: +618
11 6488 3257; e-mail: ullrich.ecker@uwa.edu.au.

12 Abstract: Misinformation often has an ongoing effect on people’s memory and inferential
13 reasoning even after clear corrections are provided; this is known as the continued influence
14 effect. In pursuit of more effective corrections, one factor that has not yet been investigated is
15 the narrative versus non-narrative format of the correction. Some scholars have suggested
16 that a narrative format facilitates comprehension and retention of complex information, and
17 may serve to overcome resistance to worldview-dissonant corrections. It is, therefore, a
18 possibility that misinformation corrections are more effective if they are presented in a
19 narrative format versus a non-narrative format. The present study tests this possibility. We
20 designed corrections that are either narrative or non-narrative, while minimizing differences
21 in informativeness. We compared narrative and non-narrative corrections in three pre-
22 registered experiments. Experiment 1 targeted misinformation contained in fictional event
23 reports; Experiment 2 used false claims commonly encountered in the real world;
24 Experiment 3 used real-world claims that are controversial, in order to test the notion that a
25 narrative format may facilitate corrective updating primarily when it serves to reduce
26 resistance to correction. In all experiments, we also manipulated test delay (immediate vs.
27 two days), as any potential benefit of the narrative format may only arise in the short term (if
28 the story format aids primarily with initial comprehension and updating of the relevant
29 mental model) or after a delay (if the story format aids primarily with later correction
30 retrieval). It was found that narrative corrections are (no) more effective than non-narrative
31 corrections. [...]

32 Keywords: Misinformation; Continued influence effect; Myth debunking; Narrative
33 processing; Stories

34 Significance statement: Misinformation often has an ongoing effect on people’s reasoning
35 even after they receive corrections. Therefore, to reduce the impact of misinformation, it is
36 important to design corrections that are as effective as possible. One suggestion often made
37 by front-line communicators is to use stories to convey complex information. The rationale is
38 that humans are uniquely “tuned” to stories, such that the narrative format facilitates
39 understanding and retention of complex information. Some scholars have also suggested that
40 a story format may help overcome resistance to corrections that threaten a worldview-
41 consistent misconception. It is, therefore, a possibility that misinformation corrections are
42 more effective if they are presented in a narrative versus a non-narrative, more fact-oriented
43 format. The present study tests this possibility. We designed narrative and non-narrative
44 corrections that differ in format while conveying the same relevant information. In
45 Experiment 1, corrections targeted misinformation contained in fictional event reports. In
46 Experiment 2, the corrections targeted false claims commonly encountered in the real world.
47 Experiment 3 used real-world claims that are controversial, in order to test the notion that a
48 narrative format may facilitate corrective updating primarily when it serves to reduce
49 resistance to correction. In all experiments, we also manipulated test delay, as any benefit of
50 the narrative format may only arise in the short term (if the story format aids primarily with
51 initial understanding) or after a delay (if the story format aids primarily with later memory for
52 the correction). It was found that narrative corrections are (no) more effective than non-
53 narrative corrections. Thus, front-line communicators should be advised to [...].

54

55 You [Don't] Have to Tell a Story! A Registered Report Testing the Effectiveness of
56 Narrative versus Non-narrative Misinformation Corrections

57 The contemporary media landscape is awash with false information (Lazer et al.,
58 2018; Southwell & Thorson, 2015; Vargo, Guo, & Amazeen, 2018). Misinformation featured
59 in the media ranges from preliminary accounts of newsworthy events that are superseded by
60 more accurate accounts as evidence accrues (e.g., a wildfire is initially believed to be arson-
61 related but is later found to have been caused by a fallen power pole), to commonly
62 encountered “myths” about causal relations (e.g., alleged links between childhood
63 vaccinations and various negative health outcomes), to strategically disseminated
64 disinformation that intends to deceive, confuse, and sow social division (e.g., doctored stories
65 intended to discredit or denigrate a political opponent during an election campaign; see
66 Lewandowsky, Ecker, & Cook, 2017).

67 From a psychological perspective, an insidious aspect of misinformation is that it
68 often continues to influence people’s reasoning after a clear correction has been provided,
69 even when there are no motivational reasons to dismiss the correction; this is known as the
70 continued influence effect (CIE; Johnson & Seifert, 1994; Rapp & Salovich, 2018; Rich &
71 Zaragoza, 2016; Thorson, 2016; for reviews see Chan, Jones, Hall Jamieson, & Albarracín,
72 2017; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Walter & Tukachinsky, 2020).
73 Theoretically, the CIE is thought to arise either from failure to integrate the corrective
74 information into the mental model of the respective event or causal relationship, or from
75 selective retrieval of the misinformation (e.g., familiarity-driven retrieval of the
76 misinformation accompanied by failure to recollect the correction; see Ecker, Lewandowsky,
77 & Tang, 2010; Gordon, Brooks, Quadflieg, Ecker, & Lewandowsky, 2017; Gordon,
78 Quadflieg, Brooks, Ecker, & Lewandowsky, 2019; Rich & Zaragoza, 2016; Walter &
79 Tukachinsky, 2020).

80 Given the omnipresence of misinformation, it is of great importance to investigate the
81 factors that make corrections more effective. For example, corrections are more effective if
82 they come from a more credible source (Guillory & Geraci, 2013; Vraga, Bode, & Tully,
83 2020), contain greater detail (Chan et al., 2017; Swire, Ecker, & Lewandowsky, 2017), or a
84 greater number of counterarguments (Ecker, Lewandowsky, Jayawardana, & Mladenovic,
85 2019). However, even optimized debunking messages typically cannot eliminate the
86 continued influence of misinformation, not even if reasoning is tested immediately after a
87 correction is provided, let alone after a delay (see Ecker et al., 2010; Ecker, O'Reilly, Reid, &
88 Chang, 2020; Paynter et al., 2019; Rich & Zaragoza, 2016; Swire et al., 2017; Walter &
89 Tukachinsky, 2020). Thus, additional factors to enhance the effectiveness of corrections need
90 to be identified. The present paper is thus concerned with one particular avenue that might
91 make corrections more effective, which is important because greater correction effects mean
92 smaller continued influence effects.

93 Specifically, one piece of advice often given by educators and science communicators
94 regarding the communication of complex information, such as misinformation corrections, is
95 to use stories (e.g., Brewer, Chapman, Rothman, Leask, & Kempe, 2017; Caulfield et al.,
96 2019; Dahlstrom, 2014; Klassen, 2010; Marsh, Butler, & Umanath, 2012; Shelby & Ernst,
97 2013). For example, Shelby and Ernst (2013) argued that part of the reason why some
98 misconceptions are common amongst the public is that disinformants use the power of
99 storytelling, while fact-checkers often rely exclusively on facts and evidence. Indeed, people
100 seem to be influenced by anecdotes and stories more so than stated facts or statistical
101 evidence in their medical decision-making (Bakker, Kerstholt, van Bommel, & Giebels,
102 2019; Fagerlin, Wang, & Ubel, 2005), risk perceptions (Betsch, Renkewitz, & Haase, 2013;
103 de Wit, Das, & Vet, 2008; Haase, Betsch, & Renkewitz, 2015), behavioral intentions and

104 choices (Borgida & Nisbett, 1977; Dillard, Ferrer, & Welch, 2018), and attitudes (Lee &
105 Leets, 2002).

106 Despite some fragmentation in defining what constitutes a story, researchers generally
107 agree that stories are defined by their chronology and causality: they depict characters
108 pursuing goals over time, and may feature access to characters' thoughts and emotions
109 (Brewer & Liechtenstein, 1982; Bruner, 1986; Pennington & Hastie, 1988; Shen, Ahern, &
110 Baker, 2014; van Krieken & Sanders, 2019). Research on narrative processing often contrasts
111 narrative messages with non-narrative formats (such as those that feature statistics or facts,
112 descriptive passages, or list-based, informative format; sometimes these are also called
113 "expository" or "informational" texts; Ratcliff & Sun, 2020; Reinhart, 2006; Shen et al.,
114 2014; Zebregs, van den Putte, Neijens, & de Graaf, 2015). Though non-narrative formats
115 may differ in form and substance, they often share an abstract, logic-based, decontextualized
116 message style (relative to narratives), and tend to evoke analytical processing. Research from
117 advertising and consumer psychology suggests that even short messages featuring several
118 lines of text can evoke narrative or analytical processing styles, based on their content
119 (Chang, 2009; Escalas, 2007; Kim, Ratneshwar, & Thorson, 2017).

120 Stories can impact reasoning and decision making through several mechanisms (see
121 Hamby, Brinberg, & Jaccard, 2018; Schaffer, Focella, Hathaway, Scherer, & Zikmund-
122 Fisher, 2018). Compared to processing of non-narrative messages, narrative processing is
123 usually associated with greater emotional involvement in the message (Busselle & Bilandzic,
124 2008; Golke, Hagen, & Wittwer, 2019; Green & Brock, 2000; Ratcliff & Sun, 2020). While
125 narrative and non-narrative messages can be cognitively engaging, the nature of engagement
126 differs. Readers of narratives apply more imagery and visualization, and may even report
127 feelings of transportation into to the world of the story in which they experience story events
128 as though they were happening to them personally (Bower & Morrow, 1990; Green & Brock,

129 2000; Hamby et al., 2018; Mar & Oatley, 2008). Additionally, narrative processing tends to
130 reduce resistance to message content; not only are narratives usually less overtly persuasive
131 than their non-narrative counterparts, but audiences are often less motivated to generate
132 counterarguments when processing narratives, as this would disrupt the enjoyable experience
133 of immersion in the story (Green & Brock, 2000; Krakow, Yale, Jensen, Carcioppolo, &
134 Ratcliff, 2018; Slater & Rouner, 1996). Stories may thus lead to stronger encoding and
135 comprehension of information embedded within because of the cognitive and emotional
136 involvement they tend to evoke (Browning & Hohenstein, 2015; Romero, Paris, & Brem,
137 2005; Zabrocky & Moore, 1999).

138 In addition, a story format may facilitate information retrieval (Bower & Clark, 1969;
139 Graesser, Hautt-Smith, Cohen, & Pyles, 1980). This may arise from the afore-mentioned
140 enhanced processing at encoding, to the extent that enhanced encoding results in a more vivid
141 and coherently integrated memory representation (Graesser & McNamara, 2011). Bruner
142 (1986) argued that the story format provides the most fundamental means by which people
143 construct reality, and enhanced retrieval of information presented in story format may
144 therefore also result from the fact that stories typically offer a structured series of retrieval
145 cues (e.g., markers of spatio-temporal context or characters' emotional states or
146 introspections) that are consistent with the way in which people generally think. A more
147 easily-retrieved correction during a subsequent reasoning task will naturally promote use of
148 correct information and reduce reliance on the corrected misinformation (see Ecker,
149 Lewandowsky, Swire, & Chang, 2011).

150 However, the evidence regarding the persuasive superiority of the story format over
151 non-narrative text is not entirely consistent. Some studies contrasting narrative and non-
152 narrative formats of health-related messages found both formats equally able to effect
153 changes to attitudes and behavioral intentions (Dunlop, Wakefield, & Kashima, 2010;

154 Zebregs, van den Putte, de Graaf, Lammers, & Neijens, 2015). Greene and Brinn (2003) even
155 reported that narratives were inferior to non-narrative texts in reducing use of tanning beds.
156 Early meta-analyses found that narrative information is either less persuasive than statistical
157 information (Allen & Preiss, 1997) or that there is no clear difference in favor of either
158 approach (Reinhart, 2006). More recent meta-analyses, however, found stronger support for
159 the narrative approach (e.g., Ratcliff & Sun, 2020), but also highlighted that communication
160 effectiveness depended on persuasion context: While Zebregs, van den Putte, Neijens et al.'s
161 (2015) analysis found that narrative information was superior to statistical information when
162 it comes to changing behavioral intentions, they found that statistical evidence had stronger
163 effects on attitudes and beliefs. Shen, Sheer, and Li (2015) found that narratives were more
164 effective than non-narrative communications when it came to fostering prevention but not
165 cessation behaviors.

166 Similar to the approach taken in the present study, Golke et al. (2019) contrasted
167 standard non-narrative texts with so-called “informative narratives”—enhanced fact-based
168 texts that present essentially the same information as the standard non-narrative fact-based
169 text, but in a storyline format. They found that the narrative format did not enhance reading
170 comprehension, and even reduced comprehension in two of their three experiments. Wolfe
171 and Mienko (2007) found no retrieval benefit for informative narratives, and Wolfe and
172 Woodwyk (2010) reported that readers showed enhanced integration of new information with
173 existing knowledge when reading non-narrative texts compared to informative narratives. In
174 the context of misinformation corrections, this may suggest that narrative elements may
175 distract the reader from the core correction, and/or that non-narrative corrections may
176 facilitate integration of the correction into the reader’s mental model, which may render them
177 more effective than informative-narrative corrections (see Kendeou, Walsh, Smith, &
178 O’Brien, 2014).

