



City Research Online

City, University of London Institutional Repository

Citation: Ruggeri, K., Stock, F., Haslam, S. A., Capraro, V., Boggio, P., Ellemers, N., Cichocka, A., Douglas, K. M., Rand, D. G., van der Linden, S., et al (2024). A synthesis of evidence for policy from behavioural science during COVID-19. *Nature*, 625(7993), pp. 134-147. doi: 10.1038/s41586-023-06840-9

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/31877/>

Link to published version: <https://doi.org/10.1038/s41586-023-06840-9>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

A synthesis of evidence for policy from behavioural science during COVID-19

<https://doi.org/10.1038/s41586-023-06840-9>

Received: 18 October 2022

Accepted: 6 November 2023

Published online: 13 December 2023

Open access

 Check for updates

Scientific evidence regularly guides policy decisions¹, with behavioural science increasingly part of this process². In April 2020, an influential paper³ proposed 19 policy recommendations ('claims') detailing how evidence from behavioural science could contribute to efforts to reduce impacts and end the COVID-19 pandemic. Here we assess 747 pandemic-related research articles that empirically investigated those claims. We report the scale of evidence and whether evidence supports them to indicate applicability for policymaking. Two independent teams, involving 72 reviewers, found evidence for 18 of 19 claims, with both teams finding evidence supporting 16 (89%) of those 18 claims. The strongest evidence supported claims that anticipated culture, polarization and misinformation would be associated with policy effectiveness. Claims suggesting trusted leaders and positive social norms increased adherence to behavioural interventions also had strong empirical support, as did appealing to social consensus or bipartisan agreement. Targeted language in messaging yielded mixed effects and there were no effects for highlighting individual benefits or protecting others. No available evidence existed to assess any distinct differences in effects between using the terms 'physical distancing' and 'social distancing'. Analysis of 463 papers containing data showed generally large samples; 418 involved human participants with a mean of 16,848 (median of 1,699). That statistical power underscored improved suitability of behavioural science research for informing policy decisions. Furthermore, by implementing a standardized approach to evidence selection and synthesis, we amplify broader implications for advancing scientific evidence in policy formulation and prioritization.

Scientific evidence has an important role in policy decisions¹. This has been increasingly true of evidence from the behavioural and social sciences², particularly for public health policy throughout the global COVID-19 pandemic^{4,5}. One broad challenge in this process is that there is no universally endorsed approach to determine which scientific insights should inform policy⁶. Recommendations may be made on an ad hoc basis, may be based on relationships between certain researchers and policymakers, and may fail to factor in an appropriate level of uncertainty⁷⁻⁹. This is further complicated by the sheer volume and heterogeneity of evidence, making appropriate identification and synthesis a substantial challenge¹⁰.

One major example of science impacting policy comes from April 2020, when 42 academics from 8 countries and multiple academic disciplines published a review containing a series of hypotheses about factors that were likely to shape collective behaviour during a pandemic³. Topics included threat and risk perception, social norms, science communication, emphasizing the importance of individual and collective interests, leadership, stress and coping. The paper also included a list of broad behavioural insights (which we refer to as 'claims') deemed most relevant to the pandemic. The article received unprecedented attention: in only 2 years, it was cited over 3,000 times, and by December 2022 had an Altmetric score in the highest 0.0001% of all articles ever published. This was, in part, because governments around the world formulated pandemic policy strategies explicitly¹¹ on the basis of the behavioural concepts highlighted in the paper¹²⁻¹⁸. However, it was

also because social and behavioural scientists viewed the pandemic as a critical focus for their attention¹⁹ and research^{20,21}.

Naturally, with such levels of visibility, concerns were raised about various claims made by Van Bavel et al.³. This led to some concerns about academics making recommendations despite not all having previous experience in the domain of public policy or public health^{22,23}. In the context of the so-called replication crisis in psychology, it has also been suggested that the article focused too much on evidence from WEIRD (Western, educated, industrialized, rich, democratic) populations, took insufficient account of the heterogeneity of effects, overstated the validity of existing evidence and was opportunistic²⁴⁻²⁷.

This paper responds to those concerns by retrospectively evaluating the quality of the claims by Van Bavel et al.³. Although, more broadly, it presents a valuable opportunity to mobilize many independent experts using a structured approach to assess the appropriateness of scientific evidence being considered for application to policy. This is important not only to inform theorizing about behaviour during a pandemic but also to establish the relevance of these claims for future emergencies and more generally across public policy. It is an equally important complement to addressing replicability in science, by scrutinizing evidence produced during a crisis while also revisiting previous claims made by experts.

Concerns related to the readiness or robustness of evidence for application to policy are neither new nor unique to COVID-19. Terms such as 'evidence-based policy' have long been used to apply research to major

Analysis

Table 1 | Nineteen claims about social and behavioural science during the COVID-19 pandemic

2020 Claim wording	Revised claim wording	Behavioural theme
(1) A shared sense of identity or purpose can be encouraged by addressing the public in collective terms and by urging 'us' to act for the common good	There is a small positive association between collective identity and behaviour for the common good, but the relationship depends on the level of identity activated (for example, nation versus European Union)	Sense of identity
(2) Identifying trusted sources (for example, local, religious or community leaders) that are credible to different audiences to share public health messages can be effective	Identifying trusted sources (for example, local, religious, political or community leaders) that are credible to different audiences to share public health messages can be effective in increasing intentions to engage in recommended health behaviours	Trust and leadership
(3) Leaders and the media might try to promote cooperative behaviour by emphasizing that cooperating is the right thing to do and that other people are already cooperating	Emphasizing cooperation and highlighting the cooperative behaviour of other people can encourage people to adhere to public health recommendations, although effects may be small	Trust and leadership
(4) Norms of prosocial behaviour are more effective when coupled with the expectation of social approval and modelled by in-group members who are central in social networks	Surveys have shown that descriptive norms, especially when enacted by close reference groups, are associated with greater compliance with public health recommendations and self-reported prosocial behaviours	Sense of identity
(5) Leaders and members of the media should highlight bipartisan support for COVID-related measures, when they exist, as such endorsements in other contexts have reduced polarization and led to less-biased reasoning	Where polarization regarding public health behaviours exists, endorsement from bipartisan coalitions can be effective in reducing polarization and increasing compliance	Messaging and language
(6) There is a need for more targeted public health information within marginalized communities and for partnerships between public health authorities and trusted organizations that are internal to these communities	Marginalized communities have very different risks and health outcomes and may receive different information through different channels, suggesting the potential benefit of targeted communication and strategies	Messaging and language
(7) Messages that (i) emphasize benefits to the recipient, (ii) focus on protecting others, (iii) align with the moral values of recipients, (iv) appeal to social consensus or scientific norms, and/or (v) highlight the prospect of social group approval tend to be persuasive	Messages may be more effective when they align closely with the moral values of recipients, appeal to social consensus or scientific norms, and highlight group approval	Messaging and language
(8) Given the importance of slowing infections, it may be helpful to make people aware that they benefit from others' access to preventative measures	There is suggestive, albeit little, empirical evidence that it can help to make people aware that they benefit from others' access to preventative measures	Sense of identity
(9) Preparing people for misinformation and ensuring they have accurate information and counterarguments against false information before they encounter conspiracy theories, fake news or other forms of misinformation can help to inoculate them against false information	Preparing people for misinformation before they encounter conspiracy theories, fake news or other forms of misinformation — for example, by ensuring that they have accurate information and counterarguments against false information, or by prompting them to consider accuracy — can help to reduce belief in, and/or sharing of, false information for a limited time	Social cohesion and misinformation
(10) Use of the term 'social distancing' might imply that one needs to cut off meaningful interactions. A preferable term is 'physical distancing', because it allows for the fact that social connection is possible even when people are physically separated	Although 'physical distancing' is a more accurate term than 'social distancing' and may encourage more social connection, there is no evidence on whether it is more effective in encouraging public health behaviours	Messaging and language
(11) As negative emotions increase, people may rely on negative information about COVID-19 more than other information to make decisions. In the case of strong emotional reactions, people may also ignore important numeric information such as probabilities and the scope of a problem	An increase in negative emotions related to the pandemic may influence behaviour and decision-making and lead people to ignore important information, such as probabilities of negative outcomes or actual risk level	Messaging and language
(12) Cultures accustomed to prioritizing freedom over security may also have more difficulty coordinating in the face of a pandemic	Strong correlations indicate that cultures accustomed to prioritizing freedom over security may also have more difficulty coordinating in the face of a pandemic	Social cohesion and misinformation
(13) Fake news, conspiracy theories and misinformation will have a negative effect on vaccine hesitancy	Evidence has shown that fake news, conspiracy theories and misinformation were negatively associated with vaccination intentions, but the effect on actual vaccination behaviour has not been shown	Social cohesion and misinformation
(14) Unmitigated political polarization will disrupt or create other negative effects on attempts to minimize or end the pandemic	Evidence has shown that divergent partisan identities lead to significantly different opinions and reported behaviours in response to the pandemic, undermining coordination efforts to minimize or end the pandemic	Social cohesion and misinformation
(15) Active use of online connections can reduce some negative mental and other health effects created by isolation policies	Active social connections online can buffer against negative mental health effects, although mitigating effects may be small	Sense of identity

Claims in the left column show the original wording from Van Bavel et al.³. The text in the right column shows updated wording after assessing the evidence in 2022. Note that claim 7 has five components, making a total of 19 for the table.

decisions in government, institutions, schools and businesses²⁸. The lack of consensus concerning what counts as sufficient evidence for policy decisions²⁹ is a particular problem in an emergency when policymakers must take urgent actions with sometimes limited evidence. As such, they may seek the lowest-risk or most-effective approach rather than a perfectly informed approach (which is often unavailable

or unclear)³⁰, or they may risk delays that create even greater harm^{31,32}. Consequently, the COVID-19 pandemic presented an opportunity to assess evidence in a way that encouraged more evaluations of academic policy recommendations in the future.

With such a substantial amount of evidence now available, our goal was to provide a descriptive synthesis of evidence available relevant

Table 2 | Evidence assessments for four claims on sense of identity

Claim (number)	Evidence	Level	Direction	Effect size	Summary of evidence		
Active use of online connections can reduce some negative mental and other health effects created by isolation policies (15)	Existing studies point towards small positive effects supporting this claim, but the number of studies is insufficient. One longitudinal study found small-to-moderate effects in the real world	Tested in real-world or field studies	Positive	Small	Articles reviewed: 15	Sample range: 110–6,523	
					Average review time: 16h (spread over 3–10 days)	Mean sample: 2,062.8	Median sample: 1,344
Norms of prosocial behaviour are more effective when coupled with the expectation of social approval and modelled by in-group members who are central in social networks (4)	Evidence generally supports the notion with a medium-to-large positive effect. However, the available studies assess the general effect of norms, not the specific context stated in the claim	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Large	Articles reviewed: 45	Sample range: 52–108,075	
					Average review time: 10h (spread over 2–5 days)	Mean sample: 10,709.3	Median sample: 2,006.5
A shared sense of identity or purpose can be encouraged by addressing the public in collective terms and by urging 'us' to act for the common good (1)	Claim is generally supported; however, it lacks real-world assessments with observable outcomes. Evidence mostly stems from survey data and online experiments	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 54	Sample range: 130–910,006	
					Average review time: 20h (spread over 2–15 days)	Mean sample: 25,532.3	Median sample: 729.5
Given the importance of slowing infections, it may be helpful to make people aware that they benefit from others' access to preventative measures (8)	Clearly supported by limited evidence available, although best-evidence focus tended to be on intentions rather than true behaviours	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 5	Sample range: 134–1,373	
					Average review time: 2h (spread over 1 day)	Mean sample: 913.0	Median sample: 1,232

Overview of ratings and assessments for four claims (1, 4, 8 and 15). In addition, compilation and screening of articles were estimated to have taken over a total of 200h. Note that summaries of sample sizes included any studies that included evidence and a rating, irrespective of overall influence on summary assessment. Articles may be double-counted if they were directly relevant for multiple claims.

to 19 claims from the Van Bavel et al. article (see the first column in Table 1 from this paper for the full statements). In essence, we evaluated the extent to which those statements provided valid policy guidance, privileging empirical evidence of consistent real-world impact. The present work is not a comprehensive evaluation of all behavioural science related to the pandemic, which would be beyond the scope of a single article, but only of the key statements made in the 2020 paper.

We assessed 19 behavioural policy recommendations through evaluating available articles based on the level of evidence they include. Ratings range from purely opinion or theory to large-scale, replicated field studies, as well as the size and direction of effects reported (see 'Procedure' in the Methods section). Our evaluations primarily focused on the scale and scope of empirical findings directly related to the claims, although the compiled data (available at <https://psyarxiv.com/58udn>) also highlights methods, geographical settings and specific behaviours. We then synthesized the evidence within each claim to formulate a summary evaluation. For this exercise, we included both original authors and an independent team of evaluators to select and assess evidence relevant to these claims (see 'Evaluation teams' in the Methods section), all of whom were blinded to names and assessments. This allowed us to leverage the expertise of the original authors while also adding a diverse group of scholars who were not involved in the original paper to provide an independent, objective evaluation of the evidence.

Our primary motivations are to (1) transparently evaluate ex post evidence for a set of highly influential claims regarding behaviour during a pandemic, (2) implement a pragmatic, expert-driven method for evaluating and synthesizing evidence that is suitable for informing public policy (both related to COVID-19 and future applications), and (3) make those assessments public in a way that promotes transparency and builds trust with the public³³. The first and second aims are broadly

relevant across all scientific research. The third aim is specifically relevant to assessing policy recommendations, which is why we decided to provide a descriptive summary rather than focus on methods or causal inference (more highly valued in science). The first aim is also especially critical given substantial concerns about public trust in science in general³⁴ and raised directly in the context of COVID-19 (refs. 35–39).

