Check for updates

OPEN

The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada

John M. Carey [1], Andrew M. Guess [2], Peter J. Loewen [3,4], Eric Merkley [4], Brendan Nyhan [1✉], Joseph B. Phillips[5] and Jason Reifler [6]

Widespread misperceptions about COVID-19 and the novel coronavirus threaten to exacerbate the severity of the pandemic. We conducted preregistered survey experiments in the United States, Great Britain and Canada examining the effectiveness of fact-checks that seek to correct these false or unsupported beliefs. Across three countries with differing levels of political conflict over the pandemic response, we demonstrate that fact-checks reduce targeted misperceptions, especially among the groups who are most vulnerable to these claims, and have minimal spillover effects on the accuracy of related beliefs. However, these reductions in COVID-19 misperception beliefs do not persist over time in panel data even after repeated exposure. These results suggest that fact-checks can successfully change the COVID-19 beliefs of the people who would benefit from them most but that their effects are ephemeral.

Misinformation about COVID-19 has circulated widely during the pandemic, confusing people about the disease and encouraging them to flout expert advice and government recommendations about how to prevent its spread[1–3]. In response, journalists and public health officials have devoted a great deal of effort to debunking misinformation about the new coronavirus. Will these efforts be successful? It is essential to determine how to most effectively counter false claims, which may not only distort beliefs and attitudes but also threaten public health.

Recent meta-analyses show that fact-checks and corrective information tend to reduce misperceptions immediately after exposure[4–6]. However, most prior studies have been conducted with American respondents, neglecting potential heterogeneity across countries in factors such as political polarization, media system structure and trust in scientific expertise. In addition, past studies typically concern longstanding political and health controversies; studying a new topic like COVID-19 allows us to estimate receptiveness to factual evidence on an important but uncertain and rapidly changing issue. Most importantly, it is critical to determine whether the corrective effects we typically observe immediately after exposure can endure[7]. Does corrective information cause lasting change in factual beliefs or do its effects quickly dissipate?

We investigate these questions using parallel preregistered experiments testing the effects of exposure to fact-checks on COVID-19 pandemic misperceptions in the United States, Great Britain and Canada. (Our British sample excludes Northern Ireland, so we use the more technically correct term—Great Britain—for our sampling frame of England, Scotland and Wales, which together make up 97% of the population of the United Kingdom.) These countries combine demographic similarity with sharp differences in the extent to which the COVID-19 response has become politicized and in the severity of the pandemic at the time the studies were

fielded[8–10], allowing us to provide one of the most systematic tests to date of whether the effects of fact-checking generalize across different political and social contexts.

Our study design also allows us to test important questions about fact-check exposure levels, effect durability and consequences for related factual beliefs. First, we exploit the multiwave panel design of our surveys of the United States and Great Britain to independently randomize exposure to fact-checks or placebo in multiple waves and to measure misperception beliefs over multiple waves. This design allows us to assess the persistence of fact-check effects over time and to test whether repeating fact-checks makes their effects more durable. (These analyses exclude the Canada surveys, which used single-wave designs.) We also examine to what extent fact-checks spill over to affect beliefs about other true and false claims related to COVID-19.

Results from all three countries demonstrate that corrective information successfully reduces misperceptions about COVID-19. Exposure to fact-checks reduces the perceived accuracy of targeted claims within a given survey wave, especially among people who are most vulnerable to those claims, and has minimal spillover effects on other related beliefs (that is, we find no evidence that fact-check messages targeting one COVID-related issue affect respondent beliefs on non-targeted issues). However, panel data from the United States and Great Britain reveal that the resulting accuracy improvements dissipate within weeks and are no longer measurable in subsequent waves. These findings suggest that durably reducing misperceptions about COVID-19 requires frequent exposure to corrective information because the effects dissipate quickly even with repetition.

The specific hypotheses we test are as follows. Our study examines the effectiveness of exposure to fact-checks at reducing misperceptions about COVID-19. Theoretically, judgments about factual

[1]Department of Government, Dartmouth College, Hanover, NH, USA. [2]Department of Politics and School of Public and International Affairs, Princeton University, Princeton, NJ, USA. [3]Department of Political Science, University of Toronto, Toronto, Ontario, Canada. [4]Munk School of Global Affairs & Public Policy, University of Toronto, Toronto, Ontario, Canada. [5]School of Psychology, University of Kent, Canterbury, UK. [6]Department of Politics, University of Exeter, Exeter, UK. ✉e-mail: nyhan@dartmouth.edu

claims like these can be subject to competing information processing goals: forming accurate beliefs versus defending existing beliefs or attitudes[11,12]. The relative weight of accuracy and directional goals varies by context. Given the threat that COVID-19 poses to human health and well-being, accuracy goals may be especially salient relative to other controversial issues. Moreover, exposure to fact-checks and other forms of corrective information tends to increase the accuracy of people's beliefs[4–7]. We therefore expect fact-checks to reduce the perceived accuracy of the misperceptions they target immediately after exposure (hypothesis H1).

We also test the hypothesis (H2) that fact-check effects on beliefs will persist on the basis of prior studies, which have found that fact-check effects can be durable over a period of weeks[13,14]. However, other research suggests that they could attenuate more quickly. Treatment effects of political messages on opinion often decay rapidly[15–17]. Even with strong accuracy motives, people may fail to incorporate information from fact-checks into long-term memory[18].

Repetition, an approach frequently used in public health campaigns, is one way to counter the potential decay of fact-check effects. Repeated exposure can make claims more familiar, creating greater fluency that leads people to be more likely to infer that a claim is true[19,20]. Multiple exposures can also help people to fully process a fact-check and incorporate its contents into memory[21]. However, repeated exposure to fact-checks could also have diminishing effects due to diminished novelty[22]. On the basis of these competing expectations, we pose a research question (RQ1) asking whether prior exposure to a fact-check changes the effect of subsequent fact-check exposure on misperception beliefs.

A second research question (RQ2) asks whether fact-checks affect the perceived accuracy of other beliefs about COVID-19. One possibility is that correcting a misperception causes people to reconsider related beliefs that are now more cognitively accessible[1,10]. Exposure to fact-checks may also increase the salience of accuracy, which has been shown to increase truth discernment in sharing news online[23]. Alternatively, correcting misperceptions could cause people to misdirect their skepticism toward accurate beliefs[24,25] or to infer that uncorrected claims are more accurate[26].

Finally, we test for heterogeneous treatment effects among groups that are more misinformed or more vulnerable to misinformation. First, we consider political predispositions (RQ3A). In the United States, Republicans and supporters of President Trump hold more misperceptions about COVID-19 (refs. [10,27]). More generally, conservatives have been found to hold more misperceptions and/or to be less supportive of pandemic mitigation behaviours than are moderates or liberals in the United States, the United Kingdom and Canada[10,28–32]. We therefore test if treatment effects vary by party identification or approval of/feelings toward the country's chief executive. Second, we consider trust in authoritative sources of information (RQ3B). People who distrust the health care system or the media or who have predispositions toward conspiracy theories may be more vulnerable to COVID-19 misinformation and/or more distrusting of fact-checks of those claims. Finally, we leverage the panel survey design in our United States and Great Britain studies to examine pretreatment misperceptions as a moderator of fact-check effectiveness (RQ3C). Each of these groups is potentially more resistant to fact-checks but also more likely to be misinformed, creating greater scope for fact-checks to be effective.