179 In sum, while there may be some rationale in using a story format to correct
180 misinformation, the question of whether corrections will be more effective when they are
181 given in a story format rather than a non-narrative format remains to be empirically tested. To
182 the best of our knowledge, only one study has investigated the effectiveness of narrative
183 corrections. Sangalang, Ophir, and Cappella (2019) explored whether narrative corrections
184 could reduce smokers' misinformed beliefs about tobacco. Results were inconclusive, as a
185 narrative correction was found to reduce misconceptions in only one of the two experiments
186 reported. Importantly, this study did not contrast narrative and non-narrative corrections. This
187 is the aim of the present study.

188 In three experiments, we will contrast corrections that focus on factual evidence with
189 corrections designed to present the same amount of relevant corrective information, but in a
190 narrative format. In designing these corrections, we took inspiration from the broader
191 literature on narrative persuasion reviewed above (in particular, Shen et al., 2014; van
192 Krieken & Sanders, 2019) to ensure narrative and non-narrative corrections differed on
193 relevant dimensions. Narrative corrections featured characters' experiences and point of
194 view, quotes, chronological structure, and/or some form of complication or climax, whereas
195 non-narrative corrections focused more on the specific facts and pieces of evidence, had a
196 less engaging and emotive writing style, and adhered more closely to an inverted-pyramid
197 format (essential facts followed by supportive evidence and more general background
198 information).

199 In order to investigate the robustness of potential narrative effects, we aim to correct
200 both fictional event misinformation and real-world misconceptions: Experiment 1 will use
201 fictional event reports of the type used in most research on the continued influence effect
202 (e.g., Ecker, Hogan, & Lewandowsky, 2017). The reports first introduce a piece of critical
203 information that relates to the cause of the event, while the correction refutes that piece of

204 critical information. Participants' inferential reasoning regarding the event, in particular their
205 reliance on the critical information, is then measured via questionnaire. Experiment 2 will
206 correct some common real-world "myths" while affirming some obscure facts (as in Swire et
207 al., 2017). We will measure change in participants' beliefs, as well as their post-treatment
208 inferential reasoning relating to the false claims. Experiment 3 will examine the effect of
209 correction format in the context of more controversial real-world claims. To the extent a
210 narrative advantage arises from reduced resistance to the corrective message (see Green &
211 Brock, 2000; Krakow et al., 2018; Slater & Rouner, 1996), it should become particularly
212 apparent with corrections of worldview-consistent misconceptions. We hypothesize that
213 narrative corrections will generally be more effective at reducing misinformation-congruent
214 reasoning and beliefs.

215 In all experiments, we will additionally manipulate retention interval (i.e., the study-
216 test delay). The rationale for this is as follows: Any potential story benefit might arise
217 immediately—to the extent that the narrative format boosts engagement with and
218 comprehension of the correction, and thus facilitates its mental-model integration. However,
219 a story benefit may only arise after a delay—to the extent that the narrative format facilitates
220 correction retrieval at test, which will be more relevant after some delay-related forgetting
221 has occurred. In other words, if the narrative format is beneficial for retrieval, this benefit
222 may not become apparent in an immediate test because participants are likely to remember
223 both the narrative and the non-narrative correction just minutes after encoding; however, a
224 story benefit may emerge with a delay, when the corrections are no longer "fresh" in one's
225 memory (see Ecker et al., 2020; Swire et al., 2017).

226

Experiment 1

Method

228 Experiment 1 will present fictional event reports in four conditions. There will be two
229 control conditions: One will feature no misinformation (noMI condition), another will feature
230 a piece of misinformation that is not corrected (noC condition). The two experimental
231 conditions will correct the initially-provided misinformation using either a non-narrative
232 (NN) or narrative (N) correction. The test phase will follow the study phase either
233 immediately or after a two-day delay. The experiment will thus use a mixed within-between
234 design, with the within-subjects factor of condition (noMI; NN; N; noC), and the between-
235 subjects factor of test delay (immediate; delayed).

236 **Participants.** Participants will be U.S.-based adults recruited via the platform
237 Prolific.¹ An a-priori power analysis (using G*Power 3; Faul, Erdfelder, Lang, & Buchner,
238 2007) suggested a minimum sample size of $N = 352$ to detect a small difference between the
239 two within-subjects experimental conditions (i.e., NN vs. N; effect size $f = 0.15$; $\alpha = 0.05$,
240 $1 - \beta = 0.8$). As the core planned analyses will test for effects in each delay condition
241 separately, and to achieve an adequate sample size post exclusions, it was thus decided to aim
242 for a total of $N = 800$ participants ($n = 400$ per delay condition). Due to inevitable dropout in
243 the delayed condition (estimated at 20%), this condition will be oversampled by a factor of
244 1.25 (i.e., 500 participants will complete the study phase).

245 **Materials.** Experiment 1 will use four fictitious event reports detailing four different
246 newsworthy events (e.g., a wildfire); each report comprises two articles. In the study phase,
247 participants will be presented with all four reports in the four different conditions. In three of
248 the conditions, the report's first article contains a piece of misinformation (e.g., the wildfire

¹ Prolific (<https://www.prolific.co/>) is a recruitment platform known for high-quality data (e.g., Peer, Brandimarte, Samat, & Acquisti, 2017).

249 was caused by arson; this will simply be omitted from the report in the no-misinformation
250 condition); in these conditions, the report’s second article either contains or does not contain
251 a correction. If a correction is provided, it will be given in either a non-narrative format (e.g.,
252 explaining that an investigation had found that a rotten power pole had fallen and the power
253 line had melted on the ground, starting the fire) or a narrative format (e.g., explaining that a
254 fire chief inspected the scene, found the power pole, noticed the rot, and discovered that the
255 power line had melted on the ground, concluding it had started the fire). Narrative and non-
256 narrative corrections thus present the same critical corrective information, but differ in the
257 way it is presented: Narrative corrections feature specific characters and a causally-ordered
258 description sequence; non-narrative corrections feature objective, generalized descriptions of
259 the events (per our definition of narrative and non-narrative format; Brewer & Liechtenstein,
260 1982; Bruner, 1986; Pennington & Hastie, 1988; Shen et al., 2014; van Krieken & Sanders,
261 2019). All reports thus exist in four versions (matching the conditions; all report versions are
262 provided in the Appendix). We aimed to keep non-narrative and narrative reports as
263 equivalent as possible in terms of informativeness, length, and reading difficulty. A pilot
264 study confirmed that our narrative corrections were perceived as more “story-like” than the
265 non-narrative corrections, and also as more vivid and more easily allowing the events to be
266 imagined. By contrast, the two correction versions were rated as relatively comparable on
267 informativeness and comprehensibility (for details, see Appendix). Assignment of events to
268 experimental conditions, as well as condition and event order, will be counterbalanced across
269 participants using four different presentation sequences in a Latin-square design, as shown in
270 Table 1.

271 The test will comprise a memory question and six inference questions per report. The
272 memory questions are four-alternative-choice questions targeting an arbitrary detail provided
273 twice in the report (once in each article; e.g., “The fire came close to the town of Cranbrook /

274 Kimberley / Lumberton / Bull River”). The sole purpose of the memory questions is to ensure
 275 adequate encoding; data from participants who do not demonstrate adequate encoding will be
 276 excluded from analysis (see exclusion criteria below). The inference questions are designed
 277 to measure misinformation-congruent inferential reasoning, following previous CIE research
 278 (e.g., Ecker et al., 2017). Five of the six inference questions per report are rating scales
 279 asking participants to rate their agreement with a misinformation-related statement on a 0-10
 280 Likert scale (e.g., “Devastating wildfire intentionally lit” would be an appropriate headline
 281 for the report). One inference question is a four-alternative-choice question targeting the
 282 misinformation directly (e.g., “What do you think caused the wildfire? Arson / Lightning /
 283 Power line / None of the above”). Such measures have been found appropriate for online CIE
 284 studies (Connor Desai & Reimers, 2019). All questions are provided in the Appendix.

285 Table 1

286 *Presentation Sequences (S1-4) to be Used in Experiment 1*

	Pos 1	Pos 2	Pos 3	Pos 4
S1	A_noMI	B_NN	C_noC	D_N
S2	B_N	A_noC	D_NN	C_noMI
S3	C_NN	D_noMI	A_N	B_noC
S4	D_noC	C_N	B_noMI	A_NN

287 *Note.* Sequences counterbalance the assignment of event reports (A-D) to conditions (no-
 288 misinformation, noMI; non-narrative correction, NN; narrative correction, N; no correction,
 289 noC) as well as event and condition order across sequence positions (Pos 1-4). Assignment of
 290 presentation sequence to participants will be randomized, with the constraint that a quarter of
 291 participants will receive each sequence.

292 All materials will be presented via experimental surveys designed and administered
 293 via Qualtrics (Qualtrics, Provo, UT). The survey file, including all materials, will be made

294 available on the Open Science Framework (OSF; www.osf.io). Surveys with immediate and
295 delayed tests will necessarily be run separately due to the need for different sign-up
296 instructions. Participants in the delayed condition will be reminded via e-mail to complete the
297 test phase 48 hours after launch of the study phase; they will have 48 hrs to complete from
298 launch of the test phase but will be encouraged to complete within 24 hrs.

299 The experiment will take approximately 10 minutes. Participants in the immediate
300 condition will be reimbursed approximately GBP1.25 (US\$1.56) via Prolific; participants in
301 the delayed condition will be reimbursed approximately GBP0.58 (US\$0.72 for the study
302 phase and GBP0.67 (US\$0.84) for the test phase [times and amounts will be specified after
303 pilot testing].

304 **Procedure.** Initially, participants will be provided with an ethics-approved
305 information form. Participants will be asked to provide an English proficiency rating (1:
306 excellent to 5: poor), gender, and age information, and indicate their country of residence.
307 The four reports will then be presented, with each article presented on a separate screen, with
308 applied fixed minimum times (set at approx. 150 ms per word).

309 The test will follow after a short (1-minute) or long (2 days) retention interval.
310 Participants will be presented with a questionnaire for each report, each comprising the
311 memory question and the six inference questions. The order of questionnaires will follow the
312 order of the reports in the study phase; the order of questions in each questionnaire will be
313 fixed (see Appendix).

314 Following the test phase, participants will be given a “data use” question asking them
315 to provide honest feedback on whether or not their data should be included in our analysis
316 (“In your honest opinion should we use your data in our analysis? This is not related to how
317 well you think you performed, but whether you put in a reasonable effort.”). This question

318 can be answered with “Yes, I put in reasonable effort (1)”; “Maybe, I was a little distracted
319 (2)”; or “No, I really wasn’t paying any attention (3)”.

320 **Results**

321 Data analysis will be pre-registered on the OSF website, where the raw data will also
322 be made available. Analysis will adhere to the following procedure. First, exclusion criteria
323 will be applied. We will exclude data from participants who (a) indicate they do not reside in
324 the U.S.; (b) indicate their English proficiency is “fair” or “poor”; (c) respond to the “data
325 use” question with “No [do not use my data], I really wasn’t paying any attention”; (d) fail at
326 least three memory questions; (e) respond in a “cynical” manner by selecting the “none of the
327 above” response option for all four multiple-choice inference questions; (f) respond
328 uniformly (a response *SD* across all 20 raw rating-scale inference-question responses < 0.5).
329 Finally, to identify inconsistent, erratic responding, we will calculate response *SD* for each
330 set of five inference questions, and then calculate mean *SD* across the four sets. We will (g)
331 exclude outliers on this measure, using the inter-quartile rule with a 2.2 multiplier (i.e., cutoff
332 $= Q3 + 2.2 \times IQR$; Hoaglin & Iglewicz, 1987). We will code the multiple-choice inference-
333 question responses as either 10 (misinformation option) or 0 (non-misinformation options).
334 We will then calculate four mean inference scores for the noC, NN, N, and noMI conditions;
335 this will be the main dependent variable.

336 We will run a two-way mixed ANOVA with factors condition (within-subjects) and
337 delay (between-subjects) on inference scores. Irrespective of the outcome of the omnibus
338 ANOVA, we will test the core hypothesis with planned contrasts, assessing the difference
339 between NN and N conditions (planned contrast: $NN > N$; i.e., narrative correction more
340 effective at reducing reliance on misinformation than non-narrative correction) in each delay
341 condition. To complement frequentist analysis (and to potentially quantify evidence in favor
342 of the null), we will run Bayesian *t*-tests comparing NN and N in both delay conditions. We

343 will also test the interaction contrast of NN vs. N \times immediate vs. delayed. The direction of a
 344 potential interaction is not pre-specified: A potential narrative benefit may only emerge after
 345 a delay if the effect reflects retrieval facilitation, or may emerge immediately if it reflects
 346 stronger correction encoding or integration into the mental event model. Finally, for the sake
 347 of completeness, we will run and report an additional series of five secondary planned
 348 contrasts for each delay condition (see Table 2). Statistical significance will be established
 349 using the Holm-Bonferroni correction, applied separately to each set of contrasts.