Evaluating predictions of academic experts is an important exercise to protect against questionable research practices, mistakes and overconfidence^{38,40,41}, which were a common concern specifically in behavioural science during COVID-19. Those concerns fed into cautions raised about systematic reviews of available evidence^{42,43}. Furthermore, by mobilizing a large group of fully independent reviewers, we ensure that no single paper (whether highly powered or merely highly visible) or person can have unchecked influence on the evaluation of policy-relevant evidence.

Evidence for 19 pandemic behaviour claims

As outlined in the Methods ('Procedure and evidence used for evaluations'), all 747 articles were reviewed by at least two reviewers from each team (four total); 518 articles received at least one rating (see Extended Data Fig. 1). One-hundred and eighty-six articles were unanimously rated as not directly relevant or informative to the claim; 43 were found to be duplicated work, typically papers that had changed titles from preprint to publication. Of the 19 claims, 18 had at least some empirical evidence to assess (see full descriptions in Tables 2–6), with only one claim lacking any empirical research. Of those 18 claims, 13 claims were assessed as having been studied empirically, although only in surveys or limited laboratory settings (see Extended Data Table 1 for a breakdown of articles).

Table 3 | Evidence assessments for two claims on trust and leadership

Claim (number)	Evidence	Level	Direction	Effect size	Summary of evidence
Identifying trusted sources (for example, local, religious or community leaders) that are credible to different audiences to share public health messages can be effective (2)	General support for the claim with a medium effect size from survey data in different samples and some applications in the real world. The core claim is generally supported by the evidence	Replicated real-world or field study evidence	Positive	Medium	Articles reviewed: 40 Average review time: 19 h (spread over 2–10 days) Sample range: 372–1,429,453 Mean sample: 46,892.5 Median sample: 1,765
Leaders and the media might try to promote cooperative behaviour by emphasizing that cooperating is the right thing to do and that other people are already cooperating (3)	Evidence for the claim stems mostly from correlational data and few experimental studies reporting small but rather inconsistent effects across contexts and outcomes	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 16 Average review time: 13 h (spread over 8–10 days) Sample range: 52–484,239 Mean sample: 40,719.5 Median sample: 1,004

Overview of ratings and assessments for two claims (2 and 3). Articles may be double-counted if they were directly relevant for multiple claims.

Thirty-four studies report samples related to number of countries, studies, secondary datasets or other indirect observations. Sample sizes of the 463 original data studies included were large. The mean sample size of 418 papers specifically involving human participants was 16,848 (median of 1,699), with individual studies ranging from 52 to 1,429,453 participants. We present somewhat conservative estimates for both mean and median by including only those specifically involving human participants. We do not include in these estimates three studies (with samples from 3.7 million to 654 million) that used social media posts or accounts. Many studies also provided only vague indicators (for example, “more than ten thousand”) or aggregated groupings (for example, numbers of states, provinces or countries). Links in Data Availability give access to raw and interactive datasets for further exploration. However, only three studies reviewed had samples of fewer than 100 participants, whereas 279 had 1,000 or more. One-hundred and forty-two countries were included in one or more studies (see Evidence used for evaluations in Methods).

As depicted in Fig. 1, the direction of effect or correlation suggested by most claims were generally supported. Of the 18 claims that had at least some empirical evidence available for evaluation, 16 (89%) claims were generally supported in the direction of the original statement. Of the 16 claims that were supported by the research literature, ten were considered to show small effects, five were considered to show medium effects and one was considered to show a large effect. We did not find any meaningful effects in support of the remaining two claims, which stated that messages that emphasize benefits to the recipient (claim 7i) and focus on protecting others (claim 7ii) tend to be persuasive. This may be because several studies showed that there were moderators of which emphasis (self versus others) was more effective⁴⁴.

Six claims (2, 4, 9, 12, 13 and 14) backed by empirical evidence demonstrated medium or large effects. Claim 4 (that norms of prosocial behaviour are more effective when coupled with the expectation of social approval and modelled by in-group members who are central in social networks) was mostly tested on observational data, but the effect sizes found in these data were notably strong.

Importantly, no effects were in the opposite direction from the original predictions. This means that no recommendations from the Van Bavel et al. paper led to a consistent backfire effect. Those 19 statements proposed behavioural domains that were likely to be of interest during the COVID-19 pandemic. Some claims were general about potentially relevant behaviours, whereas others were more prescriptive about potentially more effective intervention approaches.

Overall, our review indicates that the Van Bavel et al. article generally identified highly relevant topics of study in the pandemic and, to an extent, the direction of associated findings. In particular, it identified

(1) relevant behaviours during the pandemic (both positive and negative), (2) likely barriers to mitigating the spread of the disease, and (3) major social challenges that would be faced by policymakers. The following text summarizes these in general groups, citing articles viewed by assessing teams as being most consequential for their final assessments. The ratings for each are specified in Tables 2–6.

All behaviours studied, whether specific to a claim or not, are listed in Table 7. We have included this table as a reference in the future for considering behaviours to expect or target.

Sense of identity

Four claims made in 2020 focused on how social identities would be highly relevant during the pandemic, particularly how they aligned with either community benefits or social norms. These expectations generally appeared to be accurate (see Table 2), with scores of studies concluding that connectedness with communities or aligning with morals were a predictor of behaviours and efforts to control the spread of illness^{45–60}. However, one challenge that is typically present for research on subjective and latent constructs such as identity, prosociality and connectedness is that most research was conducted through surveys. Few studies attempted to isolate the causal effect of identity on pandemic behaviours, and no experimental studies manipulated identity or the sense of collective purpose in a real-world setting. In some cases, well-powered studies directly assessing the claims were limited to asking about intentions to receive vaccines⁶⁰. Although such findings are very valuable, there is clearly additional benefit in validating those findings in consequential settings, or even carrying out retrospective studies to determine whether behaviours or infections were measurably associated with connectedness where studies were conducted.

Trust and leadership

There was a large amount of evidence from peer-reviewed research on the role of leadership during the pandemic. Two claims from 2020 specifically outlined expectations for how trusted sources and leadership may be relevant to promoting public health guidelines. There was a substantial amount of research (see Table 3) supporting both of these expectations, although the best-quality evidence from consequential settings was replicated only in relation to the claim that the most effective messaging comes from trusted sources^{61–66}. Consistent with original expectations, evidence has supported the value of highlighting the cooperation of other people to promote health behaviours, although evidence was limited to surveys and correlational studies.

Table 4 | Evidence assessments for four claims on messaging and language

Claim (number)	Evidence	Level	Direction	Effect size	Summary of evidence		
Leaders and members of the media should highlight bipartisan support for COVID-19-related measures, when they exist, as such endorsements in other contexts have reduced polarization and led to less-biased reasoning (5)	The one reviewed paper directly testing the claim generally supports it, finding that bipartisan policy endorsements reduce polarization in views of, and increase overall support for, COVID mitigation policies	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 12	Sample range: 350–10,699	
					Average review time: 7h (spread over 1–10 days)	Mean sample: 2,724.2	Median sample: 1,995
There is a need for more targeted public health information within marginalized communities and for partnerships between public health authorities and trusted organizations that are internal to these communities (6)	Empirical evidence for the core of the claim exists; however, there is little evidence available that tests the effectiveness of the suggested approach. Existing studies have suggested a small positive effect of targeted messaging	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 19	Sample range: 54–140,184	
					Average review time: 12h (spread over 3–8 days)	Mean sample: 16,758.6	Median sample: 991
As negative emotions increase, people may rely on negative information about COVID-19 more than other information to make decisions. In the case of strong emotional reactions, people may also ignore important numerical information such as probabilities and the scope of a problem (11)	No empirical evidence that empirically tested the full claim. Existing evidence has focused more on the second part of the claim, not the first part, and might broadly point towards a small effect	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 34	Sample range: 155–125,306	
					Average review time: 7h (spread over 3–5 days)	Mean sample: 6,635.4	Median sample: 1,237
Use of the term ‘social distancing’ might imply that one needs to cut off meaningful interactions. A preferable term is ‘physical distancing’ because it allows for the fact that social connection is possible even when people are physically separated (10)	Besides a few small survey studies, support for the claim is purely based on theory and opinion. Therefore, no statement can be made about the potential effect size of this claim in application	No evidence has been identified, only discussion of the theory	NA	NA	Articles reviewed: 8	Sample range: NA	
					Average review time: 9h (spread over 3–5 days)	Mean sample: NA	Median sample: NA

Overview of ratings and assessments for four claims (5, 6, 10 and 11). Articles may be double-counted if they were directly relevant for multiple claims. NA, not applicable due to neither of two studies with data being directly relevant to claim.

Messaging and language

Perhaps the most widely studied topic during the pandemic was public health messaging. This was clearly anticipated by Van Bavel et al. as 9 of the 19 claims explicitly discussed the role of messaging and language in developing effective public health interventions. Not surprisingly, this also produced the most heterogeneous set of evidence ratings (see Tables 4 and 5). Observational studies in natural settings concluded that messages directly emphasized benefits to individuals or protecting others had no measurable effect on behaviours⁶⁷. There was some evidence for the effect of benefit-based approaches on behavioural intentions, although some studies suggested that self-benefit versus other-benefit messages were differentially effective for different types of people. Although benefit-based messaging has been found to be effective in general⁶⁸, it is possible that there may be limits to its effect on behaviour in the context of novel health threats where the benefits of preventative behaviours are not well established and recommendations are evolving.

Messaging related to partisan concerns was also widely studied, although not often in consequential settings. For the studies with the most policy-relevant evidence, messages emphasizing consensus and general agreement about public health behaviours were more effective in promoting these behaviours than those considered to be polarizing or partisan in nature (in survey studies)⁶⁹. A small number of related studies in the context of marginalized communities has found that direct engagement and direct messaging were more effective⁷⁰.

Claim 10 on the use of physical distancing being preferred to social distancing yielded no evidence. Although eight articles were identified

and required some time to review, none included direct evidence of any effect on making this change, and we have excluded it from Table 4.

Social cohesion and misinformation

Claims specifically related to polarization and flawed sources of information were widely validated, with some caveats⁷¹. Across more than 200 published articles, polarizing and disingenuous messaging were consistently associated (see Table 6) with negative outcomes in terms of the effectiveness of public health interventions⁷². However, direct causal evidence was relatively scarce⁷³. Studies with the highest levels of evidence have validated these patterns in consequential settings, often with medium effect sizes, indicating that greater division in messaging and lack of social cohesion were associated with lower effectiveness of public health messaging^{48,73–85}. Encouragingly, inoculating against manipulation techniques⁸⁶ and prompting users to consider accuracy before sharing news⁸⁷ have some positive effects. Again, both lines of study would benefit from replications in consequential settings, with some additional validating work on this emerging after our review started⁸⁸.

Major themes not explicitly assessed

Several themes emerged during the search phase of this project that were discussed by Van Bavel et al. but not necessarily formalized in terms of a specific claim. These included the clear relevance of threat and risk perception, the role of inequality⁸⁹ and racism⁹⁰, skepticism towards science⁹¹, incentivizing behaviours beyond simply describing

Analysis

Table 5 | Evidence assessments for claim 7 (five sub-claims) on messaging and language

Claim (number)	Evidence	Level	Direction	Effect size	Summary of evidence		
Messages that emphasize benefits to the recipient tend to be persuasive (7i)	Although some online experiments find limited support for the claim, the general picture is mixed. Existing applications in the real world indicate no general support for the claim	Replicated real-world or field-study evidence	Null	Null	Articles reviewed: 141 (claim 7 total)	Sample range: 208–163,627	
					Average review time: 22h (spread over 2–14 days)	Mean sample: 19,599.1	Median sample: 3,964
Messages that appeal to social consensus or scientific norms tend to be persuasive (7iv)	The evidence is mixed, with online experiments and survey studies finding small positive effects of the suggested approach. However, these findings could not be replicated in an existing field study	Tested in real-world or field studies	Positive	Small	Articles reviewed: 141 (claim 7 total)	Sample range: 324–163,627	
					Average review time: 22h (spread over 2–14 days)	Mean sample: 17,901.2	Median sample: 2,358.5
Messages that focus on protecting others tend to be persuasive (7ii)	Some real-world studies are available for the claim and point towards no effect. Evidence from online experiments has also been mixed, although some studies have found small positive effects	Tested in real-world or field studies	Null	Null	Articles reviewed: 141 (claim 7 total)	Sample range: 200–163,627	
					Average review time: 22h (spread over 2–14 days)	Mean sample: 13,887.6	Median sample: 2,459
Messages that align with the moral values of the recipient tend to be persuasive (7iii)	No real-world studies with behavioural measures exist, but the existing evidence from survey data and online experiments has mostly suggested a small positive effect	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 141 (claim 7 total)	Sample range: 246–24,682	
					Average review time: 22h (spread over 2–14 days)	Mean sample: 5,019.8	Median sample: 1,683
Messages that highlight the prospect of social group approval tend to be persuasive (7v)	Few studies have tested this claim. The online experiments that do exist indicate mixed results	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Small	Articles reviewed: 141 (claim 7 total)	Sample range: 324–24,682	
					Average review time: 22h (spread over 2–14 days)	Mean sample: 4,974.4	Median sample: 1,384

Overview of ratings and assessments for claim 7 (with five sub-claims), ordered by the level of evidence identified. Articles may be double-counted if they were directly relevant for multiple claims. Time estimates are given a single value for claim 7 due to the overlap of articles and the split in reviewers, but should be treated separately (that is, the total time spent reviewing all papers for claim 7 alone was over 100h).

benefits (for example, by providing financial rewards for vaccination)^{92–94} and the absence of clear leadership^{95–100}.