We evaluate these hypotheses and research questions using the following experimental designs. In the samples for the United States and Great Britain, we conducted 2×2 between-subjects experiments in which we measured COVID-19 beliefs over multiple waves. In wave 1, respondents provided political and demographic information and rated the accuracy of both true and false claims about COVID-19 (including ones that were fact-checked in future waves). In wave 2, respondents were randomly assigned with equal

probability to receive either four articles adapted from fact-checkers in the United States and the UK debunking myths about COVID-19 or four unrelated placebo articles (within each condition, the four articles were presented in randomized order). The fact-checks debunked the false claims that hydroxychloroquine and antibiotics can cure COVID-19 and the conspiracy theories that COVID-19 is a Chinese bioweapon or was patented by Bill Gates. In wave 3, respondents were independently randomly assigned with equal probability to receive the same fact-checks or the placebo articles using the same procedure. These randomizations yield four experimental groups: respondents who received no fact-checks, fact-checks only in wave 2, fact-checks only in wave 3 or fact-checks in both waves 2 and 3. After the randomizations in both wave 2 and wave 3, respondents in the United States and Great Britain again rated the accuracy of various claims about COVID-19, including false claims targeted by the fact-checks, false claims the fact-checks did not mention and true claims. (US respondents also rated these claims in a fourth wave that did not include fact-checks. It was completed 7 months after the completion of wave 3—a substantial time lag.)

In our Canadian single-wave samples, pretreatment measures were collected at the beginning of the study and participants were assigned with equal probability to either a fact-check or placebo condition (Methods and Supplementary Information).

Our experimental treatments did not expose people to uncorrected misinformation and instead included accurate information designed to address common misperceptions. We note that our survey did ask respondents to assess the accuracy of both true and false claims about COVID-19. (We also referred participants at the end of the survey to their relevant national health authority for more information about the new coronavirus and COVID-19.)

Results

Targeted false claims. Results from our surveys in the United States, Great Britain and Canada indicate that belief in the misperceptions targeted by our fact-checks was relatively widespread, ranging from 16% to 31% in the United States, 7% to 32% in Great Britain and 21% to 29% in Canada across samples (Supplementary Table 30).

Exposure to fact-checks debunking these claims induces substantial reductions in misperceptions in each country but the effects do not persist in later waves. Results, summarized in Fig. 1, are estimated using ordinary least squares (OLS) with robust standard errors and follow our preregistrations unless noted. (Links to each preregistration and to replication data and code are provided in Methods, Data Availability and Code Availability, respectively.)

Consistent with our first preregistered hypothesis, fact-checks decrease the perceived accuracy of an aggregate measure of false claims targeted by the fact-checks immediately after exposure in all survey waves in all countries (United States: $t(2,391) = -13.509$, $P < 0.001$, $\beta = -0.227$, 95% confidence interval (CI) $(-0.260, -0.194)$; Great Britain: $t(3,141) = -18.212$, $P < 0.001$, $\beta = -0.248$, 95% CI $(-0.275, -0.222)$; Canada 1: $t(1,074) = -4.655$, $P < 0.001$, $\beta = -0.171$, 95% CI $(-0.244, -0.099)$; Canada 2: $t(1,092) = -3.658$, $P < 0.001$, $\beta = -0.139$, 95% CI $(-0.213, -0.064)$; Supplementary Tables 35–52). The sizes of these effects range from -0.09 (Canada 2 sample) to -0.24 (Great Britain wave 3) standard deviations (standardized effects were calculated separately for each sample). In relative terms, these treatment effects represent reductions of 30.4% (United States sample), 44.7% (Great Britain sample), 19.0% (Canada 1 sample) and 15.6% (Canada 2 sample) in the perceived accuracy of targeted false claims (Supplementary Information).

An exploratory analysis of individual belief items reveals that the fact-checks are consistently effective against the bioweapon, antibiotic and hydroxychloroquine misperceptions across countries, waves and samples (Supplementary Tables 39–52). By contrast,
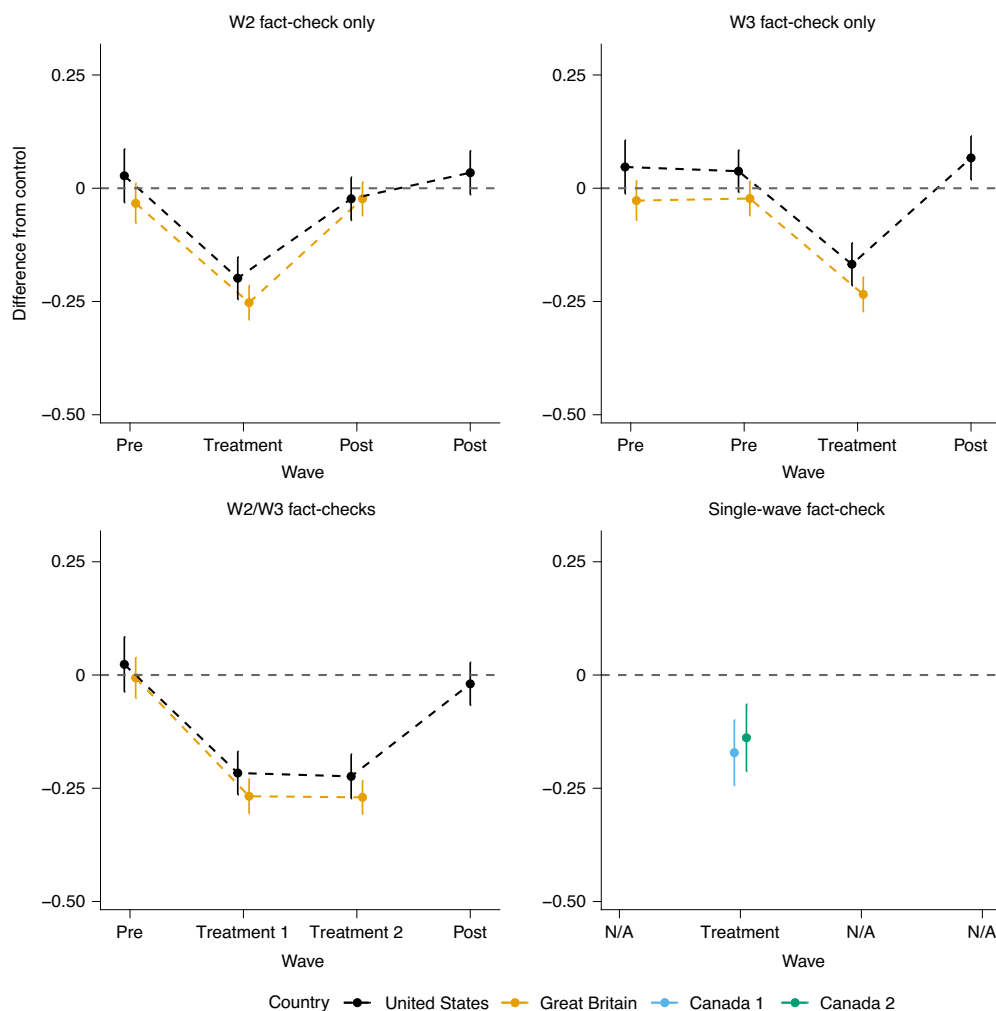
Fig. 1 | Effects of fact-check exposure on mean perceived accuracy of targeted false claims about COVID-19. United States: $n=2,397$ in wave 2 (W2), $n=2,366$ in wave 3 (W3) and $n=2,372$ in wave 4 (W4). Great Britain: $n=3,147$ in W2 and $n=3,140$ in W3. Canada 1: $n=1,147$. Canada 2: $n=1,144$. Data are presented as unstandardized regression coefficients ±1.96 s.e. from OLS regression with HC2 robust standard errors. Two-sided t-tests. No adjustments were made for multiple comparisons. Full statistical results can be found in the models in Supplementary Tables 35–38 with covariates. 'Single-wave fact-check' refers to the Canada 1 and Canada 2 experiments, which were administered in a single survey wave (Methods). 'Pre' refers to a pre-treatment survey wave and 'Post' refers to a post-treatment survey wave. 'N/A' indicates that there were no pre- or post-survey waves conducted in Canada. Fact-checks were randomly shown in wave 2 and/or wave 3 of panel surveys that was administered in the United States and Great Britain. The control condition did not receive fact-checks in either wave 2 or wave 3. See Methods and Supplementary Information for details.