350 Table 2

351 *Secondary Contrasts to be Run in Experiment 1*

#	Contrast	Effect
1	noMI < noC	Effect of uncorrected misinformation against no-misinformation baseline
2	noMI < NN	Continued influence effect of misinformation (non-narrative correction)
3	noMI < N	Continued influence effect of misinformation (narrative correction)
4	noC > NN	Effectiveness of non-narrative correction relative to no-correction baseline
5	noC > N	Effectiveness of narrative correction relative to no-correction baseline

352 Experiment 2

353 To examine the robustness of the results of Experiment 1, Experiment 2 will examine
 354 the effect of narrative versus non-narrative corrections on real-world beliefs.

355 Method

356 Experiment 2 will present claims encountered in the real world, including both true
 357 “facts” and common misconceptions, henceforth referred to as “myths”. Claims will be
 358 followed by explanations that affirm the facts and correct the myths. Corrections will either
 359 be in a non-narrative (NN) or narrative (N) form, and the test will again be either immediate
 360 or delayed. Thus, Experiment 2 will have a 2 \times 2 mixed within-between design, with the
 361 within-subjects factor of correction type (NN; N) and the between-subjects factor of test

362 delay (immediate; delayed). Fact-affirmation trials act as fillers outside of this design (even
363 though basic affirmation effects will be reported).

364 **Participants.** Experiment 2 will use the same recruitment procedures as
365 Experiment 1. Sample size will be increased by 10% to allow for the exclusion of participants
366 with any initial-belief ratings of zero (see below).² Participants who participated in
367 Experiment 1 will not be allowed to participate in Experiment 2.

368 **Materials.** Experiment 2 will use eight claims (four myths; four facts). An example
369 myth is “Gastritis and stomach ulcers are caused by excessive stress.” The non-narrative
370 corrections explain the evidence suggesting that the claim is false (e.g., that there is evidence
371 that gastritis and stomach ulcers are primarily caused by the bacterium *Helicobacter pylori*
372 and that this discovery earned the scientists involved a Nobel Prize); the narrative correction
373 details the story behind this discovery (e.g., that one of the scientists drank a broth
374 contaminated with the bacterium to prove his theory, which earned him and his colleague a
375 Nobel Prize). Again, a pilot study confirmed that the narrative corrections were perceived as
376 more story-like and vivid than the non-narrative correction, while being relatively
377 comparable on informativeness and comprehensibility dimensions (see Appendix for details).
378 Fact affirmations are of an expository nature similar to the non-narrative corrections. All
379 claims and explanations are provided in the Appendix.

380 Each participant will receive two NN and two N corrections. Assignment of claims
381 (myths M_{A-D}) to correction type will be counterbalanced, using all six possible combinations

² Should final sample size n drop below 352 in either delay condition (the sample size suggested by power analysis) despite this oversampling, we propose to resample a total of $(352 - n) \times 1.25$ participants in the immediate condition (to again account for zero-belief and other exclusions), and/or $(352 - n) \times 1.5$ participants in the immediate condition (to account for zero-belief and other exclusions, as well as drop-out due to delay) prior to analysis. These values might be adjusted based on the actual rejection and drop-out rates we observe.

382 (presentation versions V1-6 shown in Table 3); the presentation order of the eight claims (and
 383 thus the order of corrections/affirmations as well as narrative conditions) will be randomized.

384 Table 3

385 *Presentation Versions to be Used in Experiment 2*

	M _A	M _B	M _C	M _D
V1	NN	NN	N	N
V2	NN	N	NN	N
V3	NN	N	N	NN
V4	N	NN	NN	N
V5	N	NN	N	NN
V6	N	N	NN	NN

386 *Note.* Versions (V1-6) counterbalance the assignment of myths (M_{A-D}) to conditions (non-
 387 narrative correction, NN; narrative correction, N). Assignment of presentation version to
 388 participants will be randomized, with the constraint that a sixth of participants will receive
 389 each version.

390 Participants will rate their belief in each claim on a 0-10 Likert scale immediately
 391 after its initial presentation in the study phase (pre-explanation), and again at test (post-
 392 explanation). In addition to the second belief rating, the test will comprise three inference
 393 questions per claim, each requiring a rating of agreement with a statement on a 0-10 Likert
 394 scale. The inference questions are designed to measure claim-congruent inferential reasoning
 395 (e.g., “Patients with stomach ulcers should avoid any type of stress”). All questions are
 396 provided in the Appendix.

397 Administration of the survey will proceed as in Experiment 1, and the survey file will
 398 also be made available on the OSF. The experiment will take approximately 7 minutes.
 399 Participants in the immediate condition will be reimbursed approximately GBP0.88
 400 (US\$1.10) via Prolific; participants in the delayed condition will be reimbursed

401 approximately GBP0.35 (US\$0.44) for the study phase and GBP0.53 (US\$0.66) for the test
402 phase [times and amounts will be specified after pilot testing].

403 **Procedure.** The initial part of the survey will be identical to Experiment 1. In the
404 study phase, participants will be presented with all eight claims, rating their belief in each.
405 Each rating will be followed by an affirmation or a non-narrative or narrative correction.
406 Materials will again be presented for fixed minimum times (set at approx.150 ms per word).
407 In the (immediate or delayed) test phase, participants will first be presented with the short
408 questionnaire of three inference questions per claim. The order of questionnaires will be
409 randomized; the order of questions in each questionnaire will be fixed (see Appendix).
410 Subsequently, they will rate their belief in all claims for a second time. Following the test
411 phase, participants will be presented a “data use” question as in Experiment 1.

412 **Results**

413 Data analysis will be pre-registered on the OSF, where the raw data will also be made
414 available. Analysis will adhere to the following procedure. Exclusion criteria a, b, c, f, and g
415 (calculating response *SD* for each set of four test-phase questions, then calculating mean *SD*
416 across the eight sets) from Experiment 1 will be applied. In addition, we will exclude
417 participants who have any initial belief ratings of zero; given that 0-10 scales will be used to
418 measure claim belief, it can be assumed that corrections will reduce belief even in
419 participants with relatively low levels of initial claim belief (e.g., a reduction from 2 to 1 or 1
420 to 0), but naturally no reduction is possible from a zero rating. We will then calculate four
421 dependent variables relating to myth corrections and fact affirmations, respectively: mean
422 belief-rating change (belief-rating 2 – belief-rating 1) for the NN and N conditions, and mean
423 inference scores for the NN and N conditions.

424 We will run two-way mixed ANOVAs with factors condition (within-subjects) and
425 delay (between-subjects) on myth belief-change and inference scores. Irrespective of the

426 outcome of the omnibus ANOVA, we will test the core hypothesis with planned contrasts,
427 assessing the difference between NN and N conditions ($NN > N$) for both myth belief-change
428 and myth inference scores in each delay condition. To complement frequentist analysis (and
429 to potentially quantify evidence in favor of the null), we will run Bayesian *t*-tests comparing
430 NN and N myth-correction conditions for both dependent variables in both delay conditions.
431 For the sake of completeness, we will report the fact belief-change score and test it against
432 zero (*t*-test). To test for the impact of delay on affirmation effects, we will run one-way
433 ANOVAs on belief-change and inference scores.

434 Furthermore, to take into account initial belief levels more generally, we will
435 additionally run linear mixed-effects models, with presentation version and participant ID
436 (nested in presentation version) as random effects, and experimental condition, delay, their
437 interaction, and initial belief as fixed effects, predicting test-phase belief ratings and
438 inference scores. As with the ANOVAs, we will do this for the full 2×2 design, but also
439 separately for each delay condition, thus with only condition and initial belief as fixed
440 effects.

441 **Experiment 3** [highlighting just heading due to readability]

442 Narrative and non-narrative messages may evoke various processing differences that
443 could explain differences in correction effectiveness. To the extent that any narrative
444 advantage is due to reduced resistance to the corrective message, one could argue that it
445 should only emerge with corrections that challenge a person's worldview. Experiment 3 will
446 use more controversial real-world claims, where a correction can be expected to be
447 worldview-inconsistent for the majority of participants. This study therefore enables a more
448 focused test of underlying process, as well as an examination of the effect in a context of
449 practical significance. Specifically, two myths will be used that can be expected to resonate

450 particularly with more conservative participants. Hence, we will recruit participants who
451 identified as conservative on a political-worldview screener.

452 **Method**

453 Experiment 3 will again present claims encountered in the real world, including both
454 facts and myths, that will be followed by affirmations and corrections. Corrections will again
455 be either non-narrative (NN) or narrative (N), and the test will be immediate or delayed.
456 Thus, Experiment 3 will have a 2×2 mixed within-between design, with the within-subjects
457 factor of correction type (NN; N) and the between-subjects factor of test delay (immediate;
458 delayed). Fact-affirmation trials act as fillers outside of this design (even though basic
459 affirmation effects will be reported).

460 **Participants.** Sample size will be the same as in Experiment 2, but we will use a
461 sample of adult U.S. residents who indicated that they identify as politically conservative,
462 recruited via Prolific.³ Similar to Experiment 2, oversampling (again, by 10%) and potential
463 re-sampling will be applied to account for exclusions of participants with low initial-belief
464 ratings (see below). Participants who participated in Experiment 1 or 2 will not be allowed to
465 participate in Experiment 3.

466 **Materials.** Experiment 3 will use four claims (two myths; two facts). One myth is
467 “Humans are made to eat red meat; it should be part of every person’s diet.”⁴ The other is

³ We will recruit participants who responded with “conservative” to the Prolific pre-screener “Where would you place yourself along the political spectrum?” (conservative, moderate, liberal, other). Prolific indicates approx. 2,200 active participants in this population at the time of writing. If we are unable to obtain the full sample in this manner within a week, we will additionally open up the survey to participants who responded “moderate” (approx. 5,600 active participants).

⁴ Evidence for a link between political conservatism and meat consumption has been provided by Hodson and Earle (2018). Also see Gallup (2018).

468 “Children of homosexual parents have more mental health issues.”⁵ The non-narrative
469 corrections explain the evidence suggesting that the claim is false (e.g., that there is evidence
470 that eating red meat on a regular basis will shorten people’s lifespans and that replacing it
471 with other foods could lower mortality risk by 7 to 19%); the narrative correction contains the
472 same facts but is presented as a quote from someone to whom the claim is directly relevant
473 (e.g., a meat-lover explaining how their daughter pleaded with them to eat less red meat and
474 rotate in other foods). Again, a pilot study confirmed that the narrative corrections were
475 perceived as more story-like and vivid than the non-narrative correction, while being
476 relatively comparable on informativeness and comprehensibility dimensions (see Appendix
477 for details)⁶. Fact affirmations are of an expository nature similar to the non-narrative
478 corrections. All claims and explanations are provided in the Appendix. Each participant will
479 receive a NN and a N correction. Assignment of the two myths to correction type will be
480 counterbalanced, and presentation order randomized. Measures will be implemented as in
481 Experiment 2; an example inference question is “To maintain a healthy diet, people should
482 regularly consume red meat”). All questions are provided in the Appendix.

483 Administration of the survey will proceed as in Experiment 2, and the survey file will
484 also be made available on the OSF. The experiment will take approximately 5 minutes.

485 Participants in the immediate condition will be reimbursed approximately GBP0.63
486 (US\$0.79) via Prolific; participants in the delayed condition will be reimbursed

⁵ Evidence for a link between political conservatism and negative attitudes towards homosexual individuals has been provided by Haslam and Levy (2006); McLeod, Crawford, and Zechmeister (1999); Terrizzi, Shook, and Ventis (2010).

⁶ We note that the non-narrative corrections were rated as somewhat more informative; this was not surprising given that the narrative corrections contain some conversational elements. This makes our test more conservative: if we can demonstrate that narrative corrections are more effective than non-narrative ones, it would imply that the story factor can even overcome a slight informativeness deficit.

487 approximately GBP0.25 (US\$0.31) for the study phase and GBP0.38 (US\$0.47) for the test
488 phase [times and amounts will be specified after pilot testing].

489 **Procedure.** The procedure will be identical to Experiment 2.

490 **Results**

491 Data analysis will be pre-registered on the OSF, where the raw data will also be made
492 available. Analysis will follow the procedure of Experiment 2, with the exception that we
493 will exclude participants with any initial belief rating < 2 , or both initial belief ratings < 3 .⁷

494

⁷ We acknowledge that a person can have low belief in a claim they would like to believe based on their worldview, and thus it is possible that there would still be a narrative advantage in the lower belief range. However, in Experiment 3 we aimed for the corrections to be “uncomfortable,” which will only be true if the misinformation is worldview-consistent and initial belief is at least at a moderate level.