Threat perception

Although ignoring threats and risk were concerns raised in the 2020 article, the statement that we assessed as a claim (11) did not yield a substantial amount of evidence to review. However, this was not because there was an absence of evidence relating to the general issue of threat perception. In fact, substantial research indicated that threat perception – and wilful decisions to ignore risks to self and others – were a major factor during the pandemic^{101–103}. However, we chose not to create an additional generic claim to assess evidence related to this, not only to maintain consistency in the method but also because much of the research on this topic has been heavily associated with the polarization, messaging and misinformation themes. Still, there is clear and compelling evidence that deliberate decisions to ignore health information had negative impacts during the pandemic^{102,103}.

Nudging

Nudging was a widely attempted method for behavioural interventions during the pandemic. Huge increases in attempted nudges have arisen since 2019, largely due to the highly behavioural nature of pandemic policies. Although not explicitly framed as a claim in the 2020 article, nudging was highlighted as a practice likely to have a substantial bearing on pandemic-related behaviour.

Overall, interventions presented as nudges had mixed effectiveness during the pandemic. Encouraging evidence found that simplifying choice architecture and making options salient (for example, through personalized text messages), as well as making it easy to become vaccinated led to reductions in vaccine hesitancy^{104,105}. The same has been found for improving availability of locations to receive a vaccine¹⁰¹. Accuracy prompts have also shown some promise as nudges that aimed to limit sharing of misinformation^{87,106}, although replications have found generally small effect sizes for these nudges^{87,107}. However, attempts at making use of lotteries to increase vaccination rates had no overall effect⁹², along with studies reviewed in claim 7 on messaging, which also had little impact⁶⁷.

Because of the extreme number of trials, there is no single summary assessment that would appropriately cover the highs and lows, or complexity, of nudging during the pandemic. Several systematic reviews^{108,109} have explored the overall effectiveness of nudges, and more narrow systematic reviews have considered the effectiveness of nudges that target specific behaviours, such as vaccination¹⁰⁵. In light of this mixed picture, we strongly encourage focused systematic reviews of all nudging carried out in the context of COVID-19 and urge nuance in determining which nudges work and which do not (treating them as equivalent does not seem to be supported by the data).

Stress and coping

Unfortunately, the fear that isolation and lack of social connectedness would lead to a pandemic of mental illness largely played out^{110–112}.

Table 6 | Evidence assessments for four claims on culture, politics and misinformation

Claim (number)	Evidence	Level	Direction	Effect size	Summary of evidence		
Cultures accustomed to prioritizing freedom over security may also have more difficulty coordinating in the face of a pandemic (12)	Clear correlational evidence for an effect in different contexts and on different levels. Studies differ in the cultural dimensions assessed, including freedom–security, tightness–looseness and collectivism–individualism	Widely tested in real-world settings or field studies	Positive	Medium	Articles reviewed: 34 Average review time: 18 h (spread over 3–14 days)	Sample range: 384–910,006 Mean sample: 64,641.8	Median sample: 3,569.5
Preparing people for misinformation and ensuring they have accurate information and counterarguments against false information before they encounter conspiracy theories, fake news or other forms of misinformation can help to inoculate them against false information (9)	The application of this claim shows robust positive effects in online experiments and real-world applications, although effect sizes vary. Meta-analytic assessments of the effectiveness of the interventions exist	Replicated real-world or field study evidence	Positive	Medium	Articles reviewed: 60 Average review time: 15 h (spread over 3–6 days)	Sample range: 102–33,480 Mean sample: 4,340.9	Median sample: 1,554
Unmitigated political polarization will disrupt or create other negative effects on attempts to minimize or end the pandemic (14)	Robust findings for the effects of polarization in survey studies, but very few studies including manipulation or intervention. Context is very focused on the USA	Replicated real-world or field study evidence	Positive	Medium	Articles reviewed: 54 Average review time: 10 h (spread over 2–4 days)	Sample range: 235–447,332 Mean sample: 26,389.5	Median sample: 3,145.5
Fake news, conspiracy theories and misinformation will have a negative effect on vaccine hesitancy (13) ^a	Consistent evidence from survey data and correlational evidence for the claim indicating small-to-medium effect sizes	Empirical evidence (such as surveys, laboratory experiments and controlled settings)	Positive	Medium	Articles reviewed: 60 Average review time: 16 h (spread over 3–14 days)	Sample range: 104–26,576 Mean sample: 5,041.9	Median sample: 2,220

Overview of ratings and assessments for four claims (9, 12, 13 and 14). Note that the summaries of sample sizes included any studies that included evidence and a rating, irrespective of overall influence on summary assessment. Articles may be double-counted if they were directly relevant for multiple claims. ^aThe use of 'negative effect' here may create confusion for the intended meaning of claim 13; this is clarified in the Supplementary Information section "Notes for specific claims".

Although much of daily life around the world adapted to major changes, the risk of prolonged and severe effects on mental health were widely identified, with large increases in depression, anxiety, stress and other common mental disorders reported globally¹¹³. In some cases, these effects were moderated (or at least attenuated) by being isolated along with close others¹¹⁴, whereas other studies have found dramatic increases in intimate partner violence and violence against women^{115,116}. Some positive mental health outcomes had direct links to collective mindset and perspective¹¹⁷, although not able to circumvent all aspects, consistent with the 2020 article. Those patterns indicated a need to take more multidimensional approaches to well-being and mental health to find opportunities not only to treat or prevent illness but also to promote positive outcomes¹¹⁸.

Major pandemic behavioural themes

Although matters such as polarization and vaccine hesitancy were discussed in the 2020 article and turned out to be clearly relevant, other themes not specified originally have been widely studied. For example, political divisions were not the only reasons individuals refused or delayed vaccination^{119,120} as there has also been evidence of general wilful refusal to follow public health guidelines (whether masking, social distancing, isolating when sick, avoiding unnecessary travel, vaccine hesitancy, and so on) in some individuals¹²¹. In this regard, explicit, manifest behaviours based on demographics and individual differences¹²² should also be reviewed as they have not been covered

here. Those patterns are not inconsistent with perceptions, beliefs and social division, but we have not explicitly assessed that evidence. Even though some evidence pointed to the benefits of communicating good and effective policies directly to the public¹²³, more needs to be done to explore how to achieve this when there is active, deliberate intent to criticize and disrupt those policies, without inadvertently giving greater visibility to those disruptive forces.

Other major themes not covered in Van Bavel et al. include more specific predictions about what outcomes may be associated with behaviours or policy interventions. For example, although there has been some mention of isolation impacting mental health, volumes of research looked at how school closures¹²⁴ and curfews¹²⁵ might influence children by limiting opportunities for interaction, playing and development, weighed against their likely effect on mitigating the spread of illness. Similarly, beyond social media, ways to address isolation might have involved better ways to engage communities in volunteering¹²⁶ or other civic contributions for those that desired a more active role during periods of extended isolation.

Another theme not discussed was how traditional forms of mass media might have undermined the potentially helpful role of descriptive norms by giving disproportionate attention to anti-vaccination, conspiracy and other beliefs that did not reflect expert or even majority opinion in the general public^{53,127}. Many countries, particularly those covered in the original article and where evidence was available for this paper, had vaccination rates above 70% (and sometimes above 90%). In these settings, messaging that focuses on the problem of vaccine refusal



Fig. 1 | Reviewer-assessed effect size for each claim and the qualitative rating. The y axis shows the reviewer-assessed effect size for each claim; the x axis shows the qualitative rating (from theory only to widely tested). Each set of claims is represented by a different icon. Most claims were confirmed as having small-to-medium effect sizes, including those tested and replicated in real-world

contexts. The strongest finding is indicated by the globe on the right, near the top, which shows the claim about culture (see Table 1) was widely tested in multiple studies and the results were consistent with the original Van Bavel et al. paper at roughly a medium effect size. A legend for each icon to represent the 19 claims is presented below the graph.

could mean giving a minority behaviour the same amount of attention as facts and evidence about widespread uptake and the benefits of vaccination⁶⁵. In this regard, efforts of academics and public health officials may be thwarted if media policies around ‘equal coverage’ are implemented in ways that amplify false norms and harmful, fringe ideas, given how easily it is to manipulate or control narratives that are not rooted in evidence.

General discussion

We approached this research with an appreciation that throughout the pandemic, especially early on, decisions had to be made on the basis of imperfect evidence (see ‘Building policy from imperfect evidence’ in the Supplementary information for further discussion). Our assessment focused on the quality, generalizability and policy relevance of available data rather than the average effect size of research related to the claims, using an expert-driven assessment of claims rather than formal statistical meta-analysis.

Our two teams of 72 total reviewers assessed 747 scientific articles covering 19 highly influential claims made in 2020 about human behaviour in the pandemic. Of the 747 articles, 463 studies included original empirical evidence and 418 had human participants (mean sample size of 16,848). Two independent teams evaluated the studies available for each claim (see ‘Author contributions’ for specific lists). Both teams found evidence in support of 16 of the 19 claims (84%), with no evidence available for one claim. For two claims, teams found only null effects. Overall, our review found that the Van Bavel et al. article generally anticipated meaningful topics that became relevant for research during the pandemic, and in the majority of cases, the direction of

their associated research findings. In particular, it identified relevant behaviours during the pandemic (both positive and negative), likely barriers to mitigating the spread of the disease and major social challenges that would be faced by policymakers.

Aside from cultural effects on policy effectiveness, the most strongly supported intervention claim was how combatting misinformation and polarization would be vital to promote effective public health guidelines. Effective messaging, particularly in the form of engaging trusted leaders and emphasizing positive social norms, was also heavily supported in the literature. Broadly speaking, survey data strongly supported how critical it is for policy to understand collective behaviour, shared values and effects on marginalized populations to be effective at minimizing harms during a pandemic.

We strongly endorse a full systematic review of all behaviours studied during the pandemic, which would map not only what behaviours were observed (or not) and across how many studies but also where and through what methods. However, we also share the concern that many studies carried out during the pandemic, particularly in the early stages, had low power^{42,43}. That has produced problematic meta-analyses, which was one reason why we chose an alternative approach.

Applying lessons for science and policy

To make use of additional evidence that stemmed from this very broad reading of available literature, but which did not apply only to a single claim, we have consolidated critical recommendations for science and policy in Table 8. The purpose is to help researchers and policymakers respond to future pandemics, disasters or other exogenous shocks. These range from largely scientific practices, such as more study of

Table 7 | Behaviours directly studied in the initial literature review

Washing hands
Standing 6 ft or 2 m apart
Staying at home
Reducing visits to the store
Wearing a mask
Wearing a correct mask
Wearing a mask correctly
Wearing a best-fitting mask
Avoiding crowds
Using public transportation
Exercising at home
Exercising outdoors
Dieting during isolation
Consumption during isolation
Using the Internet
Using social media
Sharing public health information
Sharing misinformation
Getting tested
Testing at home
Visiting nursing homes or the elderly
Getting vaccinated for COVID-19
Refusing a COVID-19 vaccination
Delaying a COVID-19 vaccination
Getting a second shot (of two doses)
Getting a COVID-19 booster
Getting an influenza vaccine
Choosing single-shot doses
Using fake vaccine cards
Using unsupported treatment
Working remotely
Using mental health services
Going to hospitals
Postponing treatment
Spending stimulus money
Attending school remotely
Attending school in-person
Following or ignoring guidelines
Volunteering
Promoting cooperation
Promoting division
Messaging on social media
Using dating apps
Moving home with family
Isolating within home
Joining the military during the COVID-19 pandemic
Recycling and plastic consumption during the COVID-19 pandemic
Death ^a and suicide
Downloading or using tracking apps
Trust ^b
Prosociality ^b
Intentions ^b
Continued

Beliefs^b

Preferences^b

Polarization^b

^aDeath was a heavily studied outcome that was linked to both COVID-19 and the behaviours listed. There was debate as to whether it should alone count as a behaviour, so we included it for readers to decide. ^bBehaviour-adjacent latent constructs that were commonly measured but are not possible to assess with simple binary, objective measures.

global populations, and being more specific in formulating testable questions, to primarily policy topics related to communications and managing public uncertainty. Some of these recommendations may appear obvious or generic, but were not universally adopted in the studies that we reviewed.

In light of the challenges discussed, we offer two additional recommendations. These are specifically relevant to future public health emergencies but also apply broadly towards advancing the readiness of scientific evidence being applied to policy.

Follow surveys with field research

Behavioural scientists, including those studying the pandemic, often use online data collection tools where there is substantial control over the treatment and, more generally, the research environment. We do not discount the privilege or benefits of having access to these tools or their contribution to foundational evidence that could be deployed rapidly. Online survey experiments could also offer insight on causality that, for example, descriptive or correlational field studies lack. Moreover, it is particularly encouraging that so much of that evidence converged, and this further clarified the added value of having access to those resources.

Despite those benefits, we recommend seeking opportunities to progress faster from concept testing to real-world testing and implementation in future crises. In the earliest days of the pandemic, there was no clear way to do this, and large numbers of studies were derailed, postponed, abandoned or forced to be modified substantially. This also presented challenges for validating evidence generated early in the pandemic (both at that time and during this review). In addition, new research designed to speak to the crisis had to be conducted using the tools available.