the fact-checks reduce Gates conspiracy belief only in the United States in wave 2. Belief in the Gates conspiracy was lower than other targeted misperceptions, particularly in Great Britain where a floor effect may have limited the potential impact of the fact-check (Supplementary Table 30).

However, we find no support for our second hypothesis, which predicted that fact-check effects would persist after the wave of exposure. In our panel data from the United States and Great Britain, we could not reject the null hypothesis that exposure to fact-checks in wave 2 had no effect on the perceived accuracy of targeted false claims in wave 3 (Supplementary Tables 35–52). These effects allow us to rule out reductions in the perceived accuracy of targeted claims of −0.055 or greater (−0.081 s.d.) in the United States and −0.087 or greater in Great Britain (−0.175 s.d.) at the 95% CI (exploratory analysis using two one-sided equivalence tests: United States (−0.055, 0.079), Great Britain (−0.087, −0.006)).

We also preregistered a research question asking whether exposure to fact-checks in wave 2 would change the effect of exposure

to fact-checks in wave 3. In both the United States and Great Britain, the interaction between the wave 2 and 3 treatments was not statistically significant, providing no support for this conjecture. Exploratory analyses of wave 4 of the United States sample reveal no consistent patterns of persistence or interactions between wave 2 and 3 treatments (Supplementary Tables 79–80).

Effect heterogeneity. To explore whether any of our treatment effects were moderated by individual-level characteristics or prior attitudes, we estimated a preregistered series of flexible nonlinear models designed to explore heterogeneity while minimizing false positives via regularization. We use BCF[33], an extension of tree-based approaches such as BART[34] that explicitly incorporates treatments as separate from other pretreatment covariates. BCF generate posterior estimates of individual-level treatment effects that can be summarized for subgroups of interest. In the Supplementary Information, we present graphical summaries of these treatment effect estimates plotted against the relevant moderator.

Our results provide encouraging evidence that fact-checks are often more effective at reducing misperceptions among the groups that are most vulnerable to them. In wave 2 of the United States sample, for example, we find via recursive partitioning of the Bayesian Causal Forests (BCF)-estimated posterior treatment effects that the negative effect of fact-check exposure on the perceived accuracy of false claims about COVID-19 is greater among Trump approvers in the United States sample. Specifically, CIs from BCF posterior estimates confirm that Trump approvers (wave 2, 95% CI (−0.413, −0.009); wave 3, 95% CI (−0.123, −0.018)) are more receptive to fact-checks than are disapprovers (wave 2, 95% CI (−0.216, 0.041); wave 3, 95% CI (−0.065, 0.002)). Similarly, we find that negative treatment effects in wave 2 of the British sample are greater among respondents who approve of Johnson (95% CI (−0.101, −0.013)) than among those who disapprove (95% CI (−0.089, 0.009), although we do not find evidence of heterogeneous effects by approval in wave 3 of the British sample or in either Canada sample.

To illustrate this finding, Fig. 2 presents an exploratory analysis following the same approach as Fig. 1 above but estimates treatment effects separately in a frequentist manner for respondents who either approved or disapproved of their national leader at the time the experiments were fielded (Trump in the United States, Johnson in Great Britain or Trudeau in Canada). As the figure indicates, fact-checks reduced targeted misperceptions immediately after exposure more among respondents who approved of their national leader than among those who disapproved in both the United States and Great Britain. By contrast, we find no clear evidence of heterogeneous effects by leader approval in Canada, where baseline misperceptions were uncorrelated with approval (unlike in the United States and Great Britain).

Among people who disapproved of their national leaders, the treatment produced relative reductions in the perceived accuracy of false beliefs within wave of 31.5% and 40.9% in the two treatment waves of the United States sample, 41.2% and 46.8% in the two treatment waves of the Great Britain sample, 20.5% in the Canada 1 sample and 23.8% in the Canada 2 sample. Among those who approve of their national leaders, the relative effects were 17.4% and 27.9% in the treatment waves of the United States sample, 38.7% and 49.8% in the treatment waves of the Great Britain sample, 19.7% in the Canada 1 sample and 8.1% in the Canada 2 sample. (Note that although the absolute effects are smaller among disapprovers than approvers in the United States as described above, the relative reductions are greater among disapprovers due to lower baseline levels of false belief.) The effect among approvers in the Canada 2 sample is the only case in either set in which the 95% CI includes zero (see the Supplementary Information, beginning on page 92, for *t*-statistics, *P* values, point estimates and 95% CIs for each treatment effect estimate by subgroup and country).

Recursive partitioning of the BCF-estimated posterior treatment effects also provides evidence of heterogeneous treatment effects for several other moderators, all of which indicate that the effects of fact-check exposure are often greater among more vulnerable groups. Within wave 2 of the United States sample, we find that treatment effects are higher among those with more prior misperceptions (95% credible interval (−0.457, −0.118)) than those with fewer (95% credible interval (−0.195, 0.041)). Within wave 3 of the United States sample, we find that treatment effects are higher among those with high conspiracy predispositions (95% credible interval (−0.122, −0.006)) than low conspiracy predispositions (95% credible interval (−0.079, 0.002)); among those with more rather than fewer prior misperceptions (more, 95% credible interval (−0.122, −0.005); fewer, 95% credible interval (−0.085, 0.002)); among Republicans (95% credible interval (−0.123, −0.009)) rather than Democrats (95% credible interval (−0.070, 0.002)) or Independents (95% credible interval (−0.103, 0.003)); among those with low trust in health institutions (95% credible interval (−0.127, −0.001)) rather

than high trust (95% credible interval (−0.098, 0.002)); and among those with low trust in the media (95% credible interval (−0.130, −0.006)) rather than high trust (95% credible interval (−0.099, 0.002)). Similarly, treatment effects in wave 2 of the British sample are larger among respondents who identify with right-wing parties (95% credible interval (−0.101, −0.013)) than with left-wing parties (95% credible interval (−0.079, 0.009)) or non-identifiers (95% credible interval (−0.100, 0.001)); among those with low trust in health institutions (95% credible interval (−0.108, −0.014)) rather than with high trust (95% credible interval (−0.070, 0.009)); and among those with higher prior misperceptions (95% credible interval (−0.122, −0.038)) than with lower prior misperceptions (95% credible interval (−0.090, 0.009)). Treatment effects in the Canada 1 sample are higher among those with stronger conspiracy dispositions (95% credible interval (−0.120, −0.017)) than weaker ones (95% credible interval (−0.054, 0.043)). We find the same with the Canada 2 sample (high, 95% credible interval (−0.082, −0.004); low, 95% credible interval (−0.034, 0.046)).