495 References

- 496 Allen, M., & Preiss, R. W. (1997) Comparing the persuasiveness of narrative and statistical
497 evidence using meta-analysis. *Communication Research Reports, 14*, 125-131.
498 doi:10.1080/08824099709388654
- 499 Bakker, M. H., Kerstholt, J. H., van Bommel, M., & Giebels, E. (2019) Decision-making
500 during a crisis: The interplay of narratives and statistical information before and after
501 crisis communication. *Journal of Risk Research, 22*, 1409-1424,
502 doi:10.1080/13669877.2018.1473464
- 503 Betsch, C., Renkewitz, F., & Haase, N. (2013). Effect of narrative reports about vaccine
504 adverse events and bias-awareness disclaimers on vaccine decisions: A simulation of
505 an online patient social network. *Medical Decision Making, 33*, 14-25.
506 doi:10.1177/0272989X12452342
- 507 Borgida, E., & Nisbett, R. E. (1977). The differential impact of abstract vs. concrete
508 information on decisions. *Journal of Applied Social Psychology, 7*, 258-271.
- 509 Bower, G. H., & Clark, M. C. (1969). Narrative stories as mediators for serial learning.
510 *Psychonomic Science, 14*, 181-182. doi:10.3758/BF03332778
- 511 Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science,*
512 *247*, 44-48.
- 513 Brewer, N. T., Chapman, G. B., Rothman, A. J., Leask, J., & Kempe, A. (2017). Increasing
514 vaccination: Putting psychological science into action. *Psychological Science in the*
515 *Public Interest, 18*, 149-207. doi:10.1177/1529100618760521
- 516 Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect
517 theory of stories. *Journal of Pragmatics, 6*, 473-486

- 518 Browning, E., & Hohenstein, J. (2015). The use of narrative to promote primary school
519 children's understanding of evolution. *Education 3-13*, 43, 530-547.
520 doi:10.1080/03004279.2013.837943
- 521 Bruner, J. (1986). Two modes of thought. In J. Bruner, *Actual minds, possible worlds* (pp. 11-
522 43). Cambridge, MA: Harvard University Press.
- 523 Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing
524 stories: A model of narrative comprehension and engagement. *Communication*
525 *Theory*, 18, 255-280. doi:10.1111/j.1468-2885.2008.00322.x
- 526 Caulfield, T., Marcon, A. R., Murdoch, B., Brown, J. M., Perrault, S. T., ... Hyde-Lay, R.
527 (2019). Health misinformation and the power of narrative messaging in the public
528 sphere. *Canadian Journal of Bioethics*, 2, 52-60. doi:10.7202/1060911ar
- 529 Chan, M.-P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A
530 meta-analysis of the psychological efficacy of messages countering misinformation.
531 *Psychological Science*, 28, 1531-1546. doi:10.1177/0956797617714579
- 532 Chang, C. (2009). "Being hooked" by editorial content: The implications for processing
533 narrative advertising. *Journal of Advertising*, 38, 21-34. doi:10.2753/JOA0091-
534 3367380102
- 535 Connor Desai, S., & Reimers, S. (2019). Comparing the use of open and closed questions for
536 Web-based measures of the continued-influence effect. *Behavior Research Methods*,
537 51, 1426-1440. doi:10.3758/s13428-018-1066-z
- 538 de Wit, J. B. F., Das, E., & Vet, R. (2008). What works best: Objective statistics or a personal
539 testimonial? An assessment of the persuasive effects of different types of message
540 evidence on risk perception. *Health Psychology*, 27, 110-115. doi:10.1037/0278-
541 6133.27.1.110

- 542 Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with
543 nonexpert audiences. *Proceedings of the National Academy of Sciences of the United*
544 *States of America, 111*, 13614-13620. doi:10.1073/pnas.1320645111
- 545 Dillard, A. J., Ferrer, R. A., & Welch, J. D. (2018). Associations between narrative
546 transportation, risk perception and behaviour intentions following narrative messages
547 about skin cancer. *Psychology and Health, 33*, 573-593.
548 doi:10.1080/08870446.2017.1380811
- 549 Dunlop, S. M., Wakefield, M., & Kashima, Y. (2010). Pathways to persuasion: Cognitive and
550 experiential responses to health-promoting mass media messages. *Communication*
551 *Research, 37*, 133-164. doi:10.1177/0093650209351912
- 552 Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of
553 misinformation: Helping or hindering its retraction? *Journal of Applied Research in*
554 *Memory and Cognition, 6*, 185-192. doi:10.1016/j.jarmac.2017.01.014
- 555 Ecker, U. K. H., Lewandowsky, S., Jayawardana, K., Mladenovic, A. (2019). Refutations of
556 equivocal claims: No evidence for an ironic effect of counterargument number.
557 *Journal of Applied Research in Memory and Cognition, 8*, 98-107.
558 doi:10.1016/j.jarmac.2018.07.005
- 559 Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false
560 information in memory: Manipulating the strength of misinformation encoding and its
561 retraction. *Psychonomic Bulletin & Review, 18*, 570–578. doi:10.3758/s13423-011-
562 0065-1
- 563 Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but
564 do not eliminate the continued influence of misinformation. *Memory & Cognition, 38*,
565 1087-1100. doi:10.3758/MC.38.8.1087

- 566 Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-
567 format refutational fact-checks. *British Journal of Psychology*, *111*, 36-54.
568 doi:10.1111/bjop.12383
- 569 Escalas, J. E. (2007). Self-referencing and persuasion: Narrative transportation versus
570 analytical elaboration. *Journal of Consumer Research*, *33*, 421-429.
571 doi:10.1086/510216
- 572 Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning
573 on people's health care decisions: Is a picture worth a thousand statistics? *Medical*
574 *Decision Making*, *25*, 398-405. doi:10.1177/0272989X05278931
- 575 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
576 power analysis program for the social, behavioral, and biomedical sciences. *Behavior*
577 *Research Methods*, *39*, 175-191. doi:10.3758/bf03193146
- 578 Gallup (2018). [https://news.gallup.com/poll/238328/snapshot-few-americans-vegetarian-](https://news.gallup.com/poll/238328/snapshot-few-americans-vegetarian-vegan.aspx)
579 [vegan.aspx](https://news.gallup.com/poll/238328/snapshot-few-americans-vegetarian-vegan.aspx)
- 580 Golke, S., Hagen, R., & Wittwer, J. (2019). Lost in narrative? The effect of informative
581 narratives on text comprehension and metacomprehension accuracy. *Learning and*
582 *Instruction*, *60*, 1-19. doi:10.1016/j.learninstruc.2018.11.003
- 583 Gordon, A., Brooks, J. C. W., Quadflieg, S., Ecker, U. K. H., & Lewandowsky, S.
584 (2017). Exploring the neural substrates of misinformation processing.
585 *Neuropsychologia*, *106*, 216-224. doi:10.1016/j.neuropsychologia.2017.10.003
- 586 Gordon, A., Quadflieg, S., Brooks, J. C. W., Ecker, U. K. H., & Lewandowsky, S. (2019).
587 Keeping track of 'alternative facts': The neural correlates of processing
588 misinformation corrections. *NeuroImage*, *193*, 46-56.
589 doi:10.1016/j.neuroimage.2019.03.014

- 590 Graesser, A. C., Hautt-Smith, K., Cohen A. D., & Pyles, L. D. (1980). Advanced outlines,
591 familiarity, and text genre on retention of prose. *Journal of Experimental Education*,
592 48, 281-290.
- 593 Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse
594 comprehension. *Topics in Cognitive Science*, 3, 371-398. doi:10.1111/j.1756-
595 8765.2010.01081.x
- 596 Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of
597 public narratives. *Journal of Personality and Social Psychology*, 79, 701-721.
598 doi:10.1037/0022-3514.79.5.701
- 599 Greene, K., & Brinn, L. S. (2003). Messages influencing college women's tanning bed use:
600 Statistical versus narrative evidence format and a self-assessment to increase
601 perceived susceptibility. *Journal of Health Communication*, 8, 443-461.
602 doi:10.1080/713852118
- 603 Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of
604 source credibility. *Journal of Applied Research in Memory and Cognition*, 2, 201-
605 209. doi:10.1016/j.jarmac.2013.10.001
- 606 Haase, N., Betsch, C., & Renkewitz, F. (2015). Source credibility and the biasing effect of
607 narrative information on the perception of vaccination risks. *Journal of Health*
608 *Communication*, 20, 920-929. doi:10.1080/10810730.2015.1018605
- 609 Hamby, A., Brinberg, D., & Jaccard, J. (2018). A conceptual framework of narrative
610 persuasion. *Journal of Media Psychology*, 30, 113-124. doi:10.1027/1864-
611 1105/a000187
- 612 Haslam, N., & Levy, S. R. (2006). Essentialist beliefs about homosexuality: Structure and
613 implications for prejudice. *Personality and Social Psychology Bulletin*, 32, 471-485.
614 doi:10.1177/0146167205276516

- 615 Hoaglin, D. C., and Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling.
616 *Journal of American Statistical Association*, 82, 1147-1149.
- 617 Hodson, G., & Earle, M. (2018). Conservatism predicts lapses from vegetarian/vegan diets to
618 meat consumption (through lower social justice concerns and social support).
619 *Appetite*, 120, 75-81. doi:10.1016/j.appet.2017.08.027
- 620 Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When
621 misinformation in memory affects later inferences. *Journal of Experimental*
622 *Psychology: Learning, Memory, and Cognition*, 20, 1420-1436. doi:10.1037/0278-
623 7393.20.6.1420
- 624 Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision
625 processes in refutation texts. *Discourse Processes*, 51, 374-397.
626 doi:10.1080/0163853X.2014.913961
- 627 Kim, E., Ratneshwar, S., & Thorson, E. (2017). Why narrative ads work: An integrated
628 process explanation. *Journal of Advertising*, 46, 283-296.
629 doi:10.1080/00913367.2016.1268984
- 630 Klassen, S. (2010). The relation of story structure to a model of conceptual change in science
631 learning. *Science and Education*, 19, 305-317. doi:10.1007/s11191-009-9212-8
- 632 Krakow, M. M., Yale, R. N., Jensen, J. D., Carcioppolo, N., & Ratcliff, C. L. (2018).
633 Comparing mediational pathways for narrative- and argument-based messages:
634 Believability, counterarguing, and emotional reaction. *Human Communication*
635 *Research*, 44, 299-321. doi:10.1093/hcr/hqy002
- 636 Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... &
637 Schudson, M. (2018). The science of fake news. *Science*, 359, 1094-1096.
638 doi:10.1126/science.aao2998

- 639 Lee, E., & Leets, L. (2002). Persuasive storytelling by hate groups online: Examining its
640 effects on adolescents. *American Behavioral Scientist*, *45*, 927-957.
641 doi:10.1177/0002764202045006003
- 642 Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation:
643 Understanding and coping with the “post-truth” era. *Journal of Applied Research in*
644 *Memory and Cognition*, *6*, 353-369. doi:10.1016/j.jarmac.2017.07.008
- 645 Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012).
646 Misinformation and its correction: Continued influence and successful debiasing.
647 *Psychological Science in the Public Interest*, *13*, 106-131.
648 doi:10.1177/1529100612451018
- 649 Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of
650 social experience. *Perspectives on Psychological Science*, *3*, 173-192.
- 651 Marsh, E. J., Butler, A. C., & Umanath, S. (2012). Using fictional sources in the classroom:
652 Applications from cognitive psychology. *Educational Psychology Review*, *24*, 449-
653 469. doi:10.1007/s10648-012-9204-0.
- 654 McLeod, A. C., Crawford, I., & Zechmeister, J. (1999). Heterosexual undergraduates’
655 attitudes toward gay fathers and their children. *Journal of Psychology & Human*
656 *Sexuality*, *11*, 43-62. doi:10.1300/J056v11n01_03
- 657 Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C., ... & Ecker, U. K.
658 H. (2019). Evaluation of a template for countering misinformation—Real-world
659 Autism treatment myth debunking. *PLOS ONE*, *14*, e0210746.
660 doi:10.1371/journal.pone.0210746
- 661 Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative
662 platforms for crowdsourcing behavioral research. *Journal of Experimental Social*
663 *Psychology*, *70*, 153-163. doi:10.1016/j.jesp.2017.01.006

- 664 Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory
665 structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and*
666 *Cognition, 14*, 521-533.
- 667 Rapp, D. N., & Salovich, N. A. (2018). Can't we just disregard fake news? The consequences
668 of exposure to inaccurate information. *Policy Insights from the Behavioral and Brain*
669 *Sciences, 5*, 232-239. doi:10.1177/2372732218785193
- 670 Ratcliff, C. L., & Sun, Y. (2020). Overcoming resistance through narratives: Findings from a
671 meta-analytic review. *Human Communication Research*. doi:10.1093/hcr/hqz017
- 672 Reinhart, A. M. (2006). *Comparing the Persuasive Effects of Narrative versus Statistical*
673 *Messages: A Meta-analytic Review*. Buffalo, NY State University of New York at
674 Buffalo. Available from <https://search.proquest.com/docview/304937594>
- 675 Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly
676 stated misinformation in news reports. *Journal of Experimental Psychology:*
677 *Learning, Memory, and Cognition, 42*, 62-74. doi:10.1037/xlm0000155
- 678 Romero, F., Paris, S. G., & Brem, S. K. (2005). Children's comprehension and local-to-
679 global recall of narrative and expository texts. *Current Issues in Education, 8*, 1-20.
- 680 Sangalang, A., Ophir, Y., & Cappella, J. N. (2019). The potential for narrative correctives to
681 combat misinformation. *Journal of Communication, 69*, 298-319.
682 doi:10.1093/joc/jqz014
- 683 Shaffer, V. A., Focella, E. S., Hathaway, A., Scherer, L. D., & Zikmund-Fisher, B. J. (2018).
684 On the usefulness of narratives: An interdisciplinary review and theoretical model.
685 *Annals of Behavioral Medicine, 52*, 429-442. doi:10.1093/abm/kax008
- 686 Shelby, A., & Ernst, K. (2013). Story and science: How providers and parents can utilize
687 storytelling to combat anti-vaccine misinformation. *Human Vaccines and*
688 *Immunotherapeutics, 9*, 1795-1801. doi:10.4161/hv.24828