It is commendable that so much valuable and applicable social science emerged despite the chaotic circumstances, loss of resources and the uncertainty in the early days and months of the pandemic. Nevertheless, of the 518 studies given a ratings assessment as part of the present exercise, more than 400 were empirical studies conducted in laboratories or online settings (which may have also been skewed by both the nature of the claims and disciplines of individuals identifying relevant studies). Unfortunately, this also meant that many claims were only ever tested in surveys. Rather than this being a criticism, we state this as a strong encouragement to seek, promote and fund partnerships that can function in consequential settings, even (and perhaps especially) in the face of public emergency.

Forge alliances

There are myriad challenges to linking scientific research to real-world practice, which were amplified during the pandemic. Although academics and institutions found ways to overcome these, practical constraints were evident throughout, whether based on resource and personnel limitations, or threats to health and safety. Other limitations may have included simply not knowing the appropriate communication channels to link researchers and policymakers.

In the future, we encourage academics that have not previously worked in such consequential environments to proactively engage with organizations delivering public services to find out where and

Analysis

Table 8 | Recommendations for researchers and policymakers

	Recommended future applications for science and policy	Summary	Example/application
Science	Formulate testable claims	Many prospective articles did not specify the directions of anticipated effects or identify causal factors. Others proposed complex components that were too nuanced to test. Meaningful precision in specifying claims and hypotheses would make it easier for practitioners to trial and incorporate evidence into policies	Prospective recommendations often indicated behaviours that might be 'important', but this creates an arbitrary standard as anything could conceivably be determined important or unimportant depending on the desired interpretation
	Study non-WEIRD populations	Recommendations made on the basis of evidence from WEIRD samples should not be uncritically generalized beyond these populations	Countries may need specific, local strategies that differ from those implemented in countries where much of the research has been done, because of differences in socioeconomic, political and macroeconomic conditions
	Think inside-out and outside-in	Claims from Van Bavel et al. primarily focused on the likely relevance of beliefs, perceptions, identity and other latent constructs. Research during pandemics should also focus on knowing what behaviours are most critical and the best ways to promote them, as well as identify interventions that consider structural contexts, not only psychological constructs	Early-stage recommendations from behavioural science in future pandemics should cover latent constructs (for example, identity, perceptions and norms), objective behaviours (for example, getting vaccinated and wearing a mask) and systemic factors (for example, access to the Internet, availability of healthcare and local legislation)
	Avoid the 'streetlight effect' by researching what matters, not just what is easy	Research during the pandemic often focused on what or who was easy to study rather than on what was most pressing for public health or who was most affected by the pandemic. Behavioural scientists should collaborate with practitioners to develop ways to make sure research resources are deployed where they will have the greatest impact. The most-studied populations in the pandemic were those with easy, stable access to the Internet. The most widely studied topics appeared to be those that could be tested through online surveys	Hundreds of studies on messaging were carried out online in WEIRD populations, but these did not demonstrate overall higher impact findings (in this review, at least)
	Test your assertions or programme evaluation	Many articles from 2020 made strong predictions that were not tested. The lack of clear validation or rejection of potentially influential policy interventions may allow some ineffective policy interventions to crowd out or divert resources from potentially more effective policy interventions	There was very little research studying whether the term 'physical distancing' would have more positive effects than 'social distancing'
Policy	Amplify according to evidence	The interventions getting the most attention were not necessarily those best supported by the most evidence. For example, correlations based on observational data were interpreted as if they were causal	Handwashing was widely promoted as a strategy for stopping the spread of COVID-19, yet study effects were small to null, particularly compared with masking, isolation, distancing and vaccines
	Precision, error, uncertainty and reality checks are always important	During the pandemic, public health agencies relied heavily on mathematical models with implausible assumptions about human behaviour. Behavioural scientists can improve these models by focusing on risk perception, innumeracy, noise, uncertainty and barriers to health behaviours (for example, access and costs)	When building models predicting behaviours such as vaccine uptake and isolation, factor in deviations from expectations based on practical, psychological factors
	Highlight null results	There is as much to learn from effective as ineffective interventions. Failures and backfire effects warrant more visibility to reduce attention to and spending on harmful or wasteful policies	A field experiment showing no significant effect of geo-targeted vaccine lotteries received very little coverage and influence on public policy
Policy	Consider larger context	Research findings may vary depending on national, subnational and other local settings. Translating to policy interventions may require substantial adaptations to replicate effects	Appeal to national identity in liberal versus authoritarian regimes will probably differ in its behavioural consequences and ethical implications
	Do not overcommit too early	Although there are understandable pressures to issue guidelines quickly, establishing a policy position on poor or little evidence can lead to greater costs in the long term	Early guidelines in some countries suggested that wearing masks would not minimize COVID-19, but subsequent evidence has pointed to their effectiveness

Recommendations are ordered from primarily scientific to primarily policy, although to some extent, each recommendation applies to both.

what input they would value¹²⁸. This will help to develop partnerships with local government offices¹²⁹, hospitals¹³⁰, banks¹³¹, schools¹³², local military units providing emergency personnel¹³³ or other potential end-users of scientific evidence. Surveys are not a replacement for studying real behaviours in the field, such as blood donations, vaccines, assigning volunteers, facilitating remote work, keeping people safe while shopping or voting. Furthermore, researchers should recognize the potential for policy impacts at many levels: do not overly focus on highly visible policymakers or government employees who are difficult to access and may lack the subject-specific expertise to recognize all relevant research¹³⁴.

For policymakers, managers, teachers and other institutional leaders, we also strongly recommend opening lines of communications with academics that research your professional area. Professionals seeking to apply insights generated can experience frustration at the reluctance of researchers to offer practical advice for evidence-based policies. Academics are typically not trained in ways to bridge the science–practice–policy gaps, and there are opportunities for impact in the future that may be in place if institutions also take the initiative to engage experts before emergencies. This is also a message for funding bodies: invest in initiatives that support cross-domain collaboration and translational research activity.

Both parties may feel reluctant and uncomfortable about such a collaborative effort, making them hesitant to consider working together as a viable option. Adopting a broader set of tools and research contexts would help both sides to see how they can collaborate more productively by expanding and applying their knowledge of important psychological phenomena and behavioural mechanisms in practice.

Limitations

Tables 2–6 include multiple aspects of the evidence review, including the ratings, direction, effect sizes and a summary note. These were each included because no single rating or value can fully reflect the many dimensions of each behavioural domain. We do not intend or claim to offer a perfect ‘score’ of evidence or research during the pandemic. Instead, our goal was to provide a general assessment of evidence to guide future research and policy applications. Although some policy-makers were asked to participate in this study, our focus was largely on synthesizing academic expertise to inform policy. We therefore note that, at the highest level, evaluations may prioritize scientific perspectives over insights most relevant to decision-makers and practitioners. Multiple, detailed discussion on the limitations of the approach, findings and interpretations of the work is provided in Supplementary Methods.

Conclusion

Despite the absence of a consensus approach to selecting and synthesizing evidence, scientific research has a valuable role in public policy. The present study evaluated researcher claims and recommendations at the start of a global crisis. The synthesis of evidence suggests that researchers can be a viable source of policy advice in the context of a crisis and our recommendations highlight how this can still be improved across scientific disciplines.

Our synthesis and evaluation also speak to the value of revisiting the claims (predictive, indicative or otherwise) that scientists make about events that are of substantial relevance to policy once there is sufficient evidence to assess their validity^{12,24,135,136}. This has the capacity not only to contribute to the sort of transparency that builds trust in science and public health but also to directly inform the development of relevant knowledge and tools for the next pandemic or other crisis.

Our final recommendation is vital and is directed to all scientists, especially behavioural and social, as well as to policy institutions: do not wait until the next crisis to form partnerships. It is always a good time to build relationships between organizations, clinics, schools, governments, media or any institution with which there may be mutual benefits towards building effective policies. This enables us to develop a robust and relevant evidence base and to ensure that we marshal our collective energies and resources so that we are able to use science to best effect in the service of serving, protecting, promoting and prioritizing the well-being of populations.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-023-06840-9>.

- National Academies Press. *Using Science as Evidence in Public Policy* (National Academies Press, 2012).
- Fact sheet: new progress on using behavioral science insights to better serve the American people. [whitehouse.gov](https://obamawhitehouse.archives.gov/the-press-office/2016/09/15/fact-sheet-new-progress-using-behavioral-science-insights-better-serve) <https://obamawhitehouse.archives.gov/the-press-office/2016/09/15/fact-sheet-new-progress-using-behavioral-science-insights-better-serve> (2016).
- Van Bavel, J. J. et al. Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).

- Hodges, R., Caperchione, E., van Helden, J., Reichard, C. & Sorrentino, D. The role of scientific expertise in COVID-19 policy-making: evidence from four European countries. *Public Org. Rev.* **22**, 249–267 (2022).
- Dowd, J. B. et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 9696–9698 (2020).
- Parkhurst, J. O. & Abeyesinghe, S. What constitutes “good” evidence for public health and social policy-making? From hierarchies to appropriateness. *Soc. Epistemol.* **30**, 665–679 (2016).
- Sienkiewicz, M. & Mair, D. in *Science for Policy Handbook* (eds Šucha, V. & Sienkiewicz, M.) 2–13 (Elsevier, 2020).
- Manski, C. F. Communicating uncertainty in policy analysis. *Proc. Natl Acad. Sci. USA* **116**, 7634–7641 (2019).
- Moatti, J.-P. The French response to COVID-19: intrinsic difficulties at the interface of science, public health, and policy. *Lancet Public Health* **5**, e255 (2020).
- Tyler, C. et al. AI tools as science policy advisers? The potential and the pitfalls. *Nature* **622**, 27–30 (2023).
- Policy citations for ‘using social and behavioural science to support COVID-19 pandemic response’ *Altmetric* (accessed 22 November 2023) <https://nature.altmetric.com/details/80902343/policy-documents>.
- Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R. & Messerschmidt, L. COVID-19 government response event dataset (CoronaNet v1.0). *Nat. Hum. Behav.* **4**, 756–768 (2020).
- World Health Organization. Regional Office for Europe. Pandemic fatigue: reinvigorating the public to prevent COVID-19: policy framework for supporting pandemic prevention and management: revised version November 2020. *WHO* <https://apps.who.int/iris/handle/10665/337574> (2020).
- European Commission, Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies & Group of Chief Scientific Advisors. Improving pandemic preparedness and management: lessons learned and ways forward: independent expert report. *Publications Office of the European Union* <http://op.europa.eu/en/publication-detail/-/publication/a1016d77-2562-11eb-9d7e-01aa75ed71a1/language-en/format-PDF> (2020).
- Biddle, N. Behavioural economics and the COVID-induced education crisis. *OECD* https://www.oecd-ilibrary.org/education/behavioural-economics-and-the-covid-induced-education-crisis_eba2b867-en (2021).
- Principles for managing SARS-CoV-2 transmission associated with further education, 3 September 2020. *GOV.UK* <https://www.gov.uk/government/publications/principles-for-managing-sars-cov-2-transmission-associated-with-further-education-3-september-2020> (2020).
- Diminishing returns: nudging COVID-19 prevention among Colombian young adults. *IDB* <https://publications.iadb.org/en/diminishing-returns-nudging-covid-19-prevention-among-colombian-young-adults> (2021).
- Duckett, S., Mackey, W., Stobart, A. & Swerissen, H. Coming out of COVID-19 lockdown: the next steps for Australian health care. *APO* <https://apo.org.au/node/306385> (2020).
- Social scientists scramble to study pandemic, in real time. *Science* <https://www.science.org/content/article/social-scientists-scramble-study-pandemic-real-time> (2020).
- What are behavioural scientists doing about COVID-19? *The Policy Lab* (accessed 22 November 2023) <https://thepolicylab.brown.edu/projects/covid-behavioral-tracker>.
- Van Lange, P. A. M., & Rand, D. G. Human cooperation and the crises of climate change, COVID-19, and misinformation. *Annu. Rev. Psychol.* **73**, 379–402 (2022).
- Byrne-Davis, L. M. T. et al. Using behavioural science in public health settings during the COVID-19 pandemic: the experience of public health practitioners and behavioural scientists. *Acta Psychol.* **224**, 103527 (2022).
- Ioannidis, J. P. A., Saltholz-Hillel, M., Boyack, K. W. & Baas, J. The rapid, massive growth of COVID-19 authors in the scientific literature. *R. Soc. Open Sci.* **8**, 210389 (2021).
- Ijzerman, H. et al. Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* **4**, 1092–1094 (2020).
- Bryan, C. J., Tipton, E. & Yeager, D. S. Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nat. Hum. Behav.* **5**, 980–989 (2021).
- Lewis, N. A. & Wai, J. Communicating what we know and what isn’t so: science communication in psychology. *Perspect. Psychol. Sci.* **16**, 1242–1254 (2021).
- Puthillam, A. Too WEIRD, too fast: preprints about COVID-19 in the psychological sciences. *Collabra Psychol.* <https://doi.org/10.1525/collabra.74331> (2023).
- Galluccio, M. in *Science and Diplomacy: Negotiating Essential Alliances* (ed. Galluccio, M.) 65–74 (Springer International Publishing, 2021).
- Cairney, P. & Oliver, K. Evidence-based policymaking is not like evidence-based medicine, so how far should you go to bridge the divide between evidence and policy? *Health Res. Policy Syst.* **15**, 35 (2017).
- Ellemers, N. Science as collaborative knowledge generation. *Br. J. Soc. Psychol.* **60**, 1–28 (2021).
- Ruppel Shell, E. Act now, wait for perfect evidence later, says ‘high priestess’ of U.K. COVID-19 masking campaign. *Science* <https://doi.org/10.1126/science.abf2811> (2020).
- Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6**, eabd6370 (2020).
- Schwartz, J. L. Evaluating and deploying Covid-19 vaccines — the importance of transparency, scientific integrity, and public trust. *N. Engl. J. Med.* **383**, 1703–1705 (2020).
- Philipp-Muller, A., Lee, S. W. S. & Petty, R. E. Why are people antisocial, and what can we do about it? *Proc. Natl Acad. Sci. USA* **119**, e2120755119 (2022).
- Wingen, T., Berkessel, J. B. & Englich, B. No replication, no trust? How low replicability influences trust in psychology. *Soc. Psychol. Pers. Sci.* **11**, 454–463 (2020).
- Kojan, L., Burbach, L., Ziefle, M. & Calero Valdez, A. Perceptions of behaviour efficacy, not perceptions of threat, are drivers of COVID-19 protective behaviour in Germany. *Humanit. Soc. Sci. Commun.* **9**, 1–15 (2022).
- Santana, A. P., Korn, L., Betsch, C. & Böhm, R. Lessons learned about willingness to adopt various protective measures during the early COVID-19 pandemic in three countries. *PLoS ONE* **17**, e0265892 (2022).