Spillover effects. We also preregistered a second research question asking whether exposure to fact-checks on some items affects the perceived accuracy of true claims about COVID-19 or false claims that were not targeted by fact-checks. We find little evidence of spillover for true claims (Supplementary Tables 35–52 give the results, which include results estimated by item). We find no evidence of spillover in either treatment wave in the United States sample, in wave 3 of the British sample or in either Canadian sample, although fact-check exposure did minimally affect belief in true claims in wave 2 in Britain ($t(3,117) = -2.624$, $P = 0.009$, $\beta = -0.028$, 95% CI (−0.049, −0.007)).

When we examine the perceived accuracy of false claims that were not targeted by the fact-checks, we deviate from our preregistration to separately examine the perceived accuracy of the claim that 'The Chinese government is covering up the fact that the coronavirus escaped from one of its research laboratories', which was preregistered as part of our scale of non-targeted false claims. When our studies were designed, predominant expert opinion rejected the laboratory leak account[35]. Subsequent reports have increased the plausibility of the laboratory leak account, although the matter remains unresolved[36].

We therefore first examine the effects of fact-check exposure on the index of non-targeted false claims that excludes the laboratory leak item (Supplementary Tables 89–92). In two of six survey waves where fact-checks were administered, we found statistically significant but substantively small effects (United States wave 2: $t(2,398) = -3.283$, $P = 0.001$, $\beta = -0.045$, 95% CI (−0.072, −0.018); Great Britain wave 2: $t(3,146) = -3.345$, $P < 0.001$, $\beta = -0.039$, 95% CI (−0.062, −0.016)). In the other four cases, we find that fact-checks lacked a statistically significant immediate impact (for example, United States wave 3: $t(2,385) = 0.406$, $P = 0.685$, $\beta = 0.009$, 95% CI (−0.034, 0.051)).

As with the laboratory leak findings below, these effects may be driven by conceptual relationships between fact-checks and other false claims. Two of our fact-check corrections focused on potential COVID-19 cures (hydroxychloroquine and antibiotics) and in some instances exploratory results show spillover effects to other unproven remedies (colloidal silver and vitamin C) that were not targeted by fact-checks (Supplementary Tables 75–80).

By contrast, fact-checks reduce belief in the laboratory leak claim itself immediately after exposure. These effects were measurable in each sample (United States wave 2: $t(2,375) = -10.135$, $P < 0.001$, $\beta = -0.309$, 95% CI (−0.368, −0.249); United States wave 3: $t(2,368) = -5.553$, $P < 0.001$, $\beta = -0.251$, 95% CI (−0.339, −0.162); Great Britain wave 2: $t(3,167) = -0.15.781$, $P < 0.001$, $\beta = -0.435$, 95% CI (−0.489, −0.381); Great Britain wave 3: $t(3,169) = -7.444$, $P < 0.001$, $\beta = -0.295$, 95% CI (−0.373, −0.217); Canada 1:
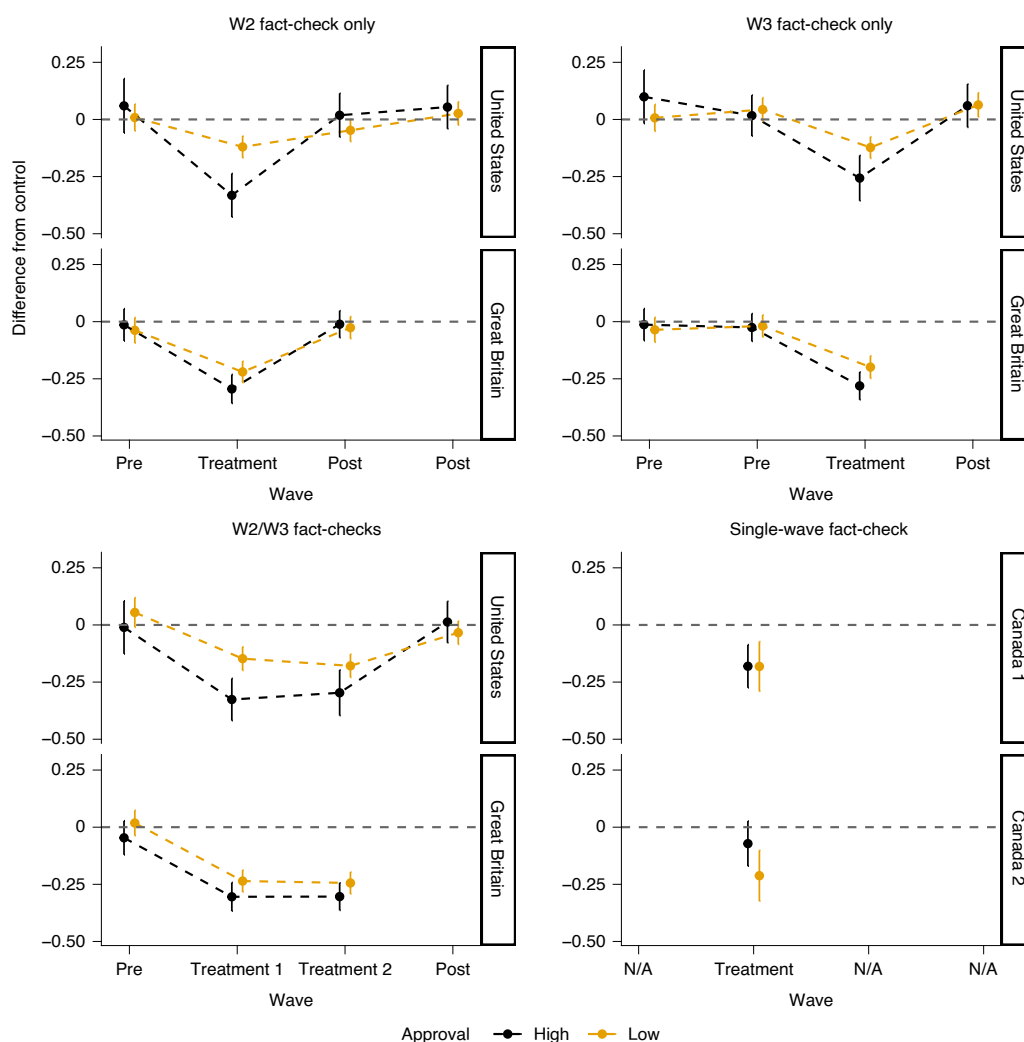
Fig. 2 | Effects of fact-check exposure on mean perceived accuracy of targeted false claims about COVID-19 by national leader approval. United States, approvers: $n=891$ in W2, $n=886$ in W3, $n=888$ in W4. United States, disapprovers: $n=1,483$ in W2, $n=1,480$ in W3, $n=1,484$ in W4. Great Britain, approvers: $n=1,308$ in W2, $n=1,304$ in W3. Great Britain, disapprovers: $n=1,841$ in W2, $n=1,835$ in W3. Canada 1: $n=697$ approvers, $n=485$ disapprovers. Canada 2: $n=697$ approvers, $n=450$ disapprovers. Error bars denote 95% CI. Data are presented as unstandardized regression coefficients ± 1.96 s.e. from OLS regression with HC2 robust standard errors. Two-sided t-tests. No adjustments were made for multiple comparisons. Full statistical results can be found in Supplementary Tables 93–98. 'Single-wave fact-check' refers to the Canada 1 and Canada 2 experiments, which were administered in a single survey wave (Methods). Fact-checks were randomly shown in wave 2 and/or wave 3 of panel surveys that was administered in the United States and Great Britain. The control condition did not receive fact-checks in either wave 2 or wave 3. See Methods and Supplementary Information for details.