- 689 Shen, F., Ahern, L., & Baker, M. (2014). Stories that Count: Influence of News Narratives on
690 Issue Attitudes. *Journalism & Mass Communication Quarterly*, *91*, 98–117.
691 doi:0.1177/1077699013514414
- 692 Shen, F., Sheer, V. C., & Li, R. (2015). Impact of narratives on persuasion in health
693 communication: A meta-analysis. *Journal of Advertising*, *44*, 105-113.
694 doi:10.1080/00913367.2015.1018467
- 695 Slater, M. D., & Rouner, D. (1996). Value-affirmative and value-protective processing of
696 alcohol education messages that include statistical evidence or anecdotes.
697 *Communication Research*, *23*, 210–235. doi:10.1177/009365096023002003
- 698 Southwell, B. G., & Thorson, E. A. (2015). The prevalence, consequence, and remedy of
699 misinformation in mass media systems. *Journal of Communication*, *65*, 589-595.
700 doi:10.1111/jcom.12168
- 701 Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting
702 inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and*
703 *Cognition*, *43*, 1948-1961. doi:10.1037/xlm0000422
- 704 Terrizzi Jr, J. A., Shook, N. J., & Ventis, W. L. (2010). Disgust: A predictor of social
705 conservatism and prejudicial attitudes toward homosexuals. *Personality and*
706 *Individual Differences*, *49*, 587-592. doi:10.1016/j.paid.2010.05.024
- 707 Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation.
708 *Political Communication*, *33*, 460-480. doi:10.1080/10584609.2015.1102187
- 709 van Krieken, K., & Sanders, J. (2019). What is narrative journalism? A systematic review
710 and an empirical agenda. *Journalism*. doi:10.1177/1464884919862056
- 711 Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A
712 big data analysis of the online media landscape from 2014 to 2016. *New Media and*
713 *Society*, *20*, 2028-2049. doi:10.1177/1461444817712086

- 714 Vraga, E. K., Bode, L., & Tully, M. (2020). Creating news literacy messages to enhance
715 expert corrections of misinformation on Twitter. *Communication Research*.
716 doi:10.1177/0093650219898094
- 717 Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued
718 influence of misinformation in the face of correction: How powerful is it, why does it
719 happen, and how to stop it? *Communication Research*, 47, 155-177.
720 doi:10.1177/0093650219854600
- 721 Wolfe, M. B. W., & Mienko, J. A. (2007). Learning and memory of factual content from
722 narrative and expository text. *British Journal of Educational Psychology*, 77, 541-
723 564. doi:10.1348/000709906X143902
- 724 Wolfe, M. B. W., & Woodwyk, J. M. (2010). Processing and memory of information
725 presented in narrative or expository texts. *British Journal of Educational Psychology*,
726 80, 341-362. doi:10.1348/000709910X485700
- 727 Zabucky, K. M., & Moore, D. (1999). Influence of text genre on adults' monitoring of
728 understanding and recall. *Educational Gerontology*, 25, 691-710.
729 doi:10.1080/036012799267440
- 730 Zebregs, S., van den Putte, B., de Graaf, A., Lammers, J., & Neijens, P. (2015). The effects
731 of narrative versus non-narrative information in school health education about alcohol
732 drinking for low educated adolescents. *BMC Public Health*, 15, article 1085.
733 doi:10.1186/s12889-015-2425-7
- 734 Zebregs, S., van den Putte, B., Neijens, P., & de Graaf, A. (2015). The differential impact of
735 statistical and narrative evidence on beliefs, attitude, and intention: A meta-analysis.
736 *Health Communication*, 30, 282-289. doi:10.1080/10410236.2013.842528

737 **Declarations**

738 Availability of data and materials: All data and survey files (which include the materials) will
739 be made available on the Open Science Framework website. For convenience, all materials
740 are additionally provided in the Appendix.

741 Competing interests: The authors declare no competing interests.

742 Funding: The research was supported by the Australian Research Council under grant
743 DP160103596, awarded to the first author.

744 Authors' contributions: UE and AH conceptualized the study; UE, LB, and AH created the
745 materials and designed the experiments; UE wrote the initial manuscript draft; AH and LB
746 contributed to the writing.

747 Acknowledgements: We thank Charles Hanich for research assistance, and Shawn Callahan
748 for early discussions regarding the “story factor.”

749 **Appendix** [lots of changes; just highlighting heading due to readability]

750 **Experiment 1**

751 **Event reports.** On average, the non-narrative corrections contained in the event
752 reports had 111 words, with a Flesch reading ease (FRE) score of 49.23 and a Flesch-Kincaid
753 grade level (FKGL) of 11.6. Narrative corrections had 111.25 words, with a reading ease
754 score of 43.05 and a grade level of 11.73.

755 **Report A: Wildfire.** (356-359 words)

756 *Article 1.*

757 VANCOUVER—Firefighters in British Columbia have been battling a wildfire that
758 raged out of control in the state’s South-East overnight. The fire came dangerously close to
759 homes in the town of Cranbrook, but it is believed that no damage was caused to property.
760 David Karle of the BC Wildfire Service indicated that authorities were looking into the cause
761 of the fire, with early evidence suggesting that the fire had been deliberately lit. Despite
762 extensive campaigns, arson remains a significant problem in the region, and a leading cause
763 of wildfires globally. Emergency services were still working tirelessly this morning to
764 extinguish the flames, but were confident that the location of the remaining fire was unlikely
765 to pose any further threat to local communities. (Word Count [WC] = 121; Flesch Reading
766 Ease [FRE] = 40.3; Flesch-Kincaid Grade Level [FKGL] = 13.6)

767 *Article 2.*

768 VANCOUVER—After working throughout the day, firefighters have managed to
769 bring a wildfire in the South-East of British Columbia under control. There have been no
770 reported casualties or damage to property, with most land damage occurring in rural fringe
771 areas and nearby forest reserves. The suspected burn area is estimated to be roughly 10,000
772 hectares. (WC = 54; FRE = 36.5; FKGL = 12.6)

773 *Non-narrative correction:* It is now clear that the fire was caused by a power line
774 from a fallen power pole. The power pole was in a condition that was substantially weakened
775 due to general rot and severe damage caused by the growth of a colony of termites. The cause
776 of the fire was announced earlier today by Cranbrook Fire and Emergency Services based on
777 new evidence that emerged from a detailed additional investigation of the ignition zone (the
778 area where the fire had started). This investigation took place shortly after the fire in that area
779 had been extinguished. A power line from the broken pole had made contact with the ground
780 and started the fire, after the power pole had fallen. (WC = 119; FRE = 58.2; FKGL = 11)

781 *Narrative correction:* An additional investigation by Fire Chief Warren Linnell
782 uncovered the true fire cause: a power line from a fallen power pole. Linnell, a 20-year
783 veteran of the Cranbrook Fire and Emergency Services, was skeptical of initial claims about
784 the fire's cause: "I've seen a lot of fires, and determining the cause of any fire always
785 requires thorough investigation." Deciding to explore further, Linnell waded through the
786 ignition zone and discovered a power pole that had snapped. Peering closely, he noticed rot
787 and severe termite damage throughout the pole. Then he noticed the broken power line.
788 When he saw that it had melted on the ground, he concluded that the broken power line
789 ignited leaf litter around the broken pole, starting the fire. (WC = 122, 1.03 ratio; FRE = 51.9;
790 FKGL = 11.1)

791 Casey Haas, a resident of Cranbrook, expressed her relief that no one had been
792 injured by the fire, saying she felt lucky that they had avoided disaster, and that her beloved
793 ponies Tom and Jerry had survived unharmed. Even so, she felt it was important for residents
794 of the community to work together to ensure they are prepared for potential future disasters.
795 (WC = 62; FRE = 43; FKGL = 14.9)

796 ***Report B: Spike in seizures.*** (347-348 words)

797 *Article 1.*

798 BRISBANE—An unprecedented spike in seizures leading to hospital admissions has
799 been reported in North Queensland (Australia). Over the past month, 17 children were
800 assessed at Townsville Hospital, with roughly half being admitted for observation and in-
801 patient treatment. According to the hospital, these are unusual numbers for the regional town,
802 which has a population of 180,000. The spike in seizures has been linked to the introduction
803 of a new compound vaccine, offered to children in the region, which combines the polio and
804 chicken pox (varicella) vaccines. It was hoped the new vaccine would increase the
805 immunization rate against chicken pox, as part of an active push to completely eradicate the
806 disease in Australia. However, seizures can be a side effect of vaccination, and administration
807 of the new vaccine has been suspended. At this stage, none of the seizures have been life-
808 threatening, although three children remain in hospital under close surveillance. (WC = 149;
809 FRE = 36.4; FKGL = 13.4)

810 *Article 2.*

811 BRISBANE—All children affected by a recent spike in seizures in North Queensland
812 have now returned home to their families. While several new cases have been reported, none
813 have required hospitalization. (WC = 30; FRE = 50.6; FKGL = 9.9)

814 *Non-narrative correction:* The spike in seizures recently seen at a North-East
815 Australian hospital has now been linked to the Kuta virus, a virus most commonly seen in
816 rural parts of South East Asia. The increase in seizures occurred at the same time as an
817 increase in the level of mosquito activity in the region. Evidence of the Kuta virus was
818 present in all examined blood samples tested. The virus is known to cause seizures in
819 children, although it is not usually present in Australia. According to experts, the unusually

820 high temperatures seen in the region over the past months could have contributed to the
821 spread of the virus. (WC = 106; FRE = 52; FKGL = 11.2)

822 *Narrative correction:* Health authorities have now linked the spike in seizures to the
823 Kuta virus. Dr. Katherine Hopkins from Townsville Hospital noticed a report about high
824 mosquito activity in the region. She became curious whether there was any connection to the
825 seizures. Running additional tests on patients' blood, she found evidence of the Kuta virus,
826 which is known to cause seizures, in all samples. "I was surprised at first, because the virus is
827 usually not present in Australia" Dr. Hopkins said, "so I called my colleague, who is an
828 epidemiologist." The epidemiologist, Dr. David Chang, confirmed that the unusually high
829 temperatures likely allowed the virus to spread. (WC = 105, .99 ratio; FRE = 44.8; FKGL =
830 11.3)

831 Locals Daniel and Tiarne Corner explained that their 5-year old son Toby had just
832 been released from hospital, and expressed their gratitude to the hospital's staff: "It was so
833 scary when the seizures started, out of the blue. The nurses and doctors took such good care
834 of us; they are amazing. We are so glad it's over, and can't wait to go home." (WC = 64;
835 FRE = 71.5; FKGL = 8.5)

836 ***Report C: Plane crash.*** (362 words)

837 *Article 1.*

838 MANCHESTER—A small business jet en route to the German town of Rostock
839 crashed on Monday morning, minutes after take-off from Manchester Airport. The two-
840 engine Zephyr ZX crashed in a field near the town of Failsworth, killing all eleven people –
841 eight passengers and three crew – on board. The passengers are believed to be the executives
842 of Manchester-based technology start-up 3RTec. Based on initial evidence and witness
843 reports, the plane stalled after hitting a drone that was flying in the area. Despite regulations,
844 drones flying near airports have been identified as a significant but difficult-to-eliminate

845 threat to air travel safety. Witnesses described that they heard a loud explosion and saw a
846 plume of black smoke when the aircraft hit the ground. “A few hundred yards further down,
847 and it would have struck my house,” local resident Liesel Mason noted. “It was frightening. I
848 really feel for the victims, it must have been terrifying.” (WC = 151; FRE = 56.4,
849 FGKL = 9.5)

850 *Article 2.*

851 MANCHESTER—The Manchester business community is still in shock after
852 Monday’s plane crash, which killed eleven people, including the entire executive team of
853 local tech company 3RTec. Alice Crane, the company’s HR manager, explained that staff are
854 absolutely devastated. “There are no words,” Ms. Crane stated. “We just don’t feel like this is
855 real.” (WC = 54; FRE = 54.5; FKGL = 8.9)

856 *Non-narrative correction:* The plane crash near Manchester has now been ruled the
857 result of a technical failure of the machinery inside the plane. In a statement put out by the
858 UK’s Civil Aviation Authority, it was revealed that the plane contained a manufacturing flaw
859 specific to Zephyr ZX aircraft manufactured recently in the company’s Aberdeen plant. One
860 of the engines’ thrust reversers accidentally deployed shortly after take-off at an altitude of
861 3,000 ft. A thrust reverser is part of an engine; it changes the direction of air flow and is used
862 by pilots to slow a plane down during or after landing. Deployment of the thrust reverser
863 caused the plane to bank to the right and enter a high-speed dive. (WC = 118; FRE = 49.9;
864 FKGL = 11.1)

865 *Narrative correction:* An additional investigation has revealed that the devastating
866 plane crash near Manchester was caused by a technical failure. Investigator Sharon Williams
867 from the UK’s Civil Aviation Authority said: “I became suspicious after learning that the
868 aircraft had been manufactured in Zephyr’s Aberdeen plant. A concerned Zephyr employee
869 previously confided in me that a manufacturing flaw had been detected in this plant. The

870 company was trying to downplay it.” Williams’ team investigated and found evidence that
871 one of the engines’ thrust reverser had malfunctioned. Williams explained: “A thrust reverser
872 acts like a brake. This one deployed shortly after take-off at an altitude of 3,000 ft. This
873 caused the plane to bank to the right and enter a high-speed dive.” (WC = 118, 1.00 ratio;
874 FRE = 41.3; FKGL = 11.1)

875 While this was the third fatal aviation accident in the UK in the past month, flying
876 continues to be a very safe mode of transportation. The overwhelming majority of aviation
877 fatalities involve small, private airplanes, and not large commercial airliners. (WC = 40;
878 FRE = 36.3; FKGL = 13.1)

879 ***Report D: Salmonella outbreak.*** (318-320 words)

880 *Article 1.*

881 ALBUQUERQUE—More than a hundred people have fallen ill—and a dozen have
882 been hospitalized—after a salmonella outbreak in New Mexico. Victims had dined at several
883 restaurants in the greater Albuquerque area. The outbreak has been traced back to a local
884 food factory, where it is believed the failure of sterilization equipment is to blame for the
885 food poisoning. The factory, which produces mayonnaise and other condiments for local
886 restaurants, has stopped production and recalled products. An estimated 1.2 million
887 salmonella cases occur in the U.S. annually. While many cases are related to food hygiene in
888 the home, larger outbreaks are often linked to technical issues during food production. While
889 the current outbreak in New Mexico is significant, the largest outbreak in U.S. history in
890 2008 saw more than 1,000 people fall ill in Texas and several other states. (WC = 139;
891 FRE = 39.3; FKGL = 12.6)

892 *Article 2.*

893 ALBUQUERQUE—The total number of victims who have fallen ill in the New
894 Mexico salmonella outbreak has risen to 137. While most victims are recovering well, a 79-

895 year-old North Valley man had to be admitted into intensive care and is in a critical
896 condition. (WC = 43; FRE = 42.2; FKGL = 12.8)

897 *Non-narrative correction:* The outbreak in the Albuquerque processing plant has now
898 been linked to intentional food contamination. This means that food had become corrupted
899 with another substance during processing. The sterilization equipment at the factory was
900 found to work adequately and reliably heat all foods to 170 degrees Fahrenheit, which is a
901 high enough temperature to destroy any biological contaminants. However, a review of the
902 CCTV footage from the factory showed a male employee in the packaging department of the
903 factory tampering with a product as it was bottled. It appears the employee's motive to do so
904 was revenge for poor treatment of staff. (WC = 102; FRE = 36.8; FKGL = 13.1)

905 *Narrative correction:* An additional investigation by inspector Stephanie Hill from
906 the Food Safety Authority has uncovered that the outbreak was the result of intentional food
907 contamination. During her inspection of the Albuquerque factory, Hill found that the
908 sterilization equipment worked adequately, heating foods to the required 170 degrees
909 Fahrenheit. "This seemed suspicious, so I decided to review the CCTV footage," Hill
910 described. What she found shocked her: the tapes showed an employee contaminating a
911 product as it was bottled. When confronted, the employee exploded with rage, describing his
912 desire to ruin the company as revenge for his boss' cruel treatment of staff. (WC = 100, .98
913 ratio; FRE = 34.2; FKGL = 13.4)

914 All restaurants remain open for business and are preparing for the upcoming
915 Albuquerque Restaurant Week, an annual event that celebrates the local food scene. Curious
916 patrons can expect fiery and creative meals, with many special offers. (WC = 36;
917 FRE = 38.1; FKGL = 12.4)

918 **Test questionnaires.**919 ***Report A.***

- 920 1. The fire came close to the town of Cranbrook / Kimberley / Lumberton / Bull
921 River
- 922 2. “Devastating wildfire intentionally lit” would be an appropriate headline for the
923 report. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 924 3. Malicious intent contributed to the fire. 0 (Strongly Disagree) – 10 (Strongly
925 Agree)
- 926 4. The person responsible for the wildfire should be identified and charged. 0
927 (Strongly Disagree) – 10 (Strongly Agree)
- 928 5. The local government should invest in measures to prevent arson. 0 (Strongly
929 Disagree) – 10 (Strongly Agree)
- 930 6. Local residents should be particularly vigilant against potential arsonists. 0
931 (Strongly Disagree) – 10 (Strongly Agree)
- 932 7. What do you think caused the wildfire? Arson / Lightning / Power line / None of
933 the above

934 ***Report B.***

- 935 1. Which Australian state was affected by the seizures? Queensland / New South
936 Wales / Victoria / Tasmania
- 937 2. “New vaccine leads to seizures, hospitalizations” would be an appropriate headline
938 for this report. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 939 3. Insufficient safety tests by pharma companies contributed to the spike in seizures.
940 0 (Strongly Disagree) – 10 (Strongly Agree)
- 941 4. There should be repercussions for the person who approved the vaccine trial. 0
942 (Strongly Disagree) – 10 (Strongly Agree)

- 943 5. The government should implement more stringent safety tests of vaccines to
944 prevent such incidents in the future. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 945 6. Based on what happened, parents should be particularly skeptical of newly
946 introduced compound vaccines. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 947 7. What do you think caused the spike in seizures? Vaccine / Lead poisoning / Virus /
948 None of the above

949 **Report C.**

- 950 1. How many people were killed in the crash? 11 / 16 / 20 / 25
- 951 2. “Drone downs plane, killing all aboard” would have been an appropriate headline
952 for the report. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 953 3. A drone collision contributed to the plane crash. 0 (Strongly Disagree) – 10
954 (Strongly Agree)
- 955 4. The person flying the drone should be identified and charged with manslaughter. 0
956 (Strongly Disagree) – 10 (Strongly Agree)
- 957 5. Following the incident, policies regarding drone usage around airports should be
958 reviewed. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 959 6. Based on this event, drone-detection hardware should be made mandatory on all
960 aircraft. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 961 7. What do you think caused the plane crash? Drone strike / Bad weather / Technical
962 fault / None of the above

963 **Report D.**

- 964 1. How many people fell ill during the New Mexico salmonella outbreak? About 50 /
965 More than 100 / More than 250 / More than 500
- 966 2. “Equipment failure causes salmonella outbreak” would be an appropriate headline
967 for this report. 0 (Strongly Disagree) – 10 (Strongly Agree)

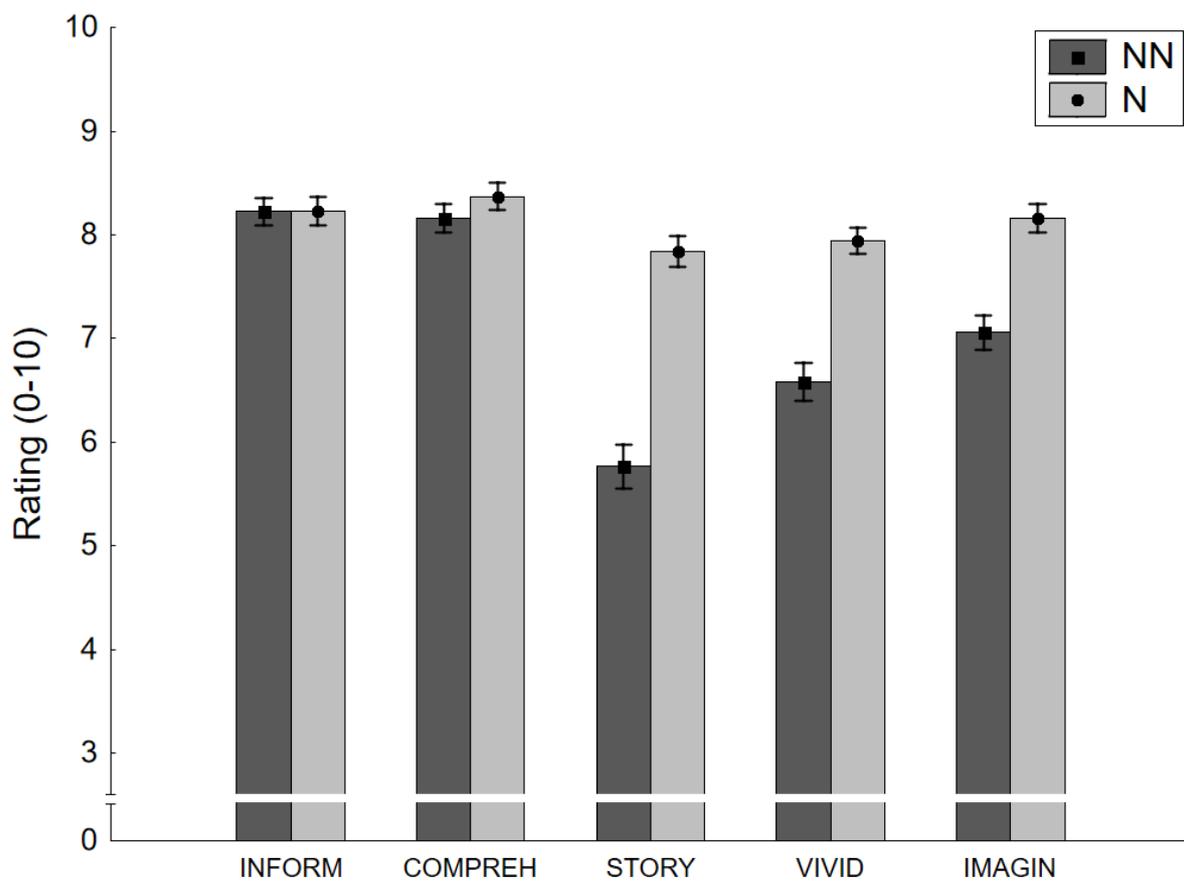
- 968 3. A technical issue contributed to the outbreak. 0 (Strongly Disagree) – 10 (Strongly
969 Agree)
- 970 4. There should be repercussions for the factory staff responsible for equipment
971 maintenance and testing. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 972 5. Based on this incident, food factories should implement more stringent safety tests
973 of sterilization equipment to prevent such incidents in the future. 0 (Strongly
974 Disagree) – 10 (Strongly Agree)
- 975 6. The affected company should consider investing in more reliable sterilization
976 equipment. 0 (Strongly Disagree) – 10 (Strongly Agree)
- 977 7. What do you think caused the outbreak? Equipment failure / Restaurant hygiene /
978 Intentional tampering / None of the above

979 **Pilot Study.** One hundred U.S.-based MTurk workers (min. 5,000 so-called Human
980 Intelligence Tasks [HITs] completed with 98%+ approval rate) were recruited to rate the non-
981 narrative and narrative corrections of all event reports. One participant was excluded due to
982 uniform responding ($SD = 0$), leaving $N = 99$ participants ($M_{age} = 40.44$ years; age range 20-
983 79; 51 males, 46 females, 2 of unspecified gender).

984 All reports were presented in randomized order. For each report, participants read
985 both corrections, also in randomized order. They were asked to rate each correction on
986 informativeness (“How informative is the correction?”), comprehensibility (“How easy to
987 understand is the correction?”), story-ness (“How story-like is the correction?”), vividness
988 (“How vivid is the correction?”), and imaginability (“While you were reading the correction,
989 how easily could you picture the events taking place?”), all on 0 (not at all) – 10 (very much)
990 scales.

991 Results are summarized in Figure A1. There was a large difference in story-ness
992 between non-narrative and narrative corrections, with substantial differences also on

993 vividness and imaginability dimensions. There was no difference between conditions on
994 comprehensibility, and only a small difference on informativeness, which was to be expected
995 given the narrative correction was designed to provide the same relevant corrective
996 information plus the story “wrapper.” We concluded that our manipulation was implemented
997 successfully.



998

999 *Figure A1.* Ratings of non-narrative (NN) and narrative (N) event-report corrections on
1000 informativeness (INFORM), comprehensibility (COMPREH), story-ness (STORY),
1001 vividness (VIVID), and imaginability (IMAGIN) in the Experiment 1 Pilot. Error bars give
1002 within-subjects standard error of the mean.

1003 **Experiment 2**

1004 **Claims and explanations.** On average, the non-narrative corrections had 101 words, with FRE = 40.83 and FKGL = 12.48; narrative
 1005 corrections had 111.5 words, with FRE = 42.15 and FKGL = 12.1 (see Table A1). Affirmations had on average 87.5 words, with FRE = 52.9
 1006 and FKGL = 10.9 (see Table A2).

Table A1

Myths and their Corresponding Non-Narrative and Narrative Corrections

Item number	Items	Non-Narrative Correction	Narrative Correction
Myth - 1	Gastritis and stomach ulcers are caused by excessive stress.	There is now strong evidence that gastritis and stomach ulcers are caused by the bacterium <i>Helicobacter pylori</i> . Scientists Barry Marshall and Robin Warren are credited with the discovery of this association, which was viewed by the broader scientific community as novel. A Nobel Prize was awarded to Marshall and Warren because of this discovery. A consequence of this discovery is that antibiotics can be used to treat these conditions. (WC = 69; FRE = 37.2; FKGL = 12.3)	Scientist Barry Marshall discovered that gastritis and stomach ulcers are caused by the bacterium <i>Helicobacter pylori</i> . At first, he was ridiculed by colleagues for his proposal. Frustrated, he intentionally drank a broth contaminated with the bacterium to prove that it caused disease. Soon after, Marshall developed gastritis as a result, and then successfully used antibiotics to treat himself. There is now strong evidence for the link, and the discovery earned Marshall and his colleague Robin Warren a Nobel Prize. (WC = 79, ratio 1.14; FRE = 39.8; FKGL = 11.6)
Myth - 2	Women talk more than men.	Numerous studies have converged on the conclusion that females do not talk more than males. Based on studies recording regular speech	Females do not talk more than males. Professor James Pennebaker of the University of Texas was leisurely reading a magazine, when he encountered a claim that

fragments from volunteers, it has been estimated that both men and women say around 16,000 words a day. This type of research is often done by using a digital device that records 30 seconds of sound every 12.5 minutes over long periods of time. From this, the total number of words spoken per day can be extrapolated with satisfactory accuracy. Results indicate that there are outliers of both genders, meaning there are some people who speak much more and others who speak much less than the average.

(WC = 108; FRE = 47.0; FKGL = 12.0)

Myth - 3 Cracking your knuckles leads to arthritis.

There is no correlation between cracking one’s knuckles and the development of arthritis, despite prevalent belief about the relationship. For example, one study demonstrated that frequent knuckle cracking did not lead to the development of arthritis in the hand, even in knuckles cracked up to 36,500 times over a time span of 50 years. The study, titled “Does knuckle cracking lead to arthritis of the fingers?”, was published in the scientific journal *Arthritis and Rheumatism*. Dr. Donald Unger, the sole author of the article, received the 2009 Ig-Nobel Prize for the work. This is a prize which is awarded for research that makes you laugh, then think.

(WC = 107; FRE = 43.5; FKGL = 12.4)

jolted his mind to action: that women are “chatterboxes” who speak three times as much as men. Dubious of the claim, he decided to test its validity. To do so, Pennebaker recorded the speech of hundreds of volunteers, who wore digital devices that recorded 30 seconds of sound every 12.5 minutes. After painstaking analysis, he found that both men and women say around 16,000 words a day, a finding that has been replicated in numerous other studies. Amusingly, the most talkative person in the study was a man, racking up 47,000 words a day!

(WC = 120, ratio 1.11; FRE = 41.3; FKGL = 12.4)

There is no correlation between cracking one’s knuckles and the development of arthritis – as was most convincingly shown by Dr. Donald Unger. When Unger was a child, his parents scolded him every time he cracked his knuckles, warning him, “you’re going to develop arthritis!” Curious about whether this was true, he began cracking his left-hand knuckles daily, while never cracking his right hand. After 50 years – cracking his left-hand knuckles about 36,500 times in the process – Unger had not developed arthritis in either hand. He published the finding in the scientific journal *Arthritis and Rheumatism*. For his work, Unger received the 2009 Ig-Nobel Prize, awarded for research that makes you laugh, then think.

(WC = 113, ratio 1.06; FRE = 45.4; FKGL = 11.5)

<p>Myth - 4</p>	<p>Delayed-onset muscle soreness is caused by build-up of lactic acid.</p>	<p>Lactic acid produced in muscles during strenuous exercise does not cause muscle soreness a day or two after exercise. Scientific evidence shows that strenuous exercise that a person is used to partaking in does not produce delayed-onset muscle soreness. Relatively easy exercise that a person is not used to, on the other hand, does produce muscle soreness. This occurs despite the fact that the relatively easier exercise often results in a lower level of lactic acid production, compared to the more strenuous but familiar exercise. Thus, delayed-onset muscle soreness is not the result of lactic acid build-up. Rather, the soreness is caused by micro-tears to muscle fibers, which are more likely to occur when engaging in new types of exercise.</p> <p>(WC = 120; FRE = 35.6; FKGL = 13.2)</p>	<p>Lactic acid produced in muscles during strenuous exercise does not cause muscle soreness. Sport scientist James Schwane, an avid runner, questioned the often-cited relationship between lactic acid and delayed-onset muscle soreness based on his own experience, and decided to test it. Schwane got participants to either run on a flat surface (which was strenuous, but involved movements the runners were used to), or downhill (which was easier, but less similar to runners' usual movements). He discovered that running downhill produced less lactic acid but caused more soreness than running on a flat surface. This led him to conclude that delayed-onset muscle soreness is not linked to lactic acid. Rather, he concluded that the soreness is caused by micro-tears to muscle fibers, which are more likely to occur when engaging in new types of exercise.</p> <p>(WC = 134, ratio 1.12; FRE = 42.1; FKGL = 12.9)</p>
-----------------	--	--	--

1007 *Note.* WC = Word Count; FRE = Flesch Reading Ease; FKGL = Flesch-Kincaid Grade Level.

Table A2

Facts and their Corresponding Affirmations

Item	Claim	Affirmation
Fact A	Stomach acid can dissolve razor blades.	<p>A study in 1997 confirmed that our gastric juices can indeed dissolve razor blades, albeit slowly. This is possible due to simple chemistry: The lining of our stomach secretes hydrochloric acid, which dissolves many metals. Razor blades are made of steel, which is an alloy of iron, and are therefore readily dissolved by hydrochloric acid. The study concluded that, if you were to swallow a razor blade, the best time for surgery would be 15 hours or so after ingestion. This is because by this time the blade will have become fragile and could be broken and removed in a piecemeal fashion.</p> <p>(WC = 102; FRE = 53.4; FKGL = 10.8)</p>
Fact B	It is not safe to talk on landline telephones when there is a thunderstorm.	<p>It is, in fact, not safe to talk on a landline during a thunderstorm. The current in a lightning bolt can exceed 100,000 volts. Electrical wires are good transmitters of electricity, so when lightning strikes a house, it has the potential to move through the interconnected cables. Usually, the energy is simply absorbed into the ground, but it is possible for the current to travel through the landline’s cables and shock the person on the end of the phone line.</p> <p>(WC = 80; FRE = 55.7; FKGL = 10.5)</p>
Fact C	Dogs can smell cancer.	<p>Dogs perform better than state-of-the-art screening tests at detecting people with lung and breast cancer. This has been tested in a scientific setting. Cancer patients have traces of chemicals (like alkanes and benzene derivatives) in their breath, which dogs can detect in concentrations as small as a few parts per trillion. A study at the University of California showed that dogs correctly detected 99% of lung cancer breath samples and made a mistake with only 1% of samples from healthy controls.</p> <p>(WC = 81; FRE = 48.4; FKGL = 11.5)</p>

Fact D We are taller in the morning than in the evening. We are taller in the mornings than the evenings due to the compression of our spine over the course of the day. When you are standing or sitting, there is pressure on the intervertebral discs, which causes water to be expelled. At night, when the spine is horizontal, water is reabsorbed by the disks. In 1935, DePukey measured 1,216 participants between 5 and 90 years old, and found the average person was more than half an inch shorter in the evening than they were in the morning.
(WC = 87; FRE = 53.2; FKGL = 10.9)

1008 *Note.* WC = Word Count; FRE = Flesch Reading Ease; FKGL = Flesch-Kincaid Grade Level.

1009 **Test questionnaire.**

Table A3

Claims and Corresponding Inference Questions

Item	Claim	Inference Question 1	Inference Question 2	Inference Question 3
Myth A	Gastritis and stomach ulcers are caused by excessive stress.	Patients with stomach ulcers should avoid any type of stress.	How effective do you think relaxation techniques are in preventing gastritis?	How likely is it that you would advise a friend or family member with stomach pains to reduce stress so they do not develop a stomach ulcer?
Myth B	Women talk more than men.	At any given time, a woman is more likely to be speaking compared to a man.	In general, jobs that require a lot of talking are a more natural fit for women.	If you met a new male-female couple, how likely is it that the woman would talk more than the man?
Myth C	Cracking your knuckles leads to arthritis.	People with a family history of arthritis should avoid cracking their knuckles.	Children should be taught not to crack their knuckles in order to	How likely is it that you would advise a friend or family member

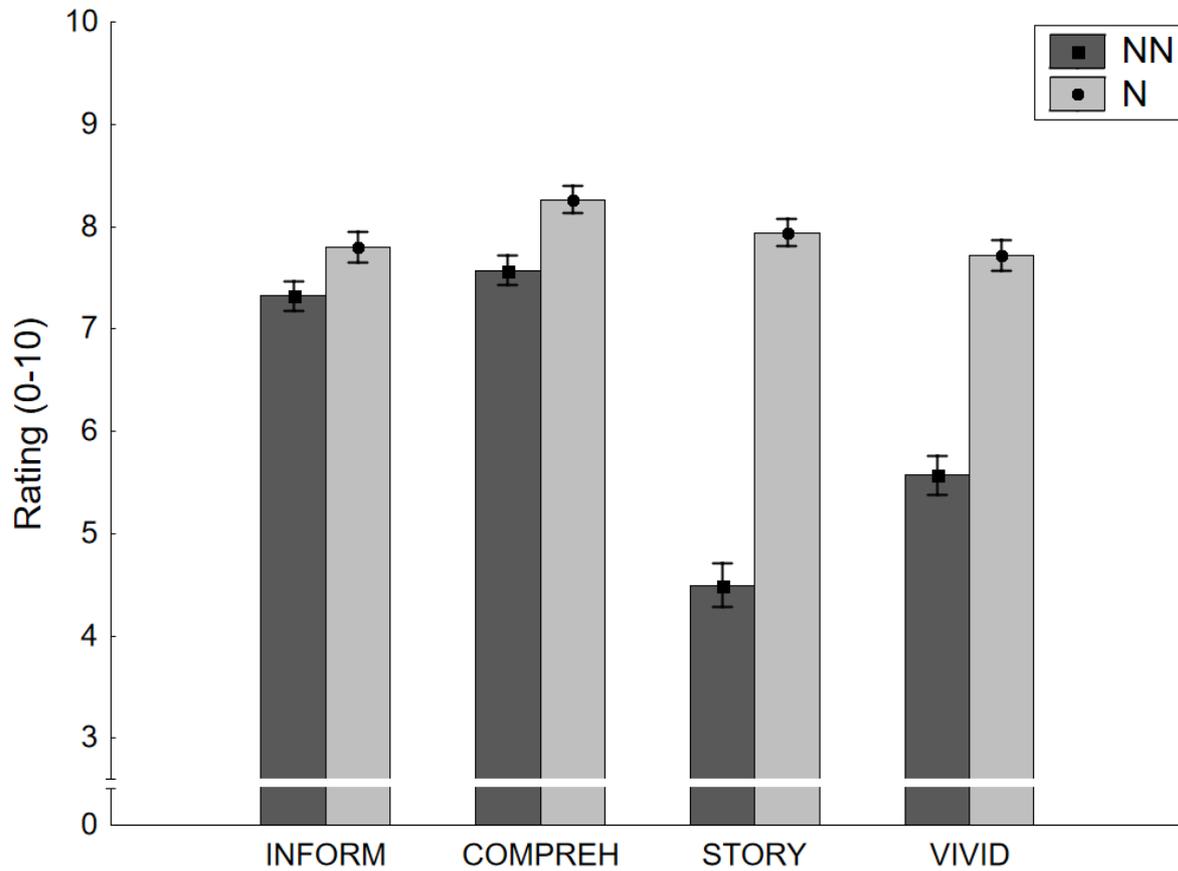
			reduce the risk of arthritis in later life.	with joint pains in their hands to avoid knuckle-cracking?
Myth D	Delayed-onset muscle soreness is caused by build-up of lactic acid.	After strenuous exercise, a warm-down routine is essential because it breaks-down the lactic acid that contributes to delayed-onset muscle soreness.	How effective do you think supplements that help break down lactic acid are in preventing exercise-induced muscle soreness?	How likely is it that you would advise a friend or family member with exercise-induced muscle soreness to avoid exercise activities that create lactic acid?
Fact A	Stomach acid can dissolve razor blades.	Teaching teenagers that our stomach acid can dissolve razor blades would be an accurate and entertaining way to inform them about chemistry.	How effective do you think stomach acid is at dissolving razor blades?	How likely is it that a razor blade would be totally intact after 48 hours in stomach acid?
Fact B	It is not safe to talk on landline telephones when there is a thunderstorm.	People should be discouraged from talking on landlines during thunderstorms to reduce their risk of being electrocuted.	Even when inside, people should opt to use mobile phones instead of landlines during a thunderstorm.	How likely is it that you would advise a friend or family member not to talk on a landline during a thunderstorm?
Fact C	Dogs can smell cancer.	Sniffer dogs are a reliable and effective way to detect some cancers.	Sniffer dogs trained to detect cancer should be utilized more in hospitals.	To what extent would you trust the response of sniffer dog over a traditional screening test of lung cancer?
Fact D	We are taller in the morning than in the evening.	If you are half an inch too short to go on a rollercoaster in the evening, how likely is it that you would be allowed to ride the following morning?	If you want to seem taller, you should measure yourself first thing in the morning.	When doctors measure their patients, they should take into account the time of day.

1010 *Note.* All inference questions are measured on 11-point Likert scales from 0 (strongly disagree) to 10 (strongly agree).

1011 **Pilot study.** A different sample of 102 U.S.-based MTurk workers (min. 5,000 HITs
1012 completed with 98%+ approval rate) was recruited to rate the non-narrative and narrative
1013 corrections of all real-world myths. One participant was excluded due to uniform responding
1014 ($SD = 0$), and one was excluded because they indicated we should not use their data due to
1015 lack of effort. This left $N = 100$ participants ($M_{age} = 37.58$ years; age range 21-65; 61 males,
1016 39 females).

1017 All myths were presented in randomized order. For each myth, participants read both
1018 corrections, also in randomized order. They were asked to rate each correction on
1019 informativeness (“How informative is the correction?”), comprehensibility (“How easy to
1020 understand is the correction?”), story-ness (“How story-like is the correction?”), and
1021 vividness (“How vivid is the correction?”), all on 0 (not at all) – 10 (very much) scales. The
1022 imaginability dimension was omitted as the non-narrative correction featured no events that
1023 could have been pictured.

1024 Results closely mirrored the findings from the Experiment 1 Pilot, and are
1025 summarized in Figure A2. Again, there was a large difference in story-ness between non-
1026 narrative and narrative corrections, with a substantial difference also on vividness. There was
1027 no difference between conditions on comprehensibility, and only a small to-be-expected
1028 difference on informativeness. We again concluded that our manipulation was implemented
1029 successfully.



1030

1031 *Figure A2.* Ratings of non-narrative (NN) and narrative (N) myth corrections on
 1032 informativeness (INFORM), comprehensibility (COMPREH), story-ness (STORY), and
 1033 vividness (VIVID) in the Experiment 2 Pilot. Error bars give within-subjects standard error
 1034 of the mean.

1035 **Experiment 3**

1036 **Claims and explanations.** On average, the non-narrative corrections had 112 words, with FRE = 45.55 and FKGL = 11.9; narrative
 1037 corrections had 117.5 words, with FRE = 55.55 and FKGL = 10 (see Table B1). Affirmations had on average 86.5 words, with FRE = 37.1 and
 1038 FKGL = 12.85 (see Table B2).

Table B1

Myths and their Corresponding Non-Narrative and Narrative Corrections

Item number	Items	Non-Narrative Correction	Narrative Correction
Myth - 1	Humans are made to eat red meat; it should be part of every person's diet.	Recent research-based evidence published in a leading journal shows that eating red meat on a regular basis may shorten people's lifespans. The findings of the study suggest that meat eaters might improve their health by making simple changes. One suggestion made is to substitute one serving of red meat (like bacon or steak) a day with another type of protein. Options include fish, chicken, legumes, low-fat dairy and whole grains. The results of the study suggest that rotating in other foods in place of red meat could lower the risk of mortality by 7 to 19%. (WC = 96; FRE = 58.6; FKGL = 9.8)	"To me, there's no finer pleasure than smelling bacon in the morning, or sinking my teeth into a perfectly cooked steak. You can imagine my panic when my daughter, who is a nurse, showed me research-based evidence that eating red meat frequently may shorten my lifespan! She asked, 'Promise me you'll make some changes? Just substitute one serving a day with another protein.' With her help, I rotated in other foods like fish, chicken, legumes, low-fat dairy, and whole grains. She says that lowers my mortality risk by 7 to 19%. I still get to enjoy a sizzling steak on special occasions!" (WC = 102; 1.06 ratio; FRE = 66.8; FKGL = 7.5)

<p>Myth - 2 Children of homosexual parents have more mental health issues.</p>	<p>A large body of research has examined the question of whether children of homosexual parents have poorer development outcomes. This research has looked at a wide range of social, emotional, health and academic outcomes. It has compared patterns of mental health and related outcomes in children with same-sex parents compared to children in more traditional households. This research shows that children or adolescents raised by same-sex parents fare equally as well as those raised by opposite-sex parents. An article published in the Journal of Marriage and Family in 2010 conducted a summary analysis of 33 individual studies on the topic. The results of the research review suggest that the strengths that are typically associated with mother-father families appear to the same degree in families with two same-sex parents.</p> <p>(WC = 128; FRE = 32.5; FKGL = 14)</p>	<p>“People sometimes ask me what it’s like to have two mothers, rather than a mom and a dad. It seems to me like my family does the same things other, “normal” families do. For a college project, I actually looked into the research, and found that children or adolescents raised by same-sex parents fare equally as well as those raised by opposite-sex parents on a wide range of social, emotional, health and academic outcomes. One study, published in the Journal of Marriage and Family in 2010, analyzed the results of 33 individual studies to assess how the gender of parents affected children. The authors found that the strengths typically associated with mother-father families appear to the same degree in families with two same-sex parents. I certainly don’t feel any different than my peers!”</p> <p>(WC = 133; 1.04 ratio; FRE = 44.3; FKGL = 12.5)</p>
--	---	---

1039 *Note.* WC = Word Count; FRE = Flesch Reading Ease; FKGL = Flesch-Kincaid Grade Level.

Table B2

Facts and their Corresponding Affirmation

Item number	Items	Affirmation
Fact - 1	Laughing regularly helps improve vascular function.	<p>It is well known that laughter reduces stress hormones and releases endorphins, yet strangely enough, it also has a positive impact on vascular function. A 2009 study found that people with heart disease were 40% less likely to laugh in a variety of situations compared to people without heart disease. A study in 2010 demonstrated the short-term benefits of laughter by showing participants either a 20-minute clip of a comedy or a documentary. Laughter led to tissue dilation in the inner lining of blood vessels, which increased blood flow.</p> <p>(WC = 90; FRE = 39.2; FKGL = 13.3)</p>
Fact - 2	U.S. citizens are the most generous people in the world.	<p>U.S. citizens are consistently rated the most generous people in the world. Be it volunteering their time, donating money to charity, or helping out a stranger in need, the World Giving Index reports that 58% of Americans regularly partake in an act of generosity. That is more people per capita than any other country. In 2018 alone, U.S. citizens donated a staggering \$292 billion dollars to charity. More than half of individuals reported that financial constraints were stopping them from donating even more!</p> <p>(WC = 83; FRE = 35.0; FKGL = 12.4)</p>

1041 *Note.* WC = Word Count; FRE = Flesch Reading Ease; FKGL = Flesch-Kincaid Grade Level.

1042

1043 *Test Questionnaires*

1044 Table B3

Myths and Facts, and Corresponding Inference Questions

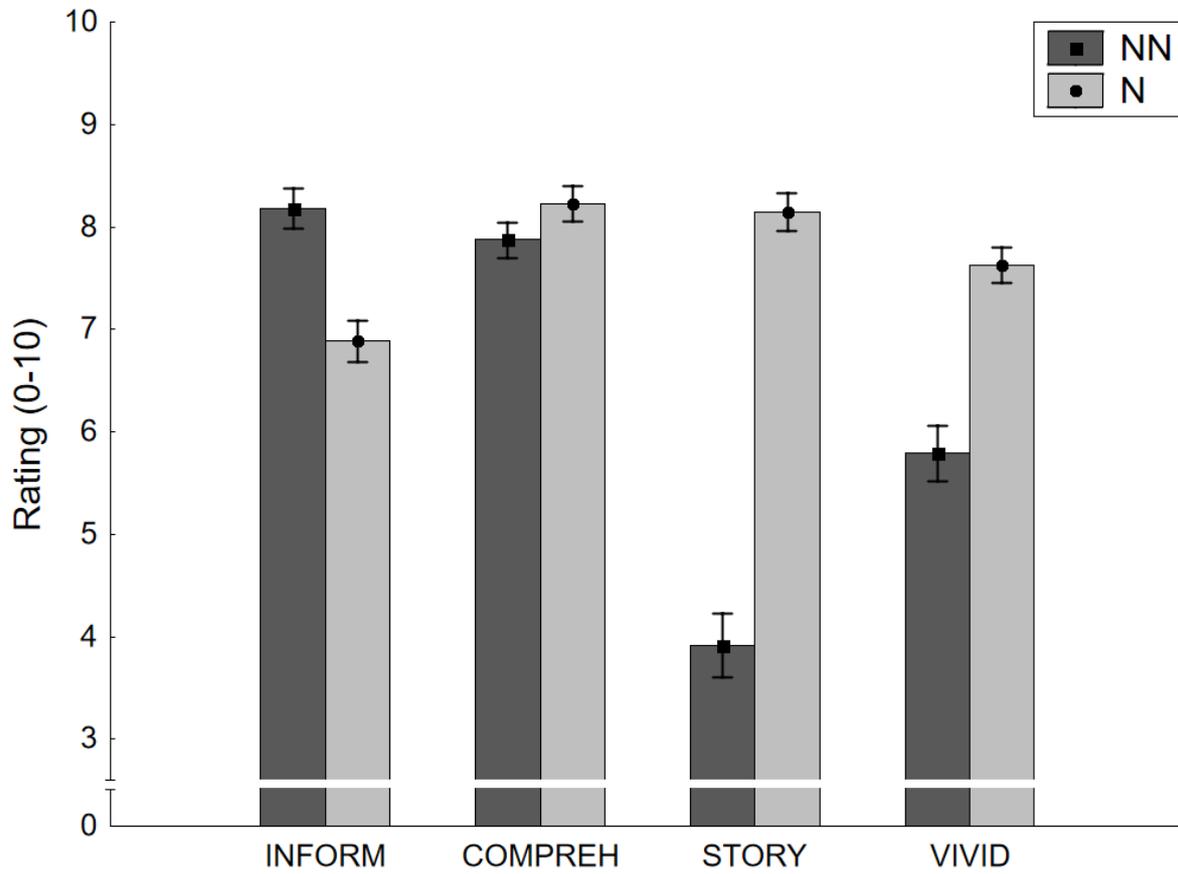
Item number	Items	Inference Question 1	Inference Question 2	Inference Question 3
Myth - 1	Humans are made to eat red meat; it should be part of every person’s diet.	Meals served to children at schools should include at least one serving of red meat every day.	To maintain a healthy diet, people should regularly consume red meat.	Diets and health care plans that do not include red meat are unsustainable for humans.
Myth - 2	Children of homosexual parents have more mental health issues.	School counsellors should be trained to look for characteristics of anxiety and depression in children of homosexual couples.	Children whose parents are homosexual are at an increased risk of experiencing mental health issues.	Homosexual couples considering adoption should consider the impact of their homosexuality on the child’s mental health.
Fact - 1	Laughing regularly helps improve vascular function.	Laughing workshops should be recommended for people with cardiovascular diseases.	The American Heart Association should run an advertisement campaign promoting laughter as a preventative measure for heart disease.	People should be advised to watch comedies as a way to improve their heart health.
Fact - 2	U.S. citizens are the most generous people in the world.	Americans should be regarded as generous people.	Americans can be proud of their generosity.	Charities seeking funds would be well suited to target Americans as potential donors.

1045 **Test questionnaire.**

1046 **Pilot study.** A separate sample of $N = 100$ U.S.-based MTurk workers (min. 5,000
1047 HITs completed with 98%+ approval rate; $M_{\text{age}} = 36.43$ years; age range 20-70; 57 males, 43
1048 females) was recruited to rate the non-narrative and narrative corrections of both
1049 controversial real-world myths.

1050 Both myths were presented in randomized order. For each myth, participants read
1051 both corrections, also in randomized order. They were asked to rate each correction on
1052 informativeness (“How informative is the correction?”), comprehensibility (“How easy to
1053 understand is the correction?”), story-ness (“How story-like is the correction?”), and
1054 vividness (“How vivid is the correction?”), all on 0 (not at all) – 10 (very much) scales.

1055 Results closely mirrored the findings from the Experiment 2 Pilot, and are
1056 summarized in Figure A3. Again, there was a large difference in story-ness between non-
1057 narrative and narrative corrections, with a substantial difference also on vividness. There was
1058 no difference between conditions on comprehensibility, and only a moderate difference on
1059 informativeness (with the non-narrative correction being rated somewhat more informative,
1060 which was expected given the narrative correction provided more arbitrary, conversational
1061 information). We again concluded that our manipulation was implemented successfully.



1062

1063

1064

1065

1066

Figure A3. Ratings of non-narrative (NN) and narrative (N) myth corrections on informativeness (INFORM), comprehensibility (COMPREH), story-ness (STORY), and vividness (VIVID) in the Experiment 3 Pilot. Error bars give within-subjects standard error of the mean.