38. Garcia-Garzon, E., Angulo-Brunet, A., Lecuona, O., Barrada, J. R. & Corradi, G. Exploring COVID-19 research credibility among Spanish scientists. *Curr. Psychol.* <https://doi.org/10.1007/s12144-022-02797-6> (2022).
39. Centers for Disease Control and Prevention. How to build COVID-19 vaccine confidence. CDC <https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/building-trust.html> (2022).
40. Nosek, B. A. et al. Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73**, 719–748 (2022).
41. Serra-Garcia, M. & Gneezy, U. Mistakes, overconfidence, and the effect of sharing on detecting lies. *Am. Econ. Rev.* **111**, 3160–3183 (2021).
42. Haber, N. A., Wieten, S. E., Smith, E. R. & Nunan, D. Much ado about something: a response to “COVID-19: underpowered randomised trials, or no randomised trials?”. *Trials* **22**, 780 (2021).
43. ‘It’s misinformation at worst.’ Weak health studies can do more harm than good, scientists say. *Science* <https://www.science.org/content/article/it-s-misinformation-worst-weak-health-studies-can-do-more-harm-good-scientists-say> (2021).
44. Luttrell, A. & Petty, R. E. Evaluations of self-focused versus other-focused arguments for social distancing: an extension of moral matching effects. *Soc. Psychol. Pers. Sci.* **12**, 946–954 (2021).
45. Korn, L., Böhm, R. & Betsch, C. Reply to Rabb et al.: why promoting COVID-19 vaccines with community immunity is not a good strategy (yet). *Proc. Natl Acad. Sci. USA* **118**, e2102054118 (2021).
46. Van Bavel, J. J. et al. National identity predicts public health support during a global pandemic. *Nat. Commun.* **13**, 517 (2022).
47. Frenzel, S. B. et al. How national leaders keep ‘us’ safe: a longitudinal, four-nation study exploring the role of identity leadership as a predictor of adherence to COVID-19 non-pharmaceutical interventions. *BMJ Open* **12**, e054980 (2022).
48. Lu, J. G., Jin, P. & English, A. S. Collectivism predicts mask use during COVID-19. *Proc. Natl Acad. Sci. USA* **118**, e2021793118 (2021).
49. Stevenson, C., Wakefield, J. R. H., Felsner, I., Drury, J. & Costa, S. Collectively coping with coronavirus: local community identification predicts giving support and lockdown adherence during the COVID-19 pandemic. *Br. J. Soc. Psychol.* **60**, 1403–1418 (2021).
50. Bentley, S. V. et al. GROUPS 2 CONNECT: an online activity to maintain social connection and well-being during COVID-19. *Appl. Psychol. Health Well Being* <https://doi.org/10.1111/aphw.12330> (2022).
51. Bicchieri, C. et al. In science we (should) trust: expectations and compliance across nine countries during the COVID-19 pandemic. *PLoS ONE* **16**, e0252892 (2021).
52. Hensel, L. et al. Global behaviors, perceptions, and the emergence of social norms at the onset of the COVID-19 pandemic. *J. Econ. Behav. Org.* **193**, 473–496 (2022).
53. Chadwick, A. et al. Online social endorsement and Covid-19 vaccine hesitancy in the United Kingdom. *Soc. Media Soc.* **7**, 20563051211008816 (2021).
54. Rabb, N., Bowers, J., Glick, D., Wilson, K. H. & Yokum, D. The influence of social norms varies with “others” groups: evidence from COVID-19 vaccination intentions. *Proc. Natl Acad. Sci. USA* **119**, e2118770119 (2022).
55. Stuart, J., O’Donnell, K., O’Donnell, A., Scott, R. & Barber, B. Online social connection as a buffer of health anxiety and isolation during COVID-19. *Cyberpsychol. Behav. Soc. Netw.* **24**, 521–525 (2021).
56. Marinucci, M., Pancani, L., Aureli, N. & Riva, P. Online social connections as surrogates of face-to-face interactions: a longitudinal study under Covid-19 isolation. *Comp. Hum. Behav.* **128**, 107102 (2022).
57. Sun, R., Rieble, C., Liu, Y. & Sauter, D. Connected despite lockdown: the role of social interactions and social media use in wellbeing. *Collabra: Psychol.* **8**, 37061 (2022).
58. Tunggenç, B. et al. Social influence matters: We follow pandemic guidelines most when our close circle does. *Br. J. Psychol.* **112**, 763–780 (2021).
59. Baal, S. T. van, Walasek, L. & Hohwy, J. Staying home so you can keep going out: a multiplayer self-isolation game modelling pandemic behaviour. *Sci. Rep.* **12**, 13466 (2022).
60. Zhu, P. et al. The efficacy of a brief, altruism-eliciting video intervention in enhancing COVID-19 vaccination intentions among a population-based sample of younger adults: randomized controlled trial. *JMIR Public Health Surveill.* **8**, e37328 (2022).
61. Abu-Akel, A., Spitz, A. & West, R. The effect of spokesperson attribution on public health message sharing during the COVID-19 pandemic. *PLoS ONE* **16**, e0245100 (2021).
62. Arroyos-Calvera, D., Drouvelis, M., Lohse, J. & McDonald, R. Improving compliance with COVID-19 guidance: a workplace field experiment. *SSRN* <https://doi.org/10.2139/ssrn.3760562> (2021).
63. Bos, B., Drupp, M. A., Meya, J. N. & Quaas, M. F. Moral suasion and the private provision of public goods: evidence from the COVID-19 pandemic. *Environ. Res. Econ.* **76**, 1117–1138 (2020).
64. Chu, J., Pink, S. L. & Willer, R. Religious identity cues increase vaccination intentions and trust in medical experts among American Christians. *Proc. Natl Acad. Sci. USA* **118**, e2106481118 (2021).
65. Larsen, B. et al. Counter-stereotypical messaging and partisan cues: moving the needle on vaccines in a polarized U.S. <http://www.nber.org/papers/w29896.pdf> (NBER, 2022).
66. Moehring, A. et al. Surfacing norms to increase vaccine acceptance. *SSRN* <https://doi.org/10.2139/ssrn.3782082> (2021).
67. Rabb, N. et al. Evidence from a statewide vaccination RCT shows the limits of nudges. *Nature* **604**, E1–E7 (2022).
68. Teeny, J. D., Siev, J. J., Briñol, P. & Petty, R. E. A review and conceptual framework for understanding personalized matching effects in persuasion. *J. Consum. Psychol.* **31**, 382–414 (2021).
69. Flores, A. et al. Politicians polarize and experts depolarize public support for COVID-19 management policies across countries. *Proc. Natl Acad. Sci. USA* **119**, e2117543119 (2022).
70. Long, A. et al. Co-created messaging for influenza vaccination in a high-risk Hispanic community provides groundwork for COVID-19 vaccine. *Health Equity* **5**, 345–352 (2021).
71. de Saint Laurent, C., Murphy, G., Hegarty, K. & Greene, C. M. Measuring the effects of misinformation exposure and beliefs on behavioural intentions: a COVID-19 vaccination study. *Cogn. Res. Princ. Implic.* **7**, 87 (2022).
72. Su, Z. et al. Mental health consequences of COVID-19 media coverage: the need for effective crisis communication practices. *Global. Health* **17**, 4 (2021).
73. 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).
74. 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).
75. Farhart, C. E., Douglas-Durham, E., Lunz Trujillo, K. & Vitriol, J. A. Vax attacks: how conspiracy theory belief undermines vaccine support. *Prog. Mol. Biol. Transl. Sci.* **188**, 135–169 (2022).
76. Freeman, D. et al. COVID-19 vaccine hesitancy in the UK: the Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. *Psychol. Med.* <https://doi.org/10.1017/S0033291720005188> (2020).
77. Gelfand, M. J. et al. The relationship between cultural tightness–looseness and COVID-19 cases and deaths: a global analysis. *Lancet Planet. Health* [https://doi.org/10.1016/S2542-5196\(20\)30301-6](https://doi.org/10.1016/S2542-5196(20)30301-6) (2021).
78. Gollwitzer, A., Martel, C., Brady, W. J., Knowles, E. D. & Van Bavel, J. Partisan differences in physical distancing predict infections and mortality during the coronavirus pandemic. *SSRN* <https://doi.org/10.2139/ssrn.3609392> (2020).
79. Kitayama, S., Camp, N. P. & Salvador, C. E. Culture and the COVID-19 pandemic: multiple mechanisms and policy implications. *Soc. Issues Policy Rev.* **16**, 164–211 (2022).
80. Kreps, S., Dasgupta, N., Brownstein, J. S., Hswen, Y. & Kriner, D. L. Public attitudes toward COVID-19 vaccination: the role of vaccine attributes, incentives, and misinformation. *npj Vaccines* **6**, 73 (2021).
81. Marchlewska, M., Hamer, K., Baran, M., Górská, P. & Kaniasty, K. COVID-19: why do people refuse vaccination? The role of social identities and conspiracy beliefs: evidence from nationwide samples of Polish adults. *Vaccines* **10**, 268 (2022).
82. Painter, M. & Qiu, T. Political beliefs affect compliance with government mandates. *J. Econ. Behav. Org.* <https://doi.org/10.1016/j.jebo.2021.03.019> (2021).
83. Porter, E., Wood, T. J. & Velez, Y. Correcting COVID-19 vaccine misinformation in ten countries. *R. Soc. Open Sci.* **10**, 221097 (2023).
84. Rodriguez, C. G., Gadarian, S. K., Goodman, S. W. & Pepinsky, T. B. Morbid polarization: exposure to COVID-19 and partisan disagreement about pandemic response. *Polit. Psychol.* <https://doi.org/10.1111/pops.12810> (2022).
85. Roozenbeek, J. et al. Susceptibility to misinformation about COVID-19 around the world. *R. Soc. Open Sci.* **7**, 201199 (2020).
86. Basol, M. et al. Towards psychological herd immunity: cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data Soc.* **8**, 20539517211013868 (2021).
87. Pennycook, G. & Rand, D. G. Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nat. Commun.* **13**, 2333 (2022).
88. Capraro, V. & Celadin, T. “I think this news is accurate”: endorsing accuracy decreases the sharing of fake news and increases the sharing of real news. *Pers. Soc. Psychol. Bull.* <https://doi.org/10.1177/01461672221117691> (2022).
89. Elgar, F. J., Stefaniak, A. & Wohl, M. J. A. The trouble with trust: time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries. *Soc. Sci. Med.* **263**, 113365 (2020).
90. Khazanchi, R., Marcelin, J., Abdoul-Mutakabbir, J. & Essien, U. Race, racism, civil rights law, and the equitable allocation of scarce COVID-19 treatments. *Health Aff. Forefront* <https://doi.org/10.1377/forefront.20220208.453850> (2022).
91. Scheitle, C. P. & Corcoran, K. E. COVID-19 skepticism in relation to other forms of science skepticism. *Socius* **7**, 23780231211049840 (2021).
92. Milkman, K. L. et al. A citywide experiment testing the impact of geographically targeted, high-pay-off vaccine lotteries. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-022-01437-0> (2022).
93. Campos-Mercade, P. et al. Monetary incentives increase COVID-19 vaccinations. *Science* **374**, 879–882 (2021).
94. Thirumurthy, H., Milkman, K. L., Volpp, K. G., Buttenheim, A. M. & Pope, D. G. Association between statewide financial incentive programs and COVID-19 vaccination rates. *PLoS ONE* **17**, e0263425 (2022).
95. Gulati, K. & Busari, J. Vaccinating a billion people against COVID-19: India’s quest for systems leadership in exceptional times. *Leadersh. Health Serv.* **35**, 137–148 (2021).
96. Cheema, S., Ameduri, M., Abraham, A., Doraiswamy, S. & Mamtani, R. The COVID-19 pandemic: the public health reality. *Epidemiol. Infect.* **148**, e223 (2020).
97. European Observatory on Health Systems and Policies et al. Public health leadership and the COVID-19 pandemic in Europe. *Eurohealth* **27**, 4–9 (2021).
98. Nkengasong, J. et al. The global response to the COVID-19 pandemic. *Med* **1**, 3–8 (2020).
99. Schwab, A. & Seas, C. The COVID-19 pandemic in Peru: what went wrong? *Am. J. Trop. Med. Hyg.* **104**, 1176–1178 (2021).
100. Ortega, F. & Orsini, M. Governing COVID-19 without government in Brazil: ignorance, neoliberal authoritarianism, and the collapse of public health leadership. *Glob. Public Health* **15**, 1257–1277 (2020).
101. Brownstein, J., Cantor, J. H., Rader, B., Simon, K. I. & Whaley, C. M. If you build it, will they vaccinate? The impact of COVID-19 vaccine sites on vaccination rates and outcomes. <https://doi.org/10.3386/w30429> (NBER, 2022).
102. Thunström, L., Ashworth, M., Shogren, J. F., Newbold, S. & Finnoff, D. Testing for COVID-19: willful ignorance or selfless behavior? *Behav. Public Policy* **5**, 135–152 (2021).
103. Leonhardt, J. M., Ridinger, G., Rong, Y. & Talaei-Khoe, A. Invincibility threatens vaccination intentions during a pandemic. *PLoS ONE* **16**, e0258432 (2021).
104. Dai, H. et al. Behavioural nudges increase COVID-19 vaccinations. *Nature* **597**, 404–409 (2021).
105. Reñosa, M. D. C. et al. Nudging toward vaccination: a systematic review. *BMJ Glob. Health* **6**, e006237 (2021).
106. Epstein, Z. et al. Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online. *Harvard Kennedy School Misinformation Rev.* <https://doi.org/10.37016/mr-2020-71> (2021).