$t(1,142) = -4.152$, $P < 0.001$, $\beta = -0.237$, 95% CI (-0.349, -0.125); Canada 2: $t(1,149) = -3.179$, $P = 0.002$, $\beta = -0.187$, 95% CI (-0.302, -0.072); Supplementary Tables 89–92. We interpret these effects as a consequence of the conceptual relationship between the fact-checked claim that the Chinese government created COVID-19 as a bioweapon and the laboratory leak claim. Both allege that China played a role in the spread of COVID-19 and then covered it up.

Discussion

Preregistered survey experiments in the United States, Great Britain and Canada show that exposure to fact-checks decreased the perceived accuracy of targeted false claims about COVID-19 immediately after exposure. These decreases in false beliefs were often greatest among people who were previously most misinformed and/or who were potentially especially susceptible due to political affiliations or distrust of established authorities. However, we find no evi-

dence that repeated exposure to fact-checks increases their effects or that exposure to these claims has durable effects on the accuracy of people's beliefs over time.

The implications of these findings are mixed. Encouragingly, our results demonstrate that fact-checks can reduce false beliefs about COVID-19 immediately after exposure and that these effects replicate across survey waves and across countries with different levels of polarization. False claims about emerging health issues can be effectively refuted with corrective information. However, our experiments find discouragingly little evidence that fact-checks have enduring effects on beliefs about COVID-19 or are strengthened by the repetition of fact-checks. It is possible that the avalanche of COVID news and (mis)information simply overwhelmed any potential persistence of the fact-checks delivered in our experiments. Under more favourable conditions, fact-checks might create more durable belief change or demonstrate cumulative exposure effects. Further research should test whether these results translate

to other health issues outside the context of a global pandemic. More generally, scholars should investigate how to create durable changes in belief accuracy about health issues.

These results also underscore the importance of cross-national comparative analysis of messaging about controversial health and science issues. The United States is often seen as an outlier due to high levels of polarization, which have hindered its pandemic response and may limit the effectiveness of fact-checking. However, we find similar results in Great Britain and Canada, which are less polarized generally and have experienced less partisan controversy over the response to the pandemic. Moreover, we find parallel erosion of fact-check effects over time in the United States and Great Britain, which suggests that levels of controversy over the pandemic are not responsible for the attenuation of fact-check effects over time that we observe.

Of course, this study has limitations that should be noted. First, we were only able to test fact-checks of four misperceptions about COVID-19 that were salient in May 2020 in three countries that are all rich, predominantly English-speaking and highly literate. Future studies should examine a broader array of false claims across a wider range of countries and health issues to provide greater leverage on how politics, culture and frequency of exposure to false information might affect the impact and duration of fact-checks. Second, our respondents were exposed to zero to two fact-check treatments, limiting our ability to test for a nonlinear relationship between repeated exposure and the effectiveness of fact-checks. Third, our study design does not allow us to observe effects on health behaviour; future studies should test whether fact-checks affect compliance with public health recommendations. Fourth, we used a design in which respondents were not allowed to select the information to which they were exposed; future research should incorporate designs that allow us to estimate the effects of fact-check exposure when people can choose the information they consume[37]. Fifth, we find no evidence that repeated exposure to an identical message from an unspecified source increases its effects on beliefs but future research should test if such exposure would have greater effects if the message was delivered by multiple sources. Finally, the precision of our measurement of persistence and decay effects is limited by the frequency of survey waves in the studies of the United States and Great Britain. In the United States, in particular, the time lag between the second round of fact-check treatments and our final measurement of beliefs was 7 months. Future research should estimate more precisely how quickly such correction effects decay.

Nonetheless, this research highlights the difficulty of durably reducing false beliefs. Without frequent exposure to fact-checks, misconceptions may return.

Methods

Ethics and consent. This research complies with all relevant ethical regulations. Informed consent was obtained from all participants. The studies were approved by human ethics review boards at Dartmouth College (STUDY00032068), Princeton University (IRB no. 12859) and the University of Toronto (protocol no. 00040160). The University of Exeter reviewed and approved the Great Britain study (SSIS Ethics Committee no. 201920-148) and recognized the approved protocols for the North American studies.

Survey codebooks and materials. The survey codebooks and materials can be found here: https://osf.io/8h4ze/

Preregistration links. The preregistrations for the experiments in each study are available at the links below:

- United States (16 July 2020): https://osf.io/e4ptq/
- Great Britain (22 October 2020): https://osf.io/bkfje/
- Canada (16 December 2020): https://osf.io/jz86u/

Randomization procedure. Respondents were randomized as described in the main text by YouGov (United States and Great Britain) or in Qualtrics (Canada). Respondents were blind to treatment assignment.

Participants/data collection. The studies conducted in the United States and Great Britain were designed as three-wave panels (although we were later able to field a fourth wave in the United States). In Canada, we fielded two single-wave studies whose design mirrored waves 2 and 3 of the United States and Great Britain studies. No statistical methods were used to predetermine sample sizes but our sample sizes are larger than those reported in previous publications[10,13,24].

The United States and Great Britain studies used similar sampling strategies. For our American sample, 4,438 YouGov respondents were recruited to a three-wave panel survey from three sampling frames (56% women, mean age 52.35 yr). A total of 1,096 respondents came from the general population YouGov panel; 2,238 came from YouGov's Pulse panel (which contains respondents who also consent to share their web-use histories); and 1,104 came from areas with a high incidence of COVID. Although a fourth wave was not preregistered, we were later able to conduct a fourth wave. Respondents completed wave 1 from 20 May to 3 June 2020 (n = 4,438; before the summer 2020 case surge in the United States); wave 2 from 25 June to 12 July 2020 (n = 3,680; approaching the peak of the summer 2020 US case surge); wave 3 from 28 July to 19 August 2020 (n = 2,983; after the peak of the summer 2020 US case surge); and wave 4 from 9 to 23 March 2021 (n = 2,464; after the winter 2020 peak in the United States but before the spring 2021 decline).

Our British sample initially consisted of 5,456 YouGov respondents (53% women, mean age 51.06 yr) from England, Wales and Scotland. A total of 2,367 respondents came from the general YouGov panel; 1,994 came from YouGov's Pulse panel; and 1,095 came from areas with a high incidence of COVID. Respondents completed wave 1 from 11 to 29 September 2020 (n = 5,456; as the autumn 2020 UK case surge begins); wave 2 from 10 to 23 December 2020 (n = 4,170; during the surge to the winter 2020 UK peak); and wave 3 from 4 to 22 February 2021 (n = 3,190; during the decline from the winter 2020 UK peak).