107. Roozenbeek, J., Freeman, A. L. J. & van der Linden, S. How accurate are accuracy-nudge interventions? A preregistered direct replication of Pennycook et al. (2020). *Psychol. Sci.* **32**, 1169–1178 (2021).
108. Maier, M. et al. No evidence for nudging after adjusting for publication bias. *Proc. Natl Acad. Sci. USA* **119**, e2200300119 (2022).
109. DellaVigna, S. & Linos, E. RCTs to scale: comprehensive evidence from two nudge units. *Econometrica* **90**, 81–116 (2022).
110. Aguilar-Latorre, A. et al. The impact of the COVID-19 lockdown on depression sufferers: a qualitative study from the province of Zaragoza, Spain. *BMC Public Health* **22**, 780 (2022).
111. Loades, M. E. et al. Rapid systematic review: the impact of social isolation and loneliness on the mental health of children and adolescents in the context of COVID-19. *J. Am. Acad. Child Adolesc. Psychiatry* **59**, 1218–1239.e3 (2020).
112. Pancani, L., Marinucci, M., Aureli, N. & Riva, P. Forced social isolation and mental health: a study on 1,006 Italians under COVID-19 lockdown. *Front. Psychol.* **12**, 663799 (2021).
113. World Health Organization. Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief, 2 March 2022. WHO www.who.int/publications-detail-redirect/WHO-2019-nCoV-Sci-Brief-Mental_health-2022.1 (2022).
114. Sels, L. et al. Intimate relationships in times of COVID-19: a descriptive study of Belgian partners and their perceived well-being. *Psychol. Belg.* **62**, 1–16 (2022).
115. Boserup, B., McKenney, M. & Elkbuli, A. Alarming trends in US domestic violence during the COVID-19 pandemic. *Am. J. Emerg. Med.* **38**, 2753–2755 (2020).
116. *Measuring the Shadow Pandemic Violence Against Women During COVID-19* (United Nations, 2021).
117. Zion, S. R. et al. Making sense of a pandemic: mindsets influence emotions, behaviors, health, and well-being during the COVID-19 pandemic. *Soc. Sci. Med.* **301**, 114889 (2022).
118. Ruggeri, K., Garcia-Garzon, E., Maguire, A., Matz, S. & Huppert, F. A. Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. *Health Qual. Life Outcomes* **18**, 192 (2020).
119. Fieselmann, J., Annac, K., Erdsiek, F., Yilmaz-Aslan, Y. & Brzoska, P. What are the reasons for refusing a COVID-19 vaccine? A qualitative analysis of social media in Germany. *BMC Public Health* **22**, 846 (2022).
120. Bullock, J., Lane, J. E. & Shults, F. L. What causes COVID-19 vaccine hesitancy? Ignorance and the lack of bliss in the United Kingdom. *Humanit. Soc. Sci. Commun.* **9**, 1–7 (2022).
121. Schumpe, B. M. et al. Predictors of adherence to public health behaviors for fighting COVID-19 derived from longitudinal data. *Sci. Rep.* **12**, 3824 (2022).
122. Myerson, J., Strube, M. J., Green, L. & Hale, S. Individual differences in COVID-19 mitigation behaviors: the roles of age, gender, psychological state, and financial status. *PLoS ONE* **16**, e0257658 (2021).
123. Belchior, A. M. & Teixeira, C. P. Determinants of political trust during the early months of the COVID-19 pandemic: putting policy performance into evidence. *Polit. Stud. Rev.* <https://doi.org/10.1177/14789299211056193> (2021).
124. Viner, R. et al. School closures during social lockdown and mental health, health behaviors, and well-being among children and adolescents during the first COVID-19 wave: a systematic review. *JAMA Pediatr.* **176**, 400–409 (2022).
125. Baumann, M. COVID-19 and mental health in children and adolescents: a diagnostic panel to map psycho-social consequences in the pandemic context. *Discov. Ment. Health* **1**, 2 (2021).
126. Gómez-Durán, E. L. et al. COVID-19 pandemic psychological impact and volunteering experience perceptions of medical students after 2 years. *Int. J. Environ. Res. Public Health* **19**, 7532 (2022).
127. Jiang, X., Hwang, J., Shah, D. V., Ghosh, S. & Brauer, M. News attention and social-distancing behavior amid COVID-19: how media trust and social norms moderate a mediated relationship. *Health Commun.* **37**, 768–777 (2022).
128. Fields, J. et al. Community-academic partnerships to address Covid-19 inequities: lessons from the San Francisco Bay Area. *NEJM Catalyst* <https://doi.org/10.1056/CAT.21.0135> (2021).
129. Mullin, M. Learning from local government research partnerships in a fragmented political setting. *Public Adm. Rev.* **81**, 978–982 (2021).
130. DeBoer, S. et al. Building successful and sustainable academic health science partnerships: exploring perspectives of hospital leaders. *Can. Med. Educ. J.* **10**, e56–e67 (2019).
131. Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M. & Yannelis, C. Income, liquidity, and the consumption response to the 2020 Economic Stimulus Payments. *NBER* <https://www.nber.org/papers/w27097> (2020).
132. D'Agostino, E. M. et al. School-academic partnerships in support of safe return to schools during the COVID-19 pandemic. *Pediatrics* <https://doi.org/10.1542/peds.2021-054268C> (2022).
133. Ruggeri, K. et al. Role of military forces in the New York State response to COVID-19. *JAMA Health Forum* **3**, e222136 (2022).
134. Subbaraman, N. Science misinformation alarms Francis Collins as he leaves top NIH job. *Nature* **600**, 372–373 (2021).
135. Habersaat, K. B. et al. Ten considerations for effectively managing the COVID-19 transition. *Nat. Hum. Behav.* **4**, 677–687 (2020).
136. Haug, N. et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your

intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Kai Ruggeri^{1,2,3,52}, Friederike Stock^{4,5}, S. Alexander Haslam⁶, Valerio Capraro⁷, Paulo Boggio^{8,9}, Naomi Ellemers¹⁰, Aleksandra Cichocka¹¹, Karen M. Douglas¹¹, David G. Rand¹², Sander van der Linden¹³, Mina Cikara¹⁴, Eli J. Finkel^{15,16}, James N. Druckman¹⁷, Michael J. A. Wohtl¹⁸, Richard E. Petty¹⁹, Joshua A. Tucker²⁰, Azim Shariff²¹, Michele Gelfand²², Dominic Packer²³, Jolanda Jetten⁶, Paul A. M. Van Lange^{24,25}, Gordon Pennycook²⁶, Ellen Peters^{27,28}, Katherine Baicker²⁹, Alia Crum³⁰, Kim A. Weeden²⁶, Lucy Napper²³, Nassim Tabri¹⁸, Jamil Zaki²², Linda Skitka³¹, Shinobu Kitayama³², Dean Mobbs^{33,34}, Cass R. Sunstein¹⁴, Sarah Ashcroft-Jones^{1,35}, Anna Louise Todsen³⁶, Ali Hajian³⁷, Sanne Verra¹⁰, Vanessa Buehler²⁸, Maja Friedemann³⁹, Marlene Hecht^{4,5}, Rayyan S. Mobarak⁴⁰, Ralitsa Karakasheva⁴¹, Markus R. Tünte⁴², Siu Kit Yeung⁴³, R. Shayna Rosenbaum^{44,45}, Žan Lep^{46,47}, Yuki Yamada⁴⁸, Sa-kiera Tierra Jolynn Hudson⁴⁹, Lucia Macchia⁵⁰, Irina Soboleva⁵¹, Eugen Dimant^{52,53}, Sandra J. Geiger⁵⁴, Hannes Jarke², Tobias Wingen⁵⁵, Jana B. Berkessel⁵⁶, Silvana Mareva^{57,58}, Lucy McGill^{59,60}, Francesca Papa⁶¹, Bojana Veckalov⁶², Zeina Afif⁶³, Eike K. Buabang⁶⁴, Marna Landman⁶⁵, Felice Tavera⁶⁶, Jack L. Andrews^{35,67}, Aslı Bursalioğlu⁶⁸, Zorana Zupan⁶⁹, Lisa Wagner^{70,71}, Joaquín Navajas^{72,73,74}, Marek Vranka⁷⁵, David Kasdan⁷⁶, Patricia Chen^{77,78}, Kathleen R. Hudson³¹, Lindsay M. Novak³¹, Paul Teas³¹, Nikolay R. Rachev⁷⁹, Matteo M. Galizzi⁸⁰, Katherine L. Milkman⁸¹, Marija Petrović⁸², Jay J. Van Bavel⁸³ & Robb Willer⁸⁴

¹Department of Health Policy and Management, Columbia University Mailman School of Public Health, New York City, NY, USA. ²Policy Research Group, Centre for Business Research, Judge Business School, University of Cambridge, Cambridge, UK. ³274th ASOS, US Air Force/New York Air National Guard, New York, Syracuse, United States. ⁴Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. ⁵Department of Psychology, Humboldt University of Berlin, Berlin, Germany. ⁶University of Queensland, St Lucia, Queensland, Australia. ⁷University of Milan-Bicocca, Milan, Italy. ⁸Mackenzie Presbyterian University, São Paulo, Brazil. ⁹National Institute of Science and Technology on Social and Affective Neuroscience, CNPq, São Paulo, Brazil. ¹⁰Utrecht University, Utrecht, Netherlands. ¹¹University of Kent, Canterbury, UK. ¹²Massachusetts Institute of Technology, Cambridge, MA, USA. ¹³Department of Psychology, University of Cambridge, Cambridge, UK. ¹⁴Harvard University, Cambridge, MA, USA. ¹⁵Department of Psychology, Northwestern University, Evanston, IL, USA. ¹⁶Kellogg School of Management, Northwestern University, Evanston, IL, USA. ¹⁷Northwestern University, Evanston, IL, USA. ¹⁸Department of Psychology, Carleton University, Ottawa, Ontario, Canada. ¹⁹Department of Psychology, Ohio State University, Columbus, OH, USA. ²⁰Department of Politics & Center for Social Media and Politics, New York University, New York, NY, USA. ²¹Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada. ²²Stanford University, Stanford, CA, USA. ²³Lehigh University, Bethlehem, PA, USA. ²⁴Institute for Brain and Behavior Amsterdam, Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. ²⁵Global Faculty, Social and Economic Behavior, University of Cologne, Cologne, Germany. ²⁶Cornell University, Ithaca, NY, USA. ²⁷Center for Science Communication Research, School of Journalism and Communication, University of Oregon, Eugene, OR, USA. ²⁸Psychology Department, University of Oregon, Eugene, OR, USA. ²⁹University of Chicago, Chicago, IL, USA. ³⁰Department of Psychology, Stanford University, Stanford, CA, USA. ³¹University of Illinois Chicago, Chicago, IL, USA. ³²University of Michigan, Ann Arbor, MI, USA. ³³Department of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA. ³⁴Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA, USA. ³⁵Department of Experimental Psychology, University of Oxford, Oxford, UK. ³⁶Department of Social Policy and Evaluation, University of Oxford, Oxford, UK. ³⁷University of Tehran, Tehran, Iran. ³⁸Cowry Consulting, London, UK. ³⁹University of Oxford, Oxford, UK. ⁴⁰Department of Agricultural and Resource Economics, University of Maryland, College Park, MD, USA. ⁴¹Junior Researcher Programme, Cambridge, UK. ⁴²Department of Developmental and Educational Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria. ⁴³Department of Psychology, The Chinese University of Hong Kong, Hong Kong SAR, China. ⁴⁴Department of Psychology, York University, Toronto, Ontario, Canada. ⁴⁵Rotman Research Institute, Baycrest Academy for Research and Education, Toronto, Ontario, Canada. ⁴⁶Department of Psychology, Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia. ⁴⁷Centre for Applied Epistemology, Educational Research Institute, Ljubljana, Slovenia. ⁴⁸Faculty of Arts and Science, Kyushu University, Fukuoka, Japan. ⁴⁹Haas School of Business, University of California Berkeley, Berkeley, CA, USA. ⁵⁰City, University of London, London, UK. ⁵¹Duke Kunshan University, Kunshan, China. ⁵²Center for Social Norms and Behavioral Dynamics, University of Pennsylvania, Philadelphia, PA, USA. ⁵³CESifo, Munich, Germany. ⁵⁴Environmental Psychology, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria. ⁵⁵University of Bonn, University Hospital Bonn, Institute of General Practice and Family Medicine, Bonn, Germany. ⁵⁶Mannheim Centre for European Social Research, University of Mannheim, Mannheim, Germany. ⁵⁷MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK. ⁵⁸Psychology Department, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK. ⁵⁹University College Dublin, Dublin, Ireland. ⁶⁰University of Groningen, Groningen, Netherlands. ⁶¹Organisation for Economic Co-operation and Development, Paris, France. ⁶²University of Amsterdam, Amsterdam, Netherlands. ⁶³The World Bank,