In contrast to our American or British samples, our Canadian surveys are repeated cross-sections. Our first sample consists of 1,300 Dynata (formerly Survey Sampling International) panellists (52% women, mean age 48.16 yr) who completed the study from 15 December 2020 to 13 January 2021 (at the peak of the winter 2020 case surge in Canada). Our second sample consists of 1,300 Dynata panellists (51% women, mean age 48.02 yr) who completed the study from 3 to 28 February 2021 (between the winter 2020 and spring 2021 case surges in Canada). Dynata uses quota-based sampling to approximate nationally representative samples. In this case, quotas were set on region (Atlantic, Quebec, Ontario and West) and language (French and English) along with interlocking quotas for education (degree versus no degree), age (18/34, 45–54, 55 yr and older) and gender to match population benchmarks from the 2016 Canadian census.

All respondents provided informed consent. Respondents were compensated with points from their respective survey firms, redeemable for money and gift cards.

Measures. We describe outcome measures and moderators used in the study below. The wave numbers we provide apply to the United States and Great Britain samples only. (The two Canadian samples were single-wave studies.)

Perceived accuracy of targeted false claims (waves 1, 2, 3 and 4). Respondents rated the perceived accuracy of four claims on four-point scales from 1 (not at all accurate) to 4 (very accurate): 'The Chinese government created the coronavirus as a bioweapon', 'A group funded by Bill Gates patented the coronavirus that causes COVID-19', 'Antibiotics are effective in preventing and treating the coronavirus' and 'The medication hydroxychloroquine is proven to cure or prevent COVID-19'. The order of the items was randomized. The sum of the items was used to construct a scale score on targeted misperceptions. (Analyses of scale reliability and factor loadings are provided in the Supplementary Information.)

Perceived accuracy of non-targeted false claims (waves 1, 2, 3 and 4). In the United States sample, respondents rated the perceived accuracy of six claims about COVID-19 that were not explicitly addressed in the fact-checks: 'Sneezing is a symptom of the coronavirus', 'Hand dryers effectively kill the coronavirus', 'The coronavirus is being spread by 5G cell phone technology', 'Eating garlic prevents infection with the coronavirus', 'High doses of vitamin C are an effective treatment for COVID-19' and 'Colloidal silver can cure COVID-19'. Responses were provided on four-point scales from 1 (not at all accurate) to 4 (very accurate). Space constraints limited these lists slightly in other surveys. The British samples include all items except the colloidal silver and hand dryers items. The Canada samples include all items. (Analyses of scale reliability and factor loadings are provided in the Supplementary Information.)

We initially preregistered eight such false claims. One of those, 'There is currently a vaccine proven to prevent the new coronavirus', was a false claim in waves 1–3 in the United States sample and in wave 1 in the British sample. However, it became true by the time subsequent survey waves were fielded in the United States and Great Britain and was true at the time of both Canadian samples. To maximize consistency within and across samples, we omit this item from analyses. Another item, 'The Chinese government is covering up the fact that the coronavirus escaped from one of its research laboratories', was initially categorized

as false but its status later changed to indeterminate. We thus analyse the item separately (above).

Perceived accuracy of true claims (waves 1, 2, 3 and 4). Respondents rated the perceived accuracy of nine true claims not addressed by the fact-checks on four-point scales from 1 (not at all accurate) to 4 (very accurate): 'A new loss of taste or smell is a symptom of COVID-19', 'Frequent hand washing is a way to protect against the coronavirus', 'Avoiding larger gatherings of people can help prevent the spread of the coronavirus', 'Coronavirus can be spread by people who do not show symptoms', 'A fever is a symptom of COVID-19', 'A dry cough is a symptom of COVID-19', 'Masks are an effective way to prevent the spread of the coronavirus' and 'There is no cure for COVID-19'. The British and Canada samples have identical measures. In the preregistration, we planned to include the item 'The medication remdesivir reduces the time required to recover from COVID-19'. This claim was thought true by the best-available evidence during waves 1–3 of the US survey but not in wave 4 of the US survey or the British or Canadian samples. To maximize consistency within and across samples, we therefore omitted this item from our analysis, including the index of true claims. (Analyses of scale reliability and factor loadings are provided in the Supplementary Information.)

Party (wave 1). In the United States sample, political party was a three-point scale of self-reported party identification from 1 (Democratic) to 2 (Independent) to 3 (Republican). In the United Kingdom, party consisted of two dichotomous variables. The first equals 1 if a respondent identifies with a left-wing party (Labour, Liberal Democrats, Plaid Cymru, Scottish National Party and Greens) and 0 otherwise. The second equals 1 if a respondent identifies with a right-wing party (Conservatives, Brexit Party and UKIP) and 0 otherwise. People who did not identify with a party were the omitted category. In the Canadian sample, party consisted of three dichotomous variables. The first equals 1 if a respondent identified with or leaned toward a left-wing party (Liberals, New Democratic Party, Bloc Quebecois and Green) and 0 otherwise. The second equals 1 if a respondent identified with or leaned toward the Conservative Party and 0 otherwise. People who did not identify with a party were the omitted category.

Ideology (wave 1). In the United States, we measured ideology via a seven-point scale of ideological identification from 1 (very liberal) to 4 (moderate; middle of the road) to 7 (very conservative). In the British and Canada samples, we measured ideology via an 11-point scale from 0 (left) to 10 (right).

Trust in health institutions (wave 1). In the United States sample, respondents filled out five items reporting the amount of trust they have in health institutions on a four-point scale from 1 (not at all) to 4 (a lot). Three concerned trust in governmental health institutions to handle the coronavirus outbreak. After receiving the prompt 'How much do you trust the following people and organizations to do the right thing to best handle the coronavirus outbreak?' respondents were asked about 'hospitals and doctors', 'scientists and researchers' and the 'Centers for Disease Control and Prevention (CDC)'. They also indicated their level of trust in information from governmental health institutions by responding using the same scale to the prompt 'How much, if at all, do you trust the information you get from...' for 'health experts in the state government?' and 'health experts in the federal government?' ($\alpha = 0.80$).

Participants in the British sample filled out five items measuring the amount of trust they have in health institutions on a four-point scale from 1 (not at all) to 4 (a lot). Four asked about trust in governmental health institutions using the prompt 'How much do you trust the following people and organizations to do the right thing to best handle the coronavirus outbreak?' Respondents were asked about 'hospitals and doctors', 'scientists and researchers', 'Public Health England' and 'SAGE (Scientific Advisory Group for Emergencies)'. They also were asked 'How much, if at all, do you trust information you get from health experts in the UK government?' We deviate from the preregistration to drop the item on trust in health experts in subnational governments (that is, Scotland and Wales) to make sure responses were comparable ($\alpha = 0.84$).

The Canadian sample filled out six items measuring the amount of trust they have in health institutions on a four-point scale from 1 (not at all) to 4 (a lot). Four concerned trust in governmental health institutions to handle the coronavirus outbreak. Respondents were asked 'How much do you trust the following people and organizations to do the right thing to best handle the coronavirus outbreak?' about 'hospitals and doctors', 'scientists and researchers', 'Public Health Agency of Canada' and 'Dr. Theresa Tam'. They also were asked 'How much, if at all, do you trust the information you get from...' about both 'health experts in your provincial government' and 'health experts in the Canadian government' ($\alpha = 0.91$ for each sample).

Trust in the media (wave 1). The American sample filled out two items on trust in the media on four-point scales from 1 (not at all) to 4 (a lot). Respondents were asked 'How much, if at all, do you trust the information you get from...' about 'national news organizations' and 'local news organizations' ($\alpha = 0.77$). The British sample was instead asked about 'news organizations (such as the BBC, The Daily Telegraph and The Guardian)?' and 'social media (such as Facebook, Twitter

and Instagram)?' In the Canadian sample, respondents were asked about 'news organizations (such as the CBC, Global News and the Globe and Mail)' and 'social media (such as Facebook, Twitter and Instagram)'. Because these items did not cohere into reliable indices (Great Britain, $\alpha = 0.28$; Canada sample 1, $\alpha = 0.25$; Canada sample 2, $\alpha = 0.24$), we deviated from the preregistration and solely used the 'news organizations' item in the British and Canada samples.

National leader affect (wave 1). Respondents were asked to rate their national leader on a four-point scale of approval from 1 (strongly disapprove) to 4 (strongly approve) and a 0–100 point feeling thermometer. American respondents filled out these items for then-President Donald Trump, British respondents for Prime Minister Boris Johnson and Canadian respondents for Prime Minister Justin Trudeau.

Conspiracy predispositions (wave 1). Respondents answered four questions measuring their general tendency to believe in conspiracy theories on a five-point scale from 1 (strongly disagree) to 5 (strongly agree): 'Much of our lives are being controlled by plots hatched in secret places', 'Even though we live in a democracy, a few people will always run things anyway', 'The people who really "run" the country are not known to voters' and 'Big events like wars, recessions and the outcomes of elections are controlled by small groups of people who are working in secret against the rest of us'[38]. In all samples, reliabilities were excellent (United States, $\alpha = 0.85$; Great Britain, $\alpha = 0.83$; Canada sample 1, $\alpha = 0.84$; Canada sample 2, $\alpha = 0.85$).

Analytic strategy. We estimate OLS regressions with robust standard errors. For covariate adjustment, we used a lasso variable selection procedure to determine the set of prognostic covariates to include in models for each dependent variable[39]. Specifically, we use a method that proceeds in two steps: (1) for each outcome variable, we fit a lasso-penalized regression model on the full data including all of our candidate covariates but excluding treatment assignment indicators; (2) estimate an OLS model for that DV using the variables selected in the first step as controls in addition to the treatment indicator. (Per our preregistration, we also report models estimated with no covariates in the online appendices; the results are almost always identical.) Eligible covariates were education, age group, gender, marital status, church attendance, region, party, ideology, living in a high incidence area, cognitive reflection test (CRT) score, political knowledge, race, trust in health institutions, trust in the media and lagged outcome measures from wave 1 (United States and Great Britain only; details above and in the Supplementary Information). As specified in our preregistration, all statistical tests are two-sided using a threshold of $P < 0.05$. Data distribution was assumed to be normal but this was not formally tested. Listwise deletion is used in the case of missing covariates.

To explore effect heterogeneity, we use non-parametric BCF[33]. BCFs are a sum-of-regression-trees approach, which entails partitioning the data to flexibly model the response surface. It in effect places separate informative prior distributions on the covariates (as selected above) and the treatment effect, which is modelled as a potential nonlinear function of a set of specified moderators. Since the prior distributions shrink estimates of both the prognostic covariates and the treatment effect, regularization is induced on the overall estimate of effect heterogeneity. This approach allows a principled way of guarding against false positives.

We tested RQ3 by comparing 95% CIs of BCF-estimated treatment effects among subgroups defined by median splits of a prespecified set of moderators: party identification, leader approval, leader feeling thermometer, trust in health institutions, conspiracy predispositions, media trust and prior belief in targeted misperceptions. We focus above on treatment effect CIs that do not overlap 0 for relevant moderator values.

We report estimated separate treatment effects for each claim in Supplementary Tables 39–52, 57–70 and 75–88.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Replication data are publicly available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6FIHZJ.

Code availability
Replication code are publicly available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6FIHZJ.

References
1. Bridgman, A. et al. The causes and consequences of COVID-19 misperceptions: understanding the role of news and social media. *Harvard Kennedy School Misinf. Rev.* https://doi.org/10.37016/mr-2020-028 (2020).

2. Lammers, J., Crusius, J. & Gast, A. Correcting misperceptions of exponential coronavirus growth increases support for social distancing. *Proc. Natl Acad. Sci. USA* **117**, 16264–16266 (2020).

3. Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K. & Larson, H. J. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav.* **5**, 337–348 (2021).

4. Chan, M.-pS., Jones, C. R., Hall Jamieson, K. & Albarracín, D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol. Sci.* **28**, 1531–1546 (2017).

5. Walter, N., Cohen, J., Holbert, R. L. & Morag, Y. Fact-checking: a meta-analysis of what works and for whom. *Polit. Commun.* **37**, 350–375 (2020).

6. Walter, N., Brooks, J. J., Saucier, C. J. & Suresh, S. Evaluating the impact of attempts to correct health misinformation on social media: a meta-analysis. *Health Commun.* **36**, 1776–1784 (2021).

7. Nyhan, B. Why the backfire effect does not explain the durability of political misperceptions. *Proc. Natl Acad. Sci. USA* **118**, e1912440117 (2021).

8. Merkley, E. et al. A rare moment of cross-partisan consensus: elite and public response to the COVID-19 pandemic in Canada. *Can. J. Polit. Sci.* **53**, 311–318 (2020).

9. Mordecai, M. & Connaughton, A. *Public Opinion about Coronavirus is More Politically Divided in US than in other Advanced Economies* (Pew Research Center, 2020).

10. Pennycook, G., McPhetres, J., Bago, B. & Rand, D. G. Predictors of attitudes and misperceptions about COVID-19 in Canada, the UK, and the USA. *Pers. Soc. Psychol. Bull.* https://doi.org/10.1177/01461672211023652 (2021).

11. Kunda, Z. The case for motivated reasoning. *Psychol. Bull.* **108**, 480 (1990).

12. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Polit. Sci.* **50**, 755–769 (2006).

13. Carnahan, D., Bergan, D. E. & Lee, S. Do corrective effects last? Results from a longitudinal experiment on beliefs toward immigration in the US. *Polit. Behav.* **43**, 1227–1246 (2021).

14. Porter, E. & Wood, T. J. The global effectiveness of fact-checking: evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proc. Natl Acad. Sci. USA* **118**, e2104235118 (2021).

15. Chong, D. & Druckman, J. N. Dynamic public opinion: communication effects over time. *Am. Polit. Sci. Rev.* **104**, 663–680 (2010).

16. Gerber, A. S., Gimpel, J. G., Green, D. P. & Shaw, D. R. How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *Am. Polit. Sci. Rev.* **105**, 135–150 (2011).

17. Hill, S. J., Lo, J., Vavreck, L. & Zaller, J. How quickly we forget: the duration of persuasion effects from mass communication. *Polit. Commun.* **30**, 521–547 (2013).

18. Hardt, O., Nader, K. & Nadel, L. Decay happens: the role of active forgetting in memory. *Trends Cogn. Sci.* **17**, 111–120 (2013).

19. Fazio, L. K., Brashier, N. M., Payne, B. K. & Marsh, E. J. Knowledge does not protect against illusory truth. *J. Exp. Psychol.* **144**, 993 (2015).

20. Pennycook, G., Cannon, T. D. & Rand, D. G. Prior exposure increases perceived accuracy of fake news. *J. Exp. Psychol.* **147**, 1865 (2018).

21. Cacioppo, J. T. & Petty, R. E. Effects of message repetition and position on cognitive response, recall, and persuasion. *J. Pers. Soc. Psychol.* **37**, 97 (1979).

22. Arendt, F. Toward a dose–response account of media priming. *Commun. Res.* **42**, 1089–1115 (2015).

23. Pennycook, G. et al. Shifting attention to accuracy can reduce misinformation online. *Nature* **592**, 590–595 (2021).

24. Carey, J. M., Chi, V., Flynn, D., Nyhan, B. & Zeitzoff, T. The effects of corrective information about disease epidemics and outbreaks: evidence from Zika and Yellow Fever in Brazil. *Sci. Adv.* **6**, eaaw7449 (2020).

25. Freeze, M. et al. Fake claims of fake news: political misinformation, warnings, and the tainted truth effect. *Polit. Behav.* **43**, 1433–1465 (2021).

26. Pennycook, G., Bear, A., Collins, E. T. & Rand, D. G. The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Manag. Sci.* **66**, 4944–4957 (2020).

27. Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. Affective polarization, local contexts and public opinion in America. *Nat. Hum. Behav.* **5**, 28–38 (2021).

28. Druckman, J. N. et al. The role of race, religion, and partisanship in misperceptions about COVID-19. *Group Process. Intergroup Relat.* **24**, 638–657 (2021).

29. Gadarian, S., Goodman, S. W. & Pepinsky, T. Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. *PLoS ONE* **16**, e0249596 (2021).

30. Grossman, G., Kim, S., Rexer, J. M. & Thirumurthy, H. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proc. Natl Acad. Sci. USA* **117**, 24144–24153 (2020).

31. Taylor, S. & Asmundson, G. J. Negative attitudes about facemasks during the COVID-19 pandemic: the dual importance of perceived ineffectiveness and psychological reactance. *PLoS ONE* **16**, e0246317 (2021).

32. Pickup, M., Stecula, D. & Van Der Linden, C. Novel coronavirus, old partisanship: COVID-19 attitudes and behaviours in the United States and Canada. *Can. J. Political Sci.* **53**, 357–364 (2020).

33. Hahn, P. R., Murray, J. S. & Carvalho, C. M. et al. Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Anal.* **15**, 965–1056 (2020).

34. Green, D. P. & Kern, H. L. Modeling heterogeneous treatment effects in survey experiments with Bayesian additive regression trees. *Public Opin. Q.* **76**, 491–511 (2012).

35. Funke, D. Archived fact-check: Tucker Carlson guest airs debunked conspiracy theory that COVID-19 was created in a lab. *PolitiFact* (16 September 2020); https://www.politifact.com/li-meng-yan-fact-check/

36. Kessler, G. Timeline: how the Wuhan lab-leak theory suddenly became credible. *Washington Post* (25 May 2021); https://www.washingtonpost.com/politics/2021/05/25/timeline-how-wuhan-lab-leak-theory-suddenly-became-credible/

37. De Benedictis-Kessner, J., Baum, M. A., Berinsky, A. J. & Yamamoto, T. Persuading the enemy: estimating the persuasive effects of partisan media with the preference-incorporating choice and assignment design. *Am. Polit. Sci. Rev.* **113**, 902–916 (2019).

38. Uscinski, J. E., Klofstad, C. & Atkinson, M. D. What drives conspiratorial beliefs? The role of informational cues and predispositions. *Political Res. Q.* **69**, 57–71 (2016).

39. Bloniarz, A., Liu, H., Zhang, C.-H., Sekhon, J. S. & Yu, B. Lasso adjustments of treatment effect estimates in randomized experiments. *Proc. Natl Acad. Sci. USA* **113**, 7383–7390 (2016).

Acknowledgements

Author contributions

J.M.C., A.M.G., B.N. and J.R. designed research. J.M.C., A.M.G., P.J.L., E.M., B.N., J.R. and J.P. designed country surveys. A.M.G. and J.P. analysed data. J.M.C., B.N., J.P. and J.R. wrote the paper. J.M.C., A.M.G., P.J.L., E.M., B.N., J.R. and J.P. revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01278-3.

Correspondence and requests for materials should be addressed to Brendan Nyhan.

Peer review information *Nature Human Behaviour* thanks Toby Bolsen, Dustin Carnahan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature portfolio

Corresponding author(s): Brendan Nyhan

Last updated by author(s): Nov. 12, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected via YouGov's online platform and Qualtrics.
Data analysis	Data was analyzed using R 4.0.4.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Replication data and code are available on Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6FIHZJ.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☒ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We investigate the effects of exposure to fact-checks on COVID-19 pandemic misperceptions using parallel pre-registered experiments testing in the United States (US), Great Britain (GB), and Canada. These countries combine demographic similarity with sharp differences in the extent to which the COVID-19 response has become politicized and in the severity of the pandemic at the time the studies were fielded, allowing us to provide one of the most systematic tests to date of whether the effects of fact-checking generalize across different political and social contexts.
Research sample	We collected data from opt-in Internet survey panels that either use a weighting and matching algorithm to approximate a demographically representative sample (US and GB) or use quota-based sampling (Canada). In the US and GB samples, we recruited three distinct samples: a national representative sample, an oversample of respondents from the 10% of zip codes (US) or local authorities (GB) with the highest per capita COVID mortality rates, and a sample of respondents who have additionally consented to provide digital trace data of their web traffic histories.
Sampling strategy	Participants in our samples were selected by our survey vendors.
Data collection	Participants entered survey responses on computers or portable devices. Researchers were not present as responses were collected.
Timing	United States: Wave 1: May 20-June 3, 2020 Wave 2: June 25-July 12, 2020 Wave 3: July 28-August 19, 2020 Wave 4: March 9-23, 2021 Great Britain: Wave 1: September 11-29, 2020 Wave 2: December 10-23, 2020 Wave 3: February 4-22, 2021 Canada: Sample 1: December 15, 2020-January 13, 2021 Sample 2: February 3-28, 2021
Data exclusions	We use listwise deletion of cases with missing data. Other than missing data, there are no data exclusions to report.
Non-participation	YouGov provided data on panel members who elected to take part in the survey in the United States and Great Britain. Data are not available on participants who declined to participate. The data we observe for Canada similarly only include respondents invited by Dynata who elected to participate in the survey.
Randomization	Randomization into conditions was programmed into the survey software (investigators were therefore blind to assignment until after data collection was complete).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☐	☒ Human research participants
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Adults in the US, Great Britain, and Canada
Recruitment	Recruited by online survey firms (YouGov and Dynata). Compensation was provided by survey firms. All respondents provided informed consent before participating in each wave of the survey.
Ethics oversight	These studies were approved by human ethics review boards at Dartmouth College (STUDY00032068), Princeton University (IRB #12859), the University of Toronto (Protocol #00040160). The University of Exeter reviewed and approved the Great Britain study (SSIS Ethics Committee #201920-148), and recognized the approved protocols for the North American studies.

Note that full information on the approval of the study protocol must also be provided in the manuscript.