Analysis

Washington DC, USA.⁶⁴Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland.⁶⁵Gordon Institute of Business Science, University of Pretoria, Johannesburg, South Africa.⁶⁶Department of Psychology, University of Cologne, Cologne, Germany.⁶⁷University College, Oxford, UK.⁶⁸Department of Psychology, Loyola University Chicago, Chicago, IL, USA.⁶⁹Institute of Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia.⁷⁰Jacobs Center for Productive Youth Development, University of Zurich, Zurich, Switzerland.⁷¹Department of Psychology, University of Zurich, Zurich, Switzerland.⁷²Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Buenos Aires, Argentina.⁷³Escuela de Negocios, Universidad Torcuato Di Tella, Buenos Aires, Argentina.⁷⁴Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

⁷⁵Charles University, Prague, Czech Republic.⁷⁶Sungkyunkwan University, Seoul, Republic of Korea.⁷⁷University of Texas at Austin, Austin, TX, USA.⁷⁸National University of Singapore, Singapore, Singapore.⁷⁹Department of General, Experimental, Developmental, and Health Psychology, Sofia University St. Kliment Ohridski, Sofia, Bulgaria.⁸⁰Department of Psychological and Behavioural Science, London School of Economics, London, UK.⁸¹The Wharton School of the University of Pennsylvania, Philadelphia, PA, USA.⁸²Department of Psychology & Laboratory for Research of Individual Differences, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia.⁸³Department of Psychology & Center for Neural Science, New York University, New York, NY, USA.⁸⁴Department of Sociology, Stanford University, Stanford, CA, USA.⁸⁵e-mail: kai.ruggeri@columbia.edu

Methods

Our approach assessed evidence related to the central statements or hypotheses (claims) in the original Van Bavel et al.³ article. For the purposes of evaluation, we treated these claims as testable hypotheses, then rated the level, direction and magnitude of findings relevant to each claim. Evaluation was conducted by seventy-two evaluators, including authors of the original study as well as an independent group of behavioural scientists and policymakers. Their assessments focused on whether evidence from the first 2 years of the pandemic supported, refuted or left unclear the validity of the claims (see the online database available at <https://psyarxiv.com/58udn> for details).

Claims evaluated

We evaluated the ten claims highlighted in Table 8 of the original article, as well as five additional claims made in the main text. Those additional claims related to behaviours, themes or policies that ended up being especially relevant during the pandemic, such as vaccination choices and the influence of political polarization, but which did not clearly overlap with one of the ten primary claims. All other claim-like statements in the text were either already covered in the original ten or were not precise enough to assess. One of the original ten claims actually comprised five distinct claims, creating a total of 19 claims.

Some claims were more general and not well suited to be treated as hypotheses, but rather as recommendations. For example, claim 6 (see Table 1) first suggests that targeted messaging and then that partnerships with community organizations are valuable. The first part of that statement is more suited to treat as a hypothesis, whereas the second part states only that community partnerships may be worth looking into, but with no implicit impact. In this case, we disregarded the second part and focused on evidence that could inform the first part.

Evidence used for evaluations

We identified articles and reports used for the assessment through extensive systematic and manual searches by all evaluators, with the primary criterion being that they were publicly available before 1 June 2022. Searches included using the pre-formatted systematic review code produced by PubMed-NCBI for research specifically on COVID-19 (available at <https://pubmed.ncbi.nlm.nih.gov/help/#covid19-article-filters>), as well as checking preprint servers (OSF, PsyArXiv and SSRN), multiple repository search engines (Google Scholar, PsycInfo and EconLit), crowdsourcing with forms to share articles (on social media and through targeted email lists) and snowballing of relevant articles (including articles that cited the original paper). There was no restriction for locations or language (see later for a discussion of the ways in which diversity of authorship enabled broader searching). This approach yielded approximately 3,000 articles initially. In a triage phase, team members checked for relevance of articles, removing duplicates and articles that did not meet the criteria (such as articles not relevant to the pandemic, being published outside the inclusion window or simply not relating to any of the 19 claims). Most removals were due to articles that were clearly not relevant but had been submitted for general relevance to behavioural science during the pandemic. Those articles were typically easy to identify; any ambiguous articles were left in to allow the full reviewers to determine relevance later. After that process, 747 articles were used in the initial assessment.

Across the 747 papers, 142 countries were represented (meaning at least one article existed where a country was studied directly). Total volumes by country ranged from 291 papers involving the USA to 1 (24 countries). Other countries with large numbers of papers included the UK (109), Germany (78), Italy (60), China (43) and Brazil (35). Full lists of the total geographical coverage of studies are included in the Supplementary information.

To ensure that we did not miss any major studies, all reviewers were asked to search for any potential additional articles after their

assessments were submitted to the lead author (K.R.). Those articles had to meet the same publication deadline and were only included if they substantively influenced the overall assessment. One such article was identified⁶⁹, whereas one set of interrelated studies was updated to include both original papers, letters to editors and responses to letters^{45,67,104}. The latter three articles were added to the final list for posterity, but did not impact assessments. We therefore only refer to 747 articles unless stating otherwise.

Our aim was to identify the highest level of evidence available for (or null, or against) each claim, with higher ratings going to evidence from field research (that is, studies conducted outside the laboratory or survey), particularly highly powered evidence from multiple field studies and settings. This approach addressed two dimensions of evidence quality assessment akin to ecological validity. We use the word 'support' here when discussing evidence in which the original claim appears to be valid. However, major findings could also simply inform understanding, such as a highly powered but null finding. In this case, support indicates those findings that correlate positively with the intended meaning of the original claim.

We chose not to conduct a systematic review for several reasons. First, we were primarily interested in compiling evidence related to claims and reporting those. Conducting a systematic review would have required refining each of the claims to be more specific than originally written, which risks excluding a large amount of potentially relevant evidence¹³⁷. There were also concerns that early-stage COVID-19 research was not suitable for meta-analyses due to being rushed, small samples (that is, underpowered), weak correlations or lacking appropriate methods, such as randomized controlled trials^{42,43}. Therefore, we took a structured but more pragmatic and inclusive approach to selecting studies that addressed claims. This allowed us to cover a broader range of evidence for both insights and limitations of the work conducted during the pandemic.

Similarly, we asked reviewers in both teams to decide what evidence should or should not be included as part of the process. In a policy context, this means there may be disagreement over what evidence informs or does not inform a particular issue. Those disagreements cannot be resolved through selection criteria and extracting data alone (that is, disagreements would still occur in setting the selection criteria). We also wanted to minimize the possibility that a singular criterion confound (such as only permitting messaging on social media but excluding those in emails or letters) might overly bias expert assessments in the same direction. Because of this approach, we do not provide a PRISMA diagram, as each reviewer had different articles that they felt were or were not valid indications of evidence. Further details and limitations about this are provided in the supplement.

Finally, we included preprints in the study, allowing reviewers to determine the quality and robustness of material alongside material that was published after peer review. This decision was made because, for better or for worse¹³⁸, preprints were extremely visible during the pandemic and often treated (at least by the public) as equivalent to published articles¹³⁹. In addition, by including preprints, we reduced some concerns of publication bias by ensuring articles that may have not been published due to null findings could be considered.

Evaluation teams

Two teams conducted the evidence review: 33 authors from the original paper³ were in one team and 39 independent reviewers made up the second team. The use of two independent teams was meant to minimize bias and increase the diversity of perspectives.

The lead author on this article was chosen based on experience in reviewing and reporting behavioural science in public policy contexts^{140,141}, particularly in public health¹⁴²⁻¹⁴⁵, for having coordinated large-team research¹⁴³⁻¹⁴⁵ and for having led the development of an evidence standard¹⁴⁶ for evaluating research for policy. To ensure full independence, the lead author – who was not involved in the original

Analysis

paper – only contributed to assessments by compiling all reviewer ratings, then reviewing articles identified as being ‘best evidence’. If any differences existed, the lead author then followed up to reconcile. In most instances, this involved discussion on what constituted ‘real world’, and the PI proposed a standard for discussion. There were no substantial disagreements with final ratings presented. This approach was intended to promote trust and integrity through transparency in assessment of previous work³⁵ by involving multiple reviewers and many papers in the evaluation.

Reviewers represented institutions from more than 30 countries. There was some representation from every continent, although reviewers were predominantly based in Europe and North America. Searches spanned more languages than English-only articles, particularly white papers and other institutional reports on interventions during the pandemic that might inform the review. The claims that reviewers were assigned to assess were entirely confidential to reduce bias or influence, as were the names of all experts participating in the review. Because all reviewers were trained in the methods and coding system, names were visible in small groups. However, only the lead author knew who had been assigned to assess particular claims.

Procedure

All reviewers followed a standardized approach to assessing each of the claims, which is included in the Supplementary information along with a brief tutorial. In each instance, reviewers received a set of articles assigned by the lead author. As the volume of papers varied substantially across claims (fewer than ten were found for claims 8 and 10; more than 100 were identified for claim 7), reviewers had different numbers of articles to review. For the larger volume claims, some procedural adjustments were made in which reviewers only assessed a subset of articles (a plan was established to address any issues created in the event someone then missed highly relevant material, but this only occurred once and was easily resolved). Each reviewer was required to read the articles in the list, noting four primary aspects:

- (1) Was the article relevant to the claim?
- (2) What was the level of evidence?
 - No evidence, only opinions, perspectives, general theory or anecdotes
 - Some empirical evidence but in limited settings (laboratories, surveys and online)
 - A field study in a real-world setting
 - Replicated evidence in field studies or other natural settings
 - Wide-scale evidence from multiple field studies, policy evaluations or other natural settings
- (3) If there was any empirical evidence, was it in support (positive), null or against (negative) the claim?
- (4) If there was any empirical evidence, what was the general effect size (null–small–medium–large)?

After reviewing all articles, reviewers were asked to produce a summary of evidence for the claim, covering the same four themes. To focus the summary claim evaluation on the highest levels of evidence available, reviewers were asked to rate specifically based on the highest-quality studies reviewed. In other words, rather than averaging all available evidence (as in a meta-analysis), reviewers focused their summary assessments of the highest levels of evidence. For this work, ‘highest’ was defined by the 1–5 scale, supported in terms of statistical power, the appropriateness of method and scope (breadth of settings). This procedure is more relevant in a policy context, in which it is preferable to evaluate the highest-quality evidence available than to estimate average effects from all available evidence (all individual article ratings are available). Causality was not directly factored into the rating given that it is not typically weighted more heavily in policy, and because it might have sidelined rare-but-valuable field studies during the pandemic and only taken stock of well-controlled online experiments.

All reviewers were actively encouraged to assign their own ratings for articles and claims, and there was no attempt to force unanimity across raters. Despite this, average inter-rater agreement (percent of times within each article that reviewer scores agreed) was high (77.5%). This value is generally good, but especially so considering that some claims had so many relevant papers that it was not possible for the same reviewers to assess them all. In addition, several papers were rated by some assigned reviewers but not others when the reviewers disagreed regarding the relevance of the paper to the claim (those judging a paper to be irrelevant were not asked to evaluate it).

We emphasized effect sizes and direction of associations rather than statistical significance to address major concerns regarding validity and replicability. Relying only on *P* values would have increased the likelihood that publication bias and misrepresentation could have overstated evidence in the papers that we reviewed. Following well-known concerns^{147,148}, our guidelines did not involve *P* values; only effect sizes were discussed given that they are more predictive of replicability¹⁴⁹. This approach was largely supported by the consistently large sample sizes of studies included (see Tables 2–6).

Furthermore, including preprints in the assessment reduced the potential role of publication bias. Similarly, we included the direction of effects (positive or negative) to both clarify whether the finding generally agreed with the claim and to account for possible backfire effects, and the harmful side effects that might entail (for example, if a study showed a messaging intervention did have the expected effect, but also created more severe harms, it was possible for the assessment to reflect this).

Evaluations were submitted to the lead author; no reviewer was allowed to see other reviews. The lead author anonymized evaluations to a central coordinating team that checked evaluations for mistakes or inconsistencies (for example, ratings that did not align with the article noted as highest evidence or with the summary statement). All material from these processes have been compiled and posted in an interactive format for public use (<https://tabsoft.co/3xZwIbD>). Note that all values related to sample sizes have been provided as a general indication within and across claims, reflecting a simplified summary of studies reviewed (where samples were clearly reported). As discussed in ‘Evidence for 19 pandemic behaviour claims’, there can be debate as to what should be considered a sample size as well as challenges in extracting values. Since these were not the focus of the research, we provide them for general context and have posted the compiled spreadsheet for those that may wish to revisit or recalculate.

A detailed discussion of all procedures and their limitations is provided in the Supplementary information.

Reporting summary

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

Data availability

All data and study material are provided either in the Supplementary information or through the two online repositories (OSF and Tableau Public, both accessible via <https://psyarxiv.com/58udn>). No code was used for analyses in this work.

137. Wang, Z., Nayfeh, T., Tetzlaff, J., O’Blenis, P. & Murad, M. H. Error rates of human reviewers during abstract screening in systematic reviews. *PLoS ONE* **15**, e0227742 (2020).
138. Guterman, E. L. & Braunstein, L. Z. Preprints during the COVID-19 pandemic: public health emergencies and medical literature. *J. Hosp. Med.* **15**, 634–636 (2020).
139. Wingen, T., Berkessel, J. B. & Dohle, S. Caution, preprint! Brief explanations allow nonscientists to differentiate between preprints and peer-reviewed journal articles. *Adv. Methods Pract. Psychol. Sci.* <https://doi.org/10.1177/25152459211070559> (2022).
140. Ruggeri, K. *Psychology and Behavioral Economics: Applications for Public Policy* (Routledge, 2022).

141. Ruggeri, K. (ed.). *Behavioral Insights for Public Policy: Concepts and Cases* (Routledge, 2019).
142. Ruggeri, K., Benzerga, A., Verra, S. & Folke, T. A behavioral approach to personalizing public health. *Behav. Public Policy* <https://doi.org/10.1017/bpp.2020.31> (2020).
143. Ruggeri, K. et al. Replicating patterns of prospect theory for decision under risk. *Nat. Hum. Behav.* **4**, 622–633 (2020).
144. Ruggeri, K. et al. The general fault in our fault lines. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-021-01092-x> (2021).
145. Ruggeri, K. et al. The globalizability of temporal discounting. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-022-01392-w> (2022).
146. Ruggeri, K. et al. Standards for evidence in policy decision-making. *Social Sciences* [socialsciences.nature.com/users/399005-kai-ruggeri/posts/standards-for-evidence-in-policy-decision-making](https://www.nature.com/users/399005-kai-ruggeri/posts/standards-for-evidence-in-policy-decision-making) (2020).
147. Leek, J. T. & Peng, R. D. Statistics: *P* values are just the tip of the iceberg. *Nature* **520**, 612 (2015).
148. Altman, N. & Krzywinski, M. Interpreting *P* values. *Nat. Methods* **14**, 213–214 (2017).
149. Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J. & Reinero, D. A. Contextual sensitivity in scientific reproducibility. *Proc. Natl Acad. Sci. USA* **113**, 6454–6459 (2016).

Acknowledgements We thank Z. Ji, M. Nair Dedhia, A. Lazara, G. Wilson, J. Usseglio, A. Asfa Durrani and M. Kobotis, as well as Corpus Christi College and the Centre for Business Research, University of Cambridge; and S. Kousta. K.R. reports financial support from the National Science Foundation (2218595). P.B. reports funding from (1) Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brazilian Federal Agency for the Support and Evaluation of Graduate Education) — 88887.310255/2018; (2) Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brazilian Federal Agency for the Support and Evaluation of Graduate Education) — 1133/2019; and (3) the Ministry of Science, Technology and Innovation | Conselho Nacional de Desenvolvimento Científico e Tecnológico (National Council for Scientific and Technological Development) — 309905/2019-2. E.P. reports financial support from National Science Foundation grants (SES-2017651 and SES-2022478). K.M.D. reports financial support from the European Research Council (101018262). R.S.R. reports financial support from the Canadian Institutes of Health Research (172681). Elements of Fig. 1 come from Apple Keys software.

Author contributions K.R. conceived, directed and developed all aspects of the study, blinded reviews and supervised all activities described, as well as wrote, edited and managed the manuscript, visualization and complementary materials. J.J.V.B. and R.W. share senior authorship and contributed to the methods, writing and analysis (the order was determined by the 2022 NBA playoff pool winner). S.A.H., V.C., P.B., N.E., A.C., K.M.D., D.R., S.v.d.L., M.C., E.F., J.N.D., M.J.A.W., R.E.P., J.A.T., A.S., M.G., D.P., J.J., P.V.L., G.P., E.P., K.B., A.C., K.W., L.N., N.T., J.Z., L.S., S.K., D.M., J.J.V.B. and R.W. were original authors that reviewed articles for the 19 claims. F.S., A.L.T., A.H., S.V., V.B., M.F., M.H., R.M., R.K., M.R.T., S.K.Y., R.S.R., Ž.L., Y.Y., S.T.J.H., L.M., I.S., E.D., S.G., H.J., T.W., J.B., S.M., L.M., F.P., B.V., Z.A., E.K.B., M.L., F.T., J.L.A., A.B., Z.Z., L.W., J.N., M.V., D.K., P.C., N.R., M.G., K.L.M. and M.P. were independent reviewers. K.R., F.S., S.A.H., V.C., P.B., N.E., A.C., K.M.D., S.v.d.L., M.C., E.F., J.N.D., M.J.A.W., R.P., M.G., P.V.L., E.P., K.B., K.W., L.N., L.S., C.S., A.L.T., S.K.Y., R.S.R., Y.Y., I.S., E.D., J.B., F.P., K.L.M., J.J.V.B. and R.W. contributed to writing and revision of the manuscript. C.S. was an original author that contributed to the manuscript. S.A.H. made major contributions to the revised manuscript. F.S., S.A.-J., K.R.H., L.N. and P.T. supported evidence compilation, data management and visualization. P.B. produced Fig. 1. A.H., K.R.H., L.N. and P.T. screened articles and contributed to various data collation tasks.

Competing interests D.R. reports receiving research funding in the form of gifts from Google and Meta. K.B. reports a Director role at Eli Lilly. C.S. reports being an unpaid advisor for the US Department of Homeland Security and a senior advisor to the Boston Consulting Group. J.A.T. reportings receiving administrative and grant funding from Facebook, collaborating with Meta and serving as a Senior Advisor at Kroll. The lead author (K.R.) did not permit any influence over the final version of the writing. All other authors declare no competing interests.

Additional information

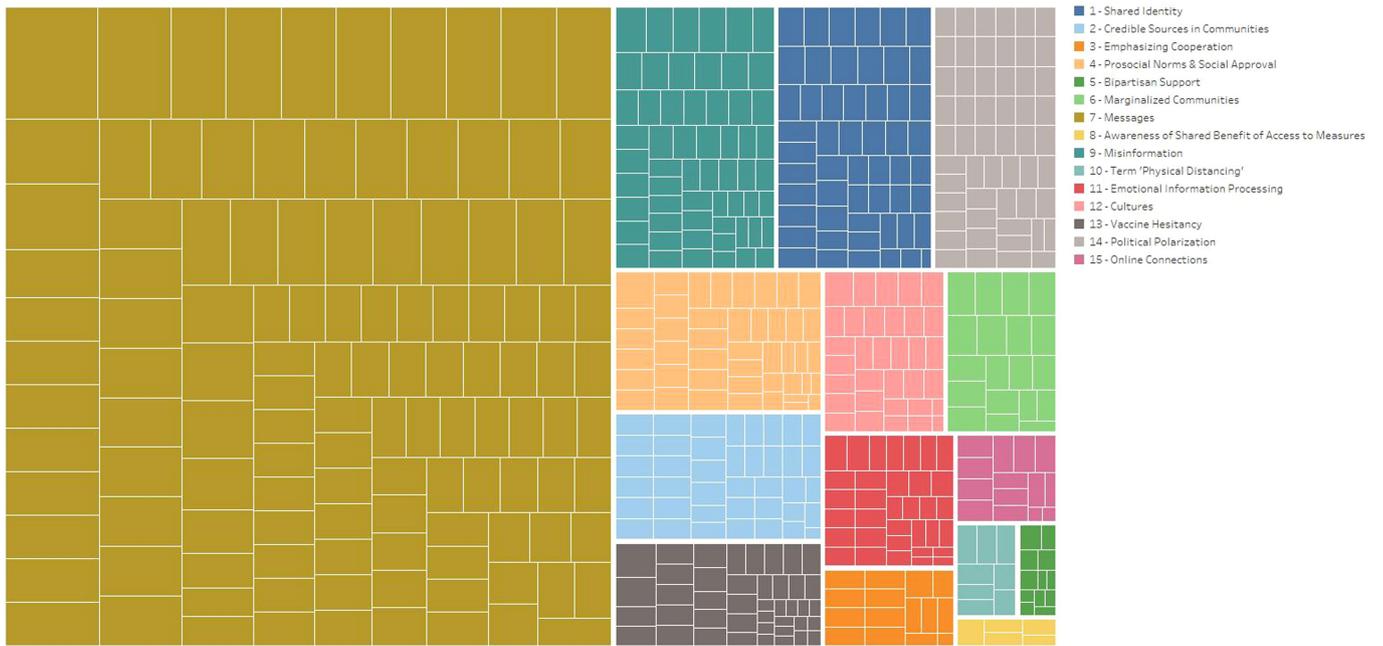
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-023-06840-9>.

Correspondence and requests for materials should be addressed to Kai Ruggeri.

Peer review information *Nature* thanks Hu Chuan-Peng, Matthew Page, Kevin Munger and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Analysis



Extended Data Fig. 1 | Breakdown of articles reviewed by claim. Each box in the figure represents an article reviewed, with the color indicating which claim it fit under. For this figure, all sections within claim 7 are treated as a single claim

due to most papers in that claim covering multiple sections. The size of each box reflects how many reviewers gave the article a rating. Specific claim colors are indicated in the legend.

Extended Data Table 1 | Articles reviewed with evidence by claim

Claim	Articles reviewed	Articles given rating	Rated + evidence
1	54	53	48
2	40	40	39
3	16	15	13
4	45	44	39
5	12	12	12
6	19	19	18
7	141		
7i		41	40
7ii		51	50
7iii		20	20
7iv		27	27
7v		15	15
8	5	5	4
9	60	58	41
10	8	8	3
11	34	34	28
12	45	33	32
13	60	45	37
14	54	54	54
15	15	14	9
Removed	174		
Total	747	518	463

Each row represents the articles reviewed by claim. Articles will not add up to the totals as 25 papers were initially reviewed under more than one claim. Of the 747 included after the triage phase (from over 3,000), 174 were later determined not to fit specifically to a claim. Of the 518 that received at least at least one rating, 118 were reviewed under more than one claim (the increase and largest impact was created by separating claims 7i-7v). Of those, 463 articles were deemed to have empirical evidence, which refers to the sample sizes described in the Main Text. Of those articles, 100 were reviewed under more than one claim. Also note that receiving a rating (with or without evidence) did not necessarily mean an article was deemed to validate or refute a claim. For example, though there were eight articles reviewed under claim 10, none of these were deemed to directly address the claim itself.

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

All data utilized in this study came from qualitative summaries of detailed reviews of published evidence. Studies were compiled using online search engines (Google Scholar, NCBI, PsycINFO, EconLit), crowdsourcing on social media (primarily Twitter and LinkedIn), and through a Qualtrics upload document. Study information were then stored on a shared spreadsheet (Google Sheets) before downloading to Excel. Interactive data were eventually uploaded on the Tableau Public software platform. Some methods details were extracted from each study, such as the country or countries where the study took place, sample size, general method, and source of publication.

Data analysis

The goal of this paper was to synthesize evidence, therefore the primary methods involved to analyze claims are described qualitatively. No inferential statistical analyses are presented in the main text. The only quantitative analyses are presentation of raw numbers, typically divided by 19 (the number of claims assessed). A small number of mean ratings are also produced but no NHST is presented with them. There are two mentions of t-tests in supplemental information, but these are only for indications of internal consistency and rely on small-n analyses.

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](#)

All data and study material is provided either in the supplementary information or through the two online repositories (OSF and Tableau Public, both accessible via <https://psyarxiv.com/58udn>). The precise OSF link will be provided with the final publication (we will upload all material once there are final versions of the manuscript and supplement). It will also be accessible via the current page (<https://psyarxiv.com/58udn>), where the pre-print and interactive tool are already available.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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

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	This study involved compiling and assessing all publications related to behavioral science during the pandemic to inform scientists and policymakers about the accuracy of expectations made in April 2020. Studies were then rated based on the level of evidence they provided, ranging from pure theory/opinion to large-scale, real-world impact.
Research sample	All data used involved publicly available scientific materials that related to or directly tested claims made the Van Bavel et al., 2020 article. No human participants were involved. All sampling involved reviewing studies published from late 2019 through May 2022. A final repository of 747 articles was used for the analyses in this study.
Sampling strategy	Articles and reports were identified and considered for the assessment through extensive systematic and manual searches by all evaluators, with the primary criterion being that they were publicly available prior to June 1, 2022. Searches included using the systematic review criteria produced by PubMed-NCBI for research specifically on COVID-19, as well as checking preprint servers, multiple repository search engines, crowdsourcing (on social media and targeted email lists), and snowballing of relevant articles (including articles that cited the original paper). There was no restriction for locations or language (see later for a discussion of the ways in which diversity of authorship enabled broader searching).
Data collection	All data utilized in this study came from qualitative summaries of detailed reviews of published evidence. Studies were compiled using online search engines (Google Scholar, NCBI, PsycINFO, EconLit), crowdsourcing on social media (primarily Twitter and LinkedIn), and through a Qualtrics upload document. Study information were then stored on a shared spreadsheet (Google Sheets) before downloading to Excel. Interactive data were eventually uploaded on the Tableau Public software platform.
Timing	The search for articles begin formally in February 2022 and was stopped on May 31, 2022. Some papers that met the full criteria were identified through snowballing or brought to the attention of the research team through other means (e.g., news reports, social media posts, papers identified in the review of studies included originally). However, studies were only permitted to be added after the 1st of June, 2022, if they met all other inclusion criteria. This approach was taken to ensure we minimized the possibility of missed studies due to the short turnaround for search engines to identify papers produced in a generally narrow timeline.

Data exclusions	No studies published on/after June 1, 2022, were permitted. Studies published in late 2019 were permitted, provided that they specifically focused on COVID-19. Some studies were published as pre-prints prior to June 1, 2022, but were later identified as having been published through peer-review after the deadline. When this was identified, reviewers were encouraged to use the peer-reviewed version as it was presumed to be more reliable evidence.
Non-participation	Not applicable.
Randomization	Not applicable.

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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging