Surfacing norms to increase vaccine acceptance

Alex Moehring[1,2], **Avinash Collis**[2,3,†], **Kiran Garimella**[2,4,†], **M. Amin Rahimian**[2,5,†], **Sinan Aral**[1,2,4], **and Dean Eckles**[1,2,4,‡]

[1]MIT Sloan School of Management; [2]MIT Initiative on the Digital Economy; [3]McCombs School of Business, The University of Texas at Austin; [4]Institute for Data, Systems, and Society, Massachusetts Institute of Technology; [5]Department of Industrial Engineering, University of Pittsburgh

Working paper
First version: 8 February 2021
This version: 18 June 2021

Despite the availability of multiple safe vaccines, vaccine hesitancy may present a challenge to successful control of the COVID-19 pandemic. As with many human behaviors, people's vaccine acceptance may be affected by their beliefs about whether others will accept a vaccine (i.e., descriptive norms). However, information about these descriptive norms may have different effects depending on the actual descriptive norm, people's baseline beliefs, and the relative importance of conformity, social learning, and free-riding. Here, using a large, pre-registered, randomized experiment (N=484,239) embedded in an international survey, we show that accurate information about descriptive norms can substantially increase intentions to accept a vaccine for COVID-19. These positive effects (e.g., reducing by 4.9% the fraction of people who are "unsure" or more negative about accepting a vaccine) are largely consistent across the 23 included countries, but are concentrated among people who were otherwise uncertain about accepting a vaccine. Providing this normative information in vaccine communications partially corrects individuals' apparent underestimation of how many other people will accept a vaccine. These results suggest that public health communications should present information about the widespread and growing acceptance of COVID-19 vaccines.

Keywords: COVID-19, descriptive norms, social influence, vaccine hesitancy, public health

† A.C., K.G., and M.A.R. contributed equally to this work and are listed alphabetically.
‡ To whom correspondence should be addressed. E-mail: eckles@mit.edu

Introduction

Nonpharmaceutical interventions in response to epidemics, such as the COVID-19 pandemic, often depend on the behavioral responses of the public for their effectiveness. Even with the availability of vaccines, success depends on people's choices to accept, or even seek out, the vaccine (1), since even low vaccine refusal rates can prevent achieving herd immunity (2, 3). Given the value of individual autonomy and the significant challenges of imposing vaccine mandates (4–6), it is important to understand how public health messaging can increase acceptance of safe and effective COVID-19 vaccines. Many messaging strategies address individual barriers to vaccination, such as complacency and inconvenience (7), as well as perceived risk of both vaccines and the disease (1, 8–10).

It may be important to look beyond individuals to consider how public health messaging can also leverage the significant roles of social networks (broadly defined) in shaping individual vaccination decisions (11–15). Rather than being a small factor, there is growing evidence that people's preventative health behaviors are dramatically influenced by many social and cultural factors, with implications for COVID-19 (16). In the United States, for example, analyses of mobility data during the COVID-19 pandemic revealed that people's mobility behaviors vary with their partisan affiliation (17) and media consumption (18, 19) and are affected by the behaviors of their social connections (20).

Acceptance of COVID-19 vaccines likely involves substantial social influence, but theory is not entirely clear on whether learning how many others are accepting a vaccine will increase or decrease acceptance. Positive peer effects can arise due to information diffusion (21, 22), conformity and injunctive norms (14), inferring vaccine safety and effectiveness from others' choices (23, 24), or pro-social motivations such as altruism (25, 26) and reciprocity (27). On the other hand, negative effects of others' acceptance can arise as a result of free-riding on vaccine-generated herd immunity, even if only partial or local (28, 29). The empirical evidence on when positive peer effects (24, 30, 31) or free-riding (28) may dominate is inconclusive. Furthermore, the effects of incorporating *truthful* information about others' into messaging strategies will depend on what that information is, i.e., how prevalent is vaccine acceptance in a given reference group? In the presence of positive peer effects, we may nonetheless wonder whether the true prevalence is high enough that emphasizing this information increases acceptance. Thus, we need further empirical guidance about scalable and effective messaging strategies leveraging social influence. That is, while some interpretations of the theoretical and empirical literature could motivate emphasizing high rates of vaccine acceptance in public health communications, little is known about how realistic interventions of using messages with factual information about others' vaccine acceptance will affect intentions to accept the

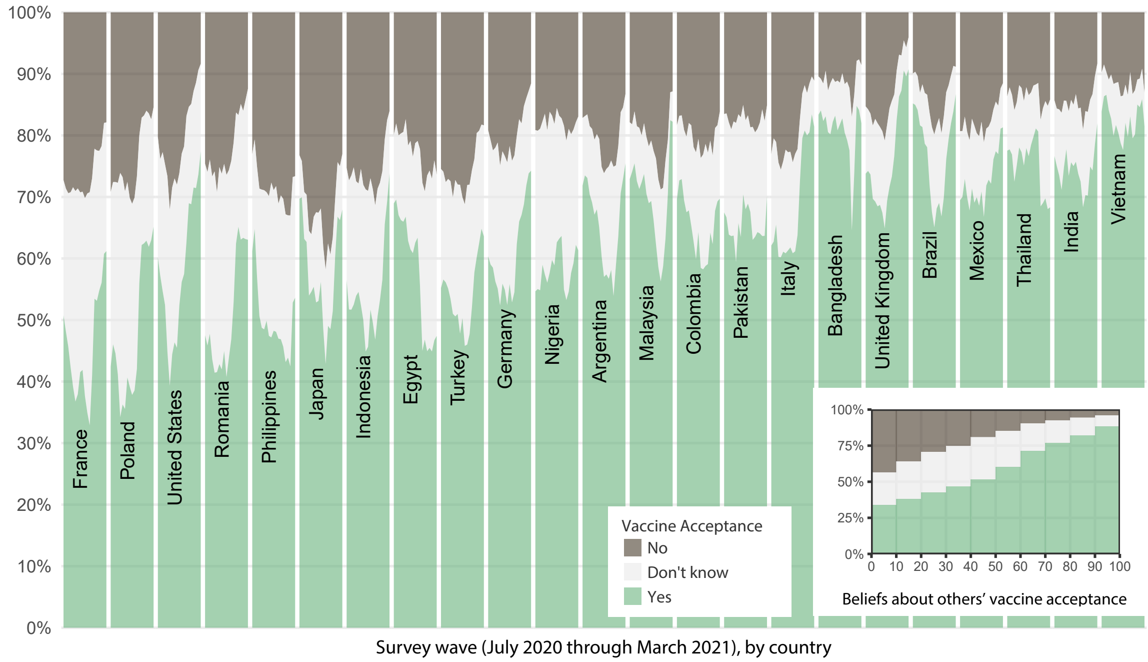


Fig. 1. Time series of COVID-19 vaccine acceptance from July 2020 to March 2021 by country. Shown are the 23 countries with repeated data collection over time. "Yes" also includes respondents indicating they already received a vaccine. Within each country, there are 19 points representing a time-series across the 19 waves of the survey. *(inset)* Pooling data from all 23 countries, people who believe a larger fraction of their community will accept a vaccine are on average more likely to say they will accept a vaccine; this is also true within each included country (Figure S15).

new COVID-19 vaccines.

Here we provide evidence, from a large-scale randomized experiment embedded in an international survey, that accurate information about descriptive norms — what other people do, believe, or say — often has positive effects on intentions to accept new vaccines for COVID-19. Furthermore, we generally rule out large negative effects of such information. To our knowledge, there are no other quantitative causal assessments of how exposure to factual descriptive norms affects intentions to receive the COVID-19 vaccines.

Through a collaboration with Facebook and Johns Hopkins University, and with input from experts at the World Health Organization and the Global Outbreak Alert and Response Network, we fielded a survey in 67 countries in their local languages, yielding over 2.0 million responses (32). This survey assessed people's knowledge about COVID-19, beliefs about and use of preventative behaviors, beliefs about others' behaviors and beliefs, and economic experiences and expectations. Recruitment to this survey was via prominent messages from Facebook to its to users that encouraged potential respondents to help with research on COVID-19 (Figure S1). While it is often impossible to account for all factors that may jointly determine selection into the sample and survey responses, our collaboration with Facebook allows using state-of-the-art, privacy-preserving weighting for non-response using rich behavioral and demographic variables, as well as further weighting to target the adult population of each country (32, 33). All analyses presented here use these survey weights to ensure our results are as representative of these countries' adult populations as possible. Additional information about the weights, and the main analyses replicated without using weights, are in the Supplementary Information (SI) Section S6.3.

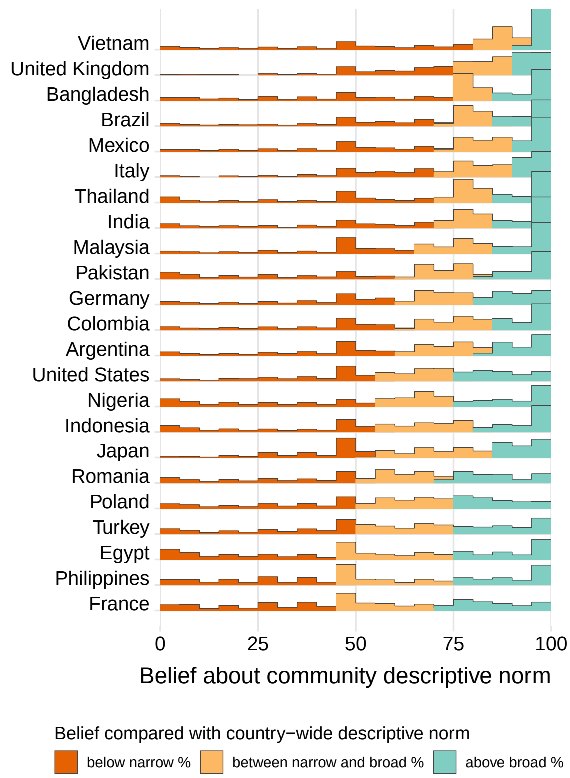


Fig. 2. Within-country distributions of beliefs about descriptive norms ("Out of 100 people in your community, how many do you think would take a COVID-19 vaccine if it were made available?") during the experimental period (October 2020 to March 2021). To enable comparison with actual country-wide potential vaccine acceptance, these histograms are colored by whether they are below (red) the narrow ("Yes" only) definition of vaccine acceptance, between (yellow) the narrow and broad ("Yes" and "Don't know") definitions, or above (teal) the broad definition.

This survey has documented substantial variation in stated intentions to take a vaccine for COVID-19 when one is available to the respondent, with, for example, substantial changes over time and some countries having much larger fractions of people saying they will take a vaccine than others (Figure 1); however, a plurality consistently say they will accept a vaccine and only a (often small) minority say they will refuse one. This is consistent with other smaller-scale national (10, 34) and international (35) surveys. There is also substantial variation in what fraction of *other people* respondents think will accept the vaccine, and these beliefs often substantially differ from country-wide levels of vaccine acceptance (Figure 2). This deviation can have multiple causes, including responding with round numbers; but we posit this is at least partially because some people have incorrect beliefs about descriptive norms. Underestimation of vaccine acceptance by others could be partially caused by processes — such as news coverage of the challenges posed by vaccine hesitancy or diffusion of anti-vaccine messages on social media — that make hesitancy more salient. Beliefs about descriptive norms are in turn positively correlated with vaccine acceptance (Figure 1 inset, Figure S15), likely reflecting many processes, such as geographic and social clustering of vaccine hesitancy, but also causal effects of beliefs about others on intentions to accept a vaccine. Public health communications could present information about norms, perhaps correcting some people's overestimation of the prevalence of vaccine hesitancy. Unlike other ongoing, frequently observable preventative behaviors, like mask wearing, people may have little information about whether others intend to or have accepted a vaccine — which suggests messages with this information could have particularly large effects.

Randomized Experiment

To learn about the effects of providing normative information about new vaccines, beginning in October 2020, for the 23 countries with ongoing data collection in this study, we presented respondents with accurate information based on how previous respondents in their country had answered a survey question about vaccine acceptance, mask wearing, or physical distancing. We randomized at what point in the survey this information was presented, which behavior the information was about, and how we summarized previous respondents' answers — enabling us to estimate the effects of presenting information about descriptive norms on people's stated intentions to accept a vaccine.

In the case of vaccine acceptance, we told some respondents, "Your responses to this survey are helping researchers in your region and around the world understand how people are responding to COVID-19. For example, we estimate from survey responses in the previous month that X% of people in your country say they will take a vaccine if one is made available", where X is the (weighted) percent of respondents saying "Yes" to a vaccine acceptance question. Other respondents received information on how many "say they *may* take a vaccine", which is the (weighted) percent who chose "Yes" or "Don't know" for that same question. (The weighted estimate is preferred to the unweighted estimate and corresponds to the methods used elsewhere in, e.g., dashboards and reports on this survey.) Whether this information occurs before or after a more detailed vaccine acceptance question and whether it uses the

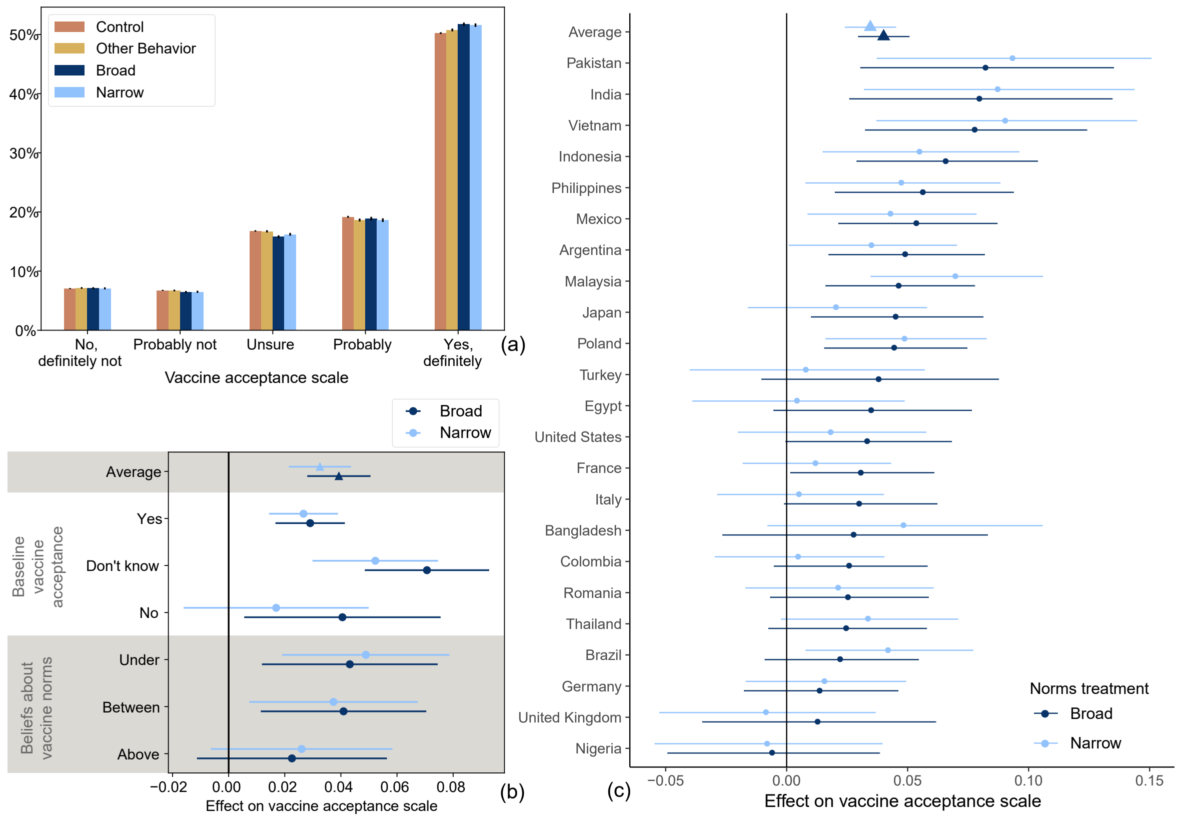


Fig. 3. (a) The normative information treatments shift people to higher levels of vaccine acceptance, whether compared with receiving no information (*control*) or information about other, non-vaccine-acceptance norms (*other behavior*). (b) These estimated effects are largest for respondents who are uncertain about accepting a vaccine at baseline and respondents with baseline beliefs about descriptive norms that are under (rather than above or between) both of the levels of normative information provided in the treatments. (c) While there is some country-level heterogeneity in these effects, point estimates of the effect of the broad normative information treatment are positive in all but one country. Error bars are 95% confidence intervals.

broad (combining "Yes" and "Don't know") or *narrow* ("Yes" only) definition of potential vaccine accepters is randomized — allowing us to estimate the causal effects of this normative information. (When the detailed vaccine acceptance question occurs after the normative information, it is always separated by at least one intervening screen with two questions, and it is often separated by several screens of questions.) Here we focus on comparisons between providing the normative information about vaccines before or after measuring outcomes (e.g., vaccine acceptance); in the SI, we also report similar results when the control group consists of those who received information about other behaviors (i.e., about mask wearing and distancing), which can avoid concerns about differential attrition and researcher demand.

Results

On average, presenting people with this normative information increases stated intentions to take a vaccine, with the broad and narrow treatments causing 0.039 and 0.033 increases on a five-point scale (95% confidence intervals: [0.028, 0.051] and [0.021, 0.044], respectively). The distribution of responses across treatments (Figure 3a) reveals that the effects of the broad (narrow) treatment are concentrated in inducing an additional 1.6% (1.1%) of people to say they will at least "probably" accept the vaccine, and moving 1.9% (1.7%) to "definitely" (Table S8). Note that these statements are about effects on the cumulative distribution of the vaccine acceptance scale (e.g. the proportion answering at least "Probably"). The proportion answering exactly "Probably" is similar across conditions (Figure 3a), consistent with the treatment shifting some respondents from "Unsure" to "Probably" but also some from "Probably" to "Definitely". For the *broad* treatment, this represents a 4.9% relative reduction in the fraction of people choosing a response that is "unsure" or more negative, a 2.4% percent relative increase in the fraction choosing at least "Probably", and a 3.8% percent relative increase in the fraction of people choosing "Yes, definitely". A post hoc analysis also concluded that these effects are largest among people who answer "Don't know" to the baseline vaccine acceptance question (Figure 3b, Table S12), consistent with the idea of targeting vaccine "fence-sitters" (36). As a comparison point, these effects are over a third of the size of the total increase in vaccine acceptance from November 2020 to January 2021 across all 23 countries (0.11 on the five-point scale) — a period that featured frequent and widely-distributed vaccine-related news. (For this comparison we restrict to the time period before vaccines were available to the public as this question was only shown to those who had reported not having already received a vaccine.)

These effects on vaccine acceptance can be at least partially attributed to changes in respondents' beliefs about these descriptive norms. We can examine this because the survey also measured respondents' beliefs about vaccine acceptance in their communities (as displayed in Figure 2), and we randomized whether this was measured before or after providing the normative information. As expected, the normative information treatment increased the fraction of people that the respondents estimate will accept a vaccine (Figure S5). Among those respondents for whom we measured these normative beliefs prior to treatment, we can examine how treatment effects varied by this baseline belief. In particular, we classify respondents according to whether their baseline belief was *above* the broad ("may take") number, *under* the narrow ("will take") number, or *between* these two numbers. (The question measuring beliefs about descriptive norms asks about "your community", while the information provided is for the country. Thus, for an individual respondent, these need not exactly match to be consistent.) Consistent with the hypothesis that this treatment works through revising beliefs about descriptive norms upwards, we find significant effects of the normative information treatment in the groups that may be underestimating vaccine acceptance — the *under* and *between* groups (Figure 3b), though the smaller sample sizes here (since these analyses are only possible for a random subset of respondents) do not provide direct evidence that the effect in the *under* group is larger than that in the *above* group ($p = 0.38$ and $p = 0.31$ for broad and narrow treatments, respectively). We had also hypothesized that the broad and narrow treatments would differ from each other in their effects on respondents in the between group, but we found no such evidence, $p = 0.87$. (In order to be truthful, these treatments also differed in their wording, which could have counterbalanced any effect of the difference in the numbers presented.) A post hoc analysis to address possible mismeasurement due to a preference to report round numbers (by removing those who reported they believe 0%, 50%, or 100% of people in their community would accept a vaccine) was likewise consistent with this hypothesis ($p = 0.03$ and $p = 0.3$ for broad and narrow treatments, respectively, Figure S8).

Having fielded this experiment in 23 countries, we can estimate and compare treatment effects internationally, which may be useful for both national and international communication efforts; however, we caution that estimates for individual countries will necessarily be less precise. Using a linear mixed-effects model, we estimate positive effects in the vast majority of countries (Figure 3c). While estimates for some countries are larger (e.g., Pakistan, Vietnam) and some are smaller (e.g., Nigeria, United Kingdom), most countries are statistically indistinguishable from the grand mean. Furthermore, point estimates of the effect of the *broad* treatment are nearly uniformly positive and we can rule out large negative effects in most countries. Thus, we summarise the results as providing evidence that accurate normative information often increases intentions to accept COVID-19 vaccines with little risk of negative effects. The heterogeneity that is observed in country level treatment effects could be partially explained by the variation in norms which are shown to respondents, with countries with higher baseline vaccine acceptance associated with larger treatment effects (Figure S10). As a more explicit post-hoc test of this, in Figure S11 we group the treatment into bins of width 20 percentage points and find providing higher normative information is associated with larger treatment effects ($p = 0.03$ and $p < 0.001$ for broad and narrow treatments, respectively).

Limitations and Robustness of the Conclusions

An important limitation is that we are only able to estimate effects on intentions to accept a vaccine against COVID-19, which could differ from effects on vaccine uptake. While it is not possible at this stage to study interventions that measure take up of the COVID-19 vaccine on a representative global population, we believe that the intervention studied here is less subject to various threats to validity — such as experimenter demand effects — that are typically a concern in survey experiments measuring intentions.

This randomized experiment was embedded in a survey with a more general advertised purpose that covers several topics, so normative information is not particularly prominent (Figure S1). In this broader survey, only 15% of questions were specific to vaccinations or social norms (32). Furthermore, unlike other sampling frames with many sophisticated study participants (e.g., country-specific survey panels, Amazon Mechanical Turk), respondents are recruited from a broader population (Facebook users). In addition, we observe null results for observable behaviors such as distancing and mask wearing, which would be surprising if researcher demand effects were driving the effects for vaccine acceptance.

A number of robustness checks increase our confidence that experimenter demand is not driving the result. As a first robustness check, we compare the outcome of subjects who receive the vaccine norm treatment to those receiving the treatment providing information about masks and distancing; the treatment effect persists (Table S6). Moreover, we may expect researcher demand effects to be smaller when the information treatment and the outcome are not immediately adjacent. In all cases, for the vaccine acceptance outcome, there is always at least one intervening screen of questions (the future mask-wearing and distancing intentions questions). Furthermore, they are often separated by more than this. We consider a subset of respondents where the treatment and the outcome are separated by at least one "block" of questions between them. Results of this analysis are presented in Figure S12 and Table S13. The estimated effects of the vaccine treatments in this smaller sample are somewhat muted and less precise, but both significantly positive. Moreover, Table S14 shows even with the larger gap between treatment and outcome the information is still moving a relatively large share of people who are unsure or more negative to at least probably accepting the vaccine.

Discussion

Framing vaccination as a social norm has been suggested as an effective approach to building COVID-19 vaccine confidence (15, 37–39), but this recommendation has lacked direct evidence on a scalable messaging strategy using accurate information, which this international randomized experiment now contributes. Brewer et al. (15) document the case of a vaccine campaign by a major pharmacy retail chain in the United States that employed negative norms messaging to emphasize risks to individuals: "Get your flu shot today because 63% of your friends didn't." Although such a strategy can reduce incentives to free-ride on vaccine herd immunity, its broader impact on social norm perceptions may render it ineffective. On the other hand, one might worry that accurate information about descriptive norms would simply feature pluralities or majorities that are too small to be effective. In general, the multimodal effects of descriptive norms on risk perceptions, pro-social motivations, and social conformity highlight the value of the evidence we provide here that accurate normative information often increases intentions to accept COVID-19 vaccines, while generally ruling out large negative effects.

For social norms to be effective it is critical that they are salient in the target population (e.g., wearing badges (40)). While in our randomized experiment norms are made salient through direct information treatments, the results have implications for communication to the public through health messaging campaigns and the news media. For example, because very high levels of vaccine uptake are needed to reach (even local) herd immunity (3), it is reasonable for news media to cover the challenges presented by vaccine hesitancy; but our results suggest that it is valuable to contextualize such reporting by highlighting the widespread norm of accepting COVID-19 vaccines. Public health campaigns to increase acceptance of safe and effective vaccines can include information about descriptive norms. In an effort to influence the public, some public figures have already documented receiving a COVID-19 vaccine in videos on television and social media. The substantial positive effects of numeric summaries of everyday people's intentions documented here suggest that simple factual information about descriptive norms can similarly leverage social influence to increased vaccine acceptance. Some negative attitudes toward vaccination put disadvantaged communities at more risk, so emphasizing country-wide vaccination norms may prove critical for removing susceptible pools and reducing the risk or prevalence of endemic disease (3, 41).

How important are the effects of the factual descriptive normative messages studied here? Smaller-scale interventions that treated individuals with misinformation (42), pro-social messages (43), demographically tailored videos (44), text message reminders (45), or other informational content (46) have yielded similar or smaller effect sizes, while lacking the scalability and practical appeal of accurate descriptive norms. The substantial effects of normative information about vaccine acceptance may reflect that people have little passive exposure to information about how many people in their communities and countries would accept a vaccine, or even have done so already. This result contrasts with other preventative behaviors (mask wearing and distancing), for which we observe smaller or no effects (see SI Section S6), that are both ongoing (i.e., respondents have repeatedly chosen whether to perform them before) and readily observable in public. However, it is possible that as people have more familiarity with social contacts choosing to accept a vaccine, this type of normative information will become less impactful.

How will our results for intentions to accept vaccines translate into vaccine receipt? Prior studies exhibit important concordance between vaccination intentions and subsequent take-up (47) — and effects of treatments on each (48, 49). In a supplementary survey conducted in the United States over two waves we found that self-reported vaccination intentions were predictive of future (self-reported) vaccination status (see SI Section S9). If respondents in our international experiment were to be vaccinated at the same rate as those in this supplementary analysis, we would see a 23.1 percentage point increase in vaccination rates among those who were unsure but were induced to say they would probably accept a vaccine and a 17.2 percentage point increase in vaccination rates among those who would probably accept a vaccine but were induced to say they would definitely accept a vaccine. Of course, to what degree effects on intentions translate into increased vaccination depends on, e.g., the ease of getting vaccinated. Thus, we encourage the use of these factual normative messages, as examined here; but we also emphasize the need for a range of interventions that lower real and perceived barriers to vaccination, remind people to get vaccinated (45), and leverage descriptive norms and social contagion more generally, such as in spreading information about how to obtain a vaccine (21).

Materials and Methods

Experiment analysis. The results presented in the main text and elaborated on in the SI each use a similar pre-registered methodology that we briefly describe here. For the results in Figure 3a, we estimate the following linear regression:

$$Y_i = \delta_0 + \sum_{j \in J} \delta_j D_i^j + \gamma X_i + \sum_{j \in J} \eta_j X_i D_i^j + \varepsilon_i, \qquad [1]$$

where Y_i is the outcome for individual i, D_i^j is an indicator if individual i received treatment $j \in J = \{\text{Broad, Narrow}\}$, and X_i is a vector of centered covariates (50, 51). All statistical inference uses heteroskedasticity-consistent Huber–White "sandwich" estimates of the variance–covariance matrix.

For heterogeneous treatment effects (Figure 3b), we estimate a similar regression where covariates are centered at their subgroup-specific means:

$$Y_i = \sum_{b \in B} 1[b_i = b] \left(\delta_0^b + \sum_{j \in J} \delta_j^b D_{ij}^b + \gamma X_i + \sum_{j \in J} \eta_j^b X_i D_{ij}^b \right) + \varepsilon_i. \qquad [2]$$

Mixed-effects model. In the main text and Figure 3c, we report results from a linear mixed-effects model with coefficients that vary by country. This model is also described in our preregistered analysis plan. Note that the coefficients for the overall (across-country) treatments effects in this model differ slightly from the estimates from the model in equation 1; that is, the "Average" points in Figure 3b and 3c do not match exactly. As noted in our analysis plan, "sandwich" standard errors are not readily available here, so reported 95% confidence intervals are obtained by estimating the standard errors via a bootstrap.

Acknowledgments. The COVID-19 Global Beliefs, Behaviors, and Norms Survey was a collaborative effort involving contributions from individuals at multiple institutions, especially the Massachusetts Institute of Technology, Johns Hopkins University, the Global Outbreak Alert and Response Network, the World Health Organization, and Facebook, with key contributions from Stella Babalola, Nina Gobat, Esther Kim, Kelsey Mulcahy, Praveen Raja, Stephanie Sasser, Dominick Shattuck, Jeni Stolow, Carlos Velasco, and Thomas Wynter. We thank Eytan Bakshy, Adam Berinsky, Daniel Bjorkegren, Alex Leavitt, Solomon Messing, David Rand, and anonymous reviewers for helpful comments on this work. We thank the millions of respondents to this survey worldwide.

Funding. This work was funded in part by a grant from Facebook to the MIT Initiative on the Digital Economy.

Competing interests. In addition to funding this work, Facebook has sponsored a conference organized by S.A. and D.E.; M.A.R. serves on the advisory committee of a vaccine confidence fund created by Facebook and Merck.

Author contributions. All authors contributed to the design of the survey and the randomized experiment. A.M. and D.E. wrote the preregistered analysis plan. A.M. led the data analysis, with contributions from all authors. All authors contributed to writing the paper.

Data and materials and availability. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Documentation of the survey instrument and aggregated data from the survey are publicly available at https://covidsurvey.mit.edu. Researchers can request access to the microdata from Facebook and MIT at https://dataforgood.fb.com/docs/preventive-health-survey-request-for-data-access/. Preregistration details are available at https://osf.io/h2gwv/. Analysis code for reproducing the results will be made public. The Committee on the Use of Humans as Experimental Subjects at MIT approved both the survey and embedded randomized experiment as exempt protocols.

References

1. A. A. Malik, S. M. McFadden, J. Elharake, S. B. Omer, Determinants of COVID-19 vaccine acceptance in the US, *EClinicalMedicine* **26**, 100495 (2020).

2. S. Sanche, *et al.*, High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2, *Emerging Infectious Diseases* **26**, 1470 (2020).

3. R. M. Anderson, C. Vegvari, J. Truscott, B. S. Collyer, Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination, *The Lancet* **396**, 1614 (2020).

4. C. Signorelli, S. Iannazzo, A. Odone, The imperative of vaccination put into practice, *The Lancet Infectious Diseases* **18**, 26 (2018).

5. S. B. Omer, C. Betsch, J. Leask, Mandate vaccination with care, *Nature* pp. 469–472 (2019).

6. C. Betsch, R. Böhm, Detrimental effects of introducing partial compulsory vaccination: Experimental evidence, *The European Journal of Public Health* **26**, 378 (2016).

7. C. Betsch, R. Böhm, G. B. Chapman, Using behavioral insights to increase vaccination policy effectiveness, *Policy Insights from the Behavioral and Brain Sciences* **2**, 61 (2015).

8. B. Nyhan, J. Reifler, S. Richey, G. L. Freed, Effective messages in vaccine promotion: a randomized trial, *Pediatrics* **133**, e835 (2014).

9. B. Nyhan, J. Reifler, Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information, *Vaccine* **33**, 459 (2015).

10. J. Green, *et al.*, Report 36: Evaluation of COVID-19 vaccine communication strategies, https://osf.io/d36nh/ (2021).

11. B. Nyhan, J. Reifler, S. Richey, The role of social networks in influenza vaccine attitudes and intentions among college students in the southeastern United States, *Journal of Adolescent Health* **51**, 302 (2012).

12. E. K. Brunson, The impact of social networks on parents' vaccination decisions, *Pediatrics* **131**, e1397 (2013).

13. C. T. Bauch, D. J. Earn, Vaccination and the theory of games, *Proceedings of the National Academy of Sciences* **101**, 13391 (2004).

14. T. Oraby, V. Thampi, C. T. Bauch, The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases, *Proceedings of the Royal Society B: Biological Sciences* **281**, 20133172 (2014).

15. N. T. Brewer, G. B. Chapman, A. J. Rothman, J. Leask, A. Kempe, Increasing vaccination: Putting psychological science into action, *Psychological Science in the Public Interest* **18**, 149 (2017).

16. S. S. DeRoo, N. J. Pudalov, L. Y. Fu, Planning for a COVID-19 vaccination program, *JAMA* **323**, 2458 (2020).

17. H. Allcott, *et al.*, Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic, *Journal of Public Economics* **191**, 104254 (2020).

18. A. Bridgman, *et al.*, The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media, *Harvard Kennedy School Misinformation Review* **1** (2020).

19. A. Simonov, S. K. Sacher, J.-P. H. Dubé, S. Biswas, The persuasive effect of Fox News: Non-compliance with social distancing during the COVID-19 pandemic, https://www.nber.org/papers/w27237 (2020).

20. D. Holtz, *et al.*, Interdependence and the cost of uncoordinated responses to COVID-19, *Proceedings of the National Academy of Sciences* **117**, 19837 (2020).

21. A. Banerjee, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, Using gossips to spread information: Theory and evidence from two randomized controlled trials, *The Review of Economic Studies* **86**, 2453 (2019).

22. V. Alatas, A. G. Chandrasekhar, M. Mobius, B. A. Olken, C. Paladines, When celebrities speak: A nationwide Twitter experiment promoting vaccination in Indonesia, https://www.nber.org/papers/w25589 (2019).

23. C. T. Bauch, S. Bhattacharyya, Evolutionary game theory and social learning can determine how vaccine scares unfold, *PLoS Comput Biol* **8**, e1002452 (2012).

24. N. Rao, M. M. Mobius, T. Rosenblat, Social networks and vaccination decisions, https://www.bostonfed.org/publications/research-department-working-paper/2007/social-networks-and-vaccination-decisions.aspx (2007).

25. J. T. Vietri, M. Li, A. P. Galvani, G. B. Chapman, Vaccinating to help ourselves and others, *Medical Decision Making* **32**, 447 (2012).

26. E. Shim, G. B. Chapman, J. P. Townsend, A. P. Galvani, The influence of altruism on influenza vaccination decisions, *Journal of The Royal Society Interface* **9**, 2234 (2012).

27. L. Korn, R. Böhm, N. W. Meier, C. Betsch, Vaccination as a social contract, *Proceedings of the National Academy of Sciences* **117**, 14890 (2020).

28. Y. Ibuka, M. Li, J. Vietri, G. B. Chapman, A. P. Galvani, Free-riding behavior in vaccination decisions: An experimental study, *PloS one* **9**, e87164 (2014).

29. R. Böhm, C. Betsch, L. Korn, Selfish-rational non-vaccination: Experimental evidence from an interactive vaccination game, *Journal of Economic Behavior & Organization* **131**, 183 (2016).

30. J. C. Hershey, D. A. Asch, T. Thumasathit, J. Meszaros, V. V. Waters, The roles of altruism, free riding, and bandwagoning in vaccination decisions, *Organizational Behavior and Human Decision Processes* **59**, 177 (1994).

31. R. Sato, Y. Takasaki, Peer effects on vaccination behavior: Experimental evidence from rural Nigeria, *Economic Development and Cultural Change* **68**, 93 (2019).

32. A. Collis, *et al.*, Global Survey on COVID-19 Beliefs, Behaviors, and Norms, https://covidsurvey.mit.edu/covid_survey_files/COVID_BBN_survey_report.pdf (2020).

33. N. Barkay, *et al.*, Weights and methodology brief for the COVID-19 symptom survey by University of Maryland and Carnegie Mellon University, in partnership with Facebook, https://arxiv.org/abs/2009.14675 (2020).

34. K. A. Fisher, *et al.*, Attitudes toward a potential SARS-CoV-2 vaccine: A survey of US adults, *Annals of internal medicine* **173**, 964 (2020).

35. J. V. Lazarus, *et al.*, A global survey of potential acceptance of a COVID-19 vaccine, *Nature Medicine* pp. 1–4 (2020).

36. C. Betsch, L. Korn, C. Holtmann, Don't try to convert the antivaccinators, instead target the fence-sitters, *Proceedings of the National Academy of Sciences* **112**, E6725 (2015).

37. W. Y. S. Chou, C. E. Burgdorf, A. Gaysynsky, C. M. Hunter, COVID-19 vaccination communication: Applying behavioral and social science to address vaccine hesitancy and foster vaccine confidence, https://obssr.od.nih.gov/wp-content/uploads/2020/12/COVIDReport_Final.pdf (2020).

38. R. Palm, T. Bolsen, J. T. Kingsland, The effect of frames on COVID-19 vaccine hesitancy, https://www.medrxiv.org/content/10.1101/2021.01.04.21249241v1.full (2021).

39. C. Chevallier, A. S. Hacquin, H. Mercier, COVID-19 vaccine hesitancy: Shortening the last mile, *Trends in Cognitive Sciences* (2021).

40. J. Riphagen-Dalhuisen, *et al.*, Hospital-based cluster randomised controlled trial to assess effects of a multi-faceted programme on influenza vaccine coverage among hospital healthcare workers and nosocomial influenza in the netherlands, 2009 to 2011, *Eurosurveillance* **18**, 20512 (2013).

41. E. Paul, A. Steptoe, D. Fancourt, Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications, *The Lancet Regional Health-Europe* p. 100012 (2020).

42. S. Loomba, A. de Figueiredo, S. J. Piatek, K. de Graaf, H. J. Larson, Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA, *Nature Human Behaviour* pp. 1–12 (2021).

43. S. Banker, J. Park, Evaluating prosocial COVID-19 messaging frames: Evidence from a field study on Facebook, *Judgment and Decision Making* **15**, 1037 (2020).

44. M. Alsan, *et al.*, Comparison of knowledge and information-seeking behavior after general COVID-19 public health messages and messages tailored for Black and Latinx communities: A randomized controlled trial, *Annals of Internal Medicine* (2020).

45. K. L. Milkman, *et al.*, A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment, *Proceedings of the National Academy of Sciences* **118** (2021).

46. O. Chanel, S. Luchini, S. Massoni, J. C. Vergnaud, Impact of information on intentions to vaccinate in a potential epidemic: Swine-origin Influenza A (H1N1), *Social Science & Medicine* **72**, 142 (2011).

47. B. A. Lehmann, R. A. Ruiter, G. Chapman, G. Kok, The intention to get vaccinated against influenza and actual vaccination uptake of dutch healthcare personnel, *Vaccine* **32**, 6986 (2014).

48. E. T. Bronchetti, D. B. Huffman, E. Magenheim, Attention, intentions, and follow-through in preventive health behavior: Field experimental evidence on flu vaccination, *Journal of Economic Behavior & Organization* **116**, 270 (2015).

49. M. Alsan, S. Eichmeyer, Persuasion in medicine: Messaging to increase vaccine demand, http://www.nber.org/papers/w28593 (2021).

50. W. Lin, Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique, *Annals of Applied Statistics* **7**, 295 (2013).

51. L. W. Miratrix, J. S. Sekhon, B. Yu, Adjusting treatment effect estimates by post-stratification in randomized experiments, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 369 (2013).

52. J. De Quidt, J. Haushofer, C. Roth, Measuring and bounding experimenter demand, *American Economic Review* **108**, 3266 (2018).

53. J. Mummolo, E. Peterson, Demand effects in survey experiments: An empirical assessment, *American Political Science Review* **113**, 517 (2019).

54. R. M. Groves, *et al.*, *Survey methodology*, vol. 561 (John Wiley & Sons, 2011).

Supplementary Information for
"Surfacing norms to increase vaccine acceptance"

Contents

S1. Experiment overview

During an update to the survey on October 28th, 2020, we introduced a prompt to all respondents that provided information about preventative behaviors in their country based on information from the survey. Although this information was provided to all respondents who completed the survey from an eligible country, the information was provided in a random order creating an experiment within the survey. For each eligible respondent, we showed the following message at a random position in the latter part of the survey:

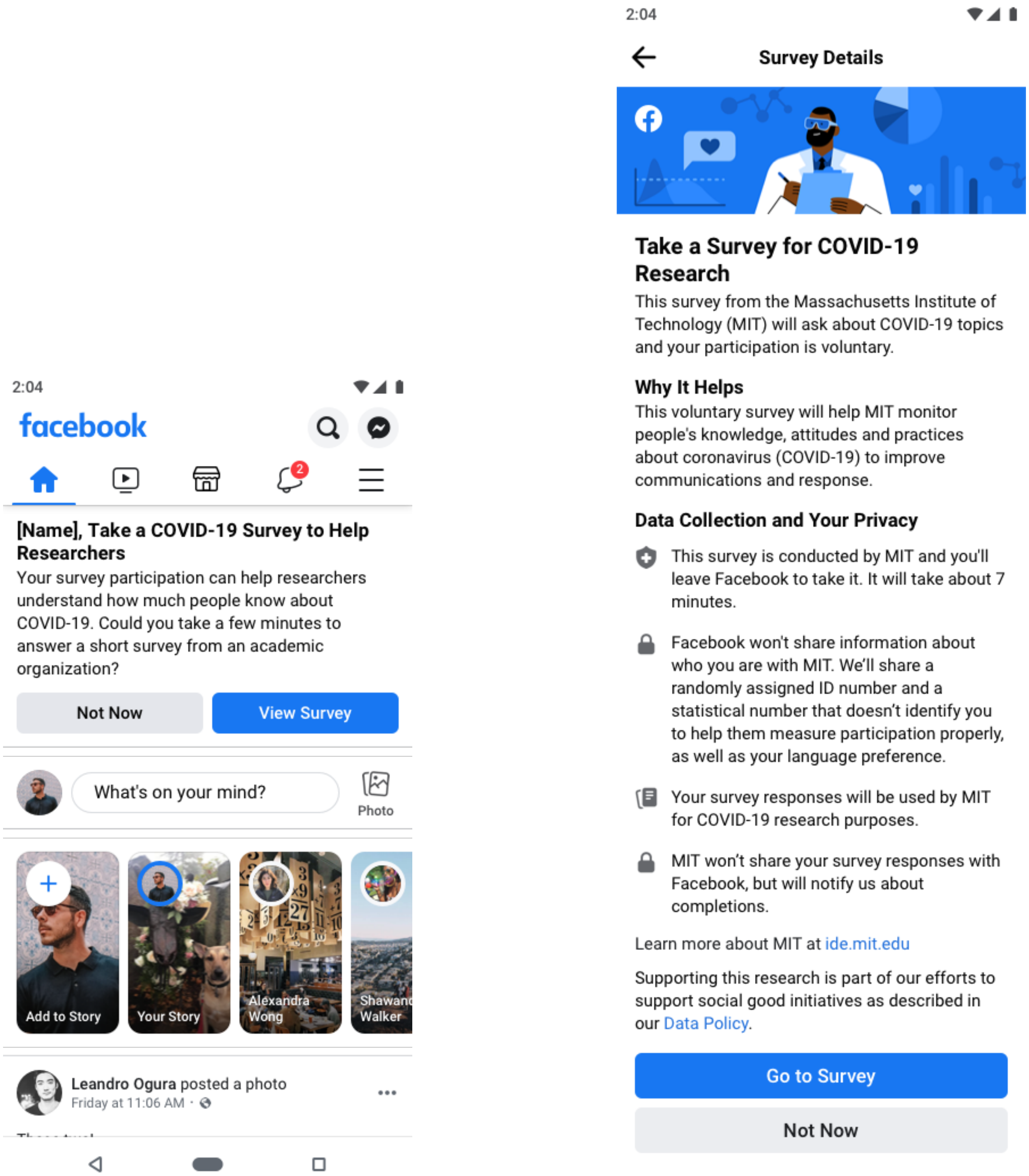
> Your responses to this survey are helping researchers in your region and around the world understand how people are responding to COVID-19. For example, we estimate from survey responses in the previous month that [[country share]]% of people in your country say they [[broad or narrow]] [[preventative behavior]].

We filled in the blanks with one randomly chosen preventative behavior, a broad or narrow definition of the activity, and the true share of responses for the respondent's country. The three behaviors were vaccine acceptance, mask wearing, and social distancing. In the broad condition, we used a more inclusive definition of the preventative behavior and the narrow condition used a more restrictive definition. For example, for vaccine acceptance we either reported the share of people responding "Yes" or the share of people responding "Yes" or "Don't know" to the baseline vaccine acceptance question. The numbers shown, which were updated with each wave, are displayed in Figure S2.

We preregistered our analysis plan, which we also updated to reflect continued data collection and our choice to eliminate the distancing information treatment in later waves. While we describe some of the main choices here, our preregistered analysis plans can be viewed at https://osf.io/h2gwv/. The analysis of the experiment in the main text that is not described in the analysis plan is labeled post hoc (in particular, heterogeneity by baseline vaccine acceptance). One set of more complex analyses speculatively described in the analysis plan (hypothesis 3, "may suggest using instrumental variables analyses") has not yet been pursued.

S2. Data construction

Our dataset is constructed from the microdata described in (32). We first code each outcome to a 5-point numerical scale. We then condition on being eligible for treatment and having a waves survey type (i.e. being in a country with continual data collection) to arrive at the full dataset of those eligible for treatment. (Respondents in the snapshot survey may have received treatment if they self-reported being in a wave country, these individuals are removed as their weights will be for the wrong country.) All randomization and balance checks



(a) Facebook Recruitment Message

(b) Facebook Interstitial

Fig. S1. Facebook Promotion and Interstitial

described as "intent-to-treat" use this dataset. In our preregistered analysis plan, we described how the sample would be restricted to those who completed the survey and for whom we received a full survey completion weight from Facebook. This removes approximately 40% of respondents, resulting in 484,239 respondents. For the main analysis comparing users who received the vaccine information treatment to control users (e.g., in Figure 3b), there are 365,593 respondents.

As in our pre-analysis plan, the following variables are used in our analysis:

1. Outcomes

   (a) Over the next two weeks, how likely are you to wear a mask when in public? [Always, Almost always, When convenient, Rarely, Never]

   (b) Over the next two weeks, how likely are you to maintain a distance of at least 1 meter from others when in public? [Always, Almost always, When convenient, Rarely, Never]

   (c) If a vaccine against COVID-19 infection is available in the market, would you take it? [Yes, definitely, Probably, Unsure, Probably not, No, definitely not]

2. Mediators & Covariates

   (a) Baseline outcomes. These questions are similar to the outcome questions. Only the vaccine question always appears before the treatment in all cases; the others are in a randomized order. Thus, for use of the other covariates for increasing precision, mean imputation is required.

      - Masks. How often are you able to wear a mask or face covering when you are in public? How effective is wearing a face mask for preventing the spread of COVID-19?
      - Distancing: How often are you able to stay at least 1 meter away from people not in your household? How important do you think physical distancing is for slowing the spread of COVID-19?
      - Vaccine: If a vaccine for COVID-19 becomes available, would you choose to get vaccinated? This will be coded as binary indicators for the possible outcomes, grouping missing outcomes with "Don't know".

   (b) Beliefs about norms. These questions will be randomized to be shown before the treatment for some respondents and after treatment for other respondents. This will allow us to study heterogeneity in baseline beliefs, as well as ensure our randomization does impact beliefs.

- Masks: Out of 100 people in your community, how many do you think do the following when they go out in public? Wear a mask or face covering.
- Distancing: Out of 100 people in your community, how many do you think do the following when they go out in public? Maintain a distance of at least 1 meter from others.
- Vaccine: Out of 100 people in your community, how many do you think would take a COVID-19 vaccine if it were made available?

When used in analysis, we require all covariates to be before both treatment and outcome. As the survey contains randomized order for these questions, this ensures that the distribution of question order is the same across treated and control groups and removes any imbalance created by differential attrition. Missing values are imputed at their (weighted) mean.

S3. Randomization checks

Table S1 presents results of a test that the treatment and control shares were equal to 50% as expected. While the final dataset does have some evidence of imbalance that could be caused by differential attrition, the "robust" dataset (described in S6.2) is well balanced and the treatment is balanced across the three behaviors information could be provided about (Table S2). According to our pre-registered analysis plan, in the presence of evidence of differential attrition, we make use of additional analyses that use the information about other behaviors as an alternative control group throughout this supplement.

Table S1. Randomization Tests

	p-val	Treated Share	Control Share
Full	0.011	0.501	0.499
Final	0.081	0.499	0.501
Robust	0.176	0.499	0.501

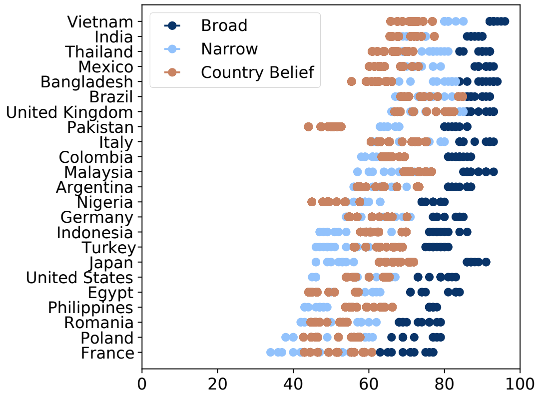
The results of a test that the treated share and control shares equal 50%. The first row uses intent-to-treat on the full set of eligible respondents, the second row uses the final data set after conditioning on eligibility and completing the survey, and the third row uses the subset of responses in the final dataset that have at least one block between treatment and outcome.

In addition, baseline covariates measured before both treatment and the outcome are balanced across treatment and control groups (Table S3). The covariates are also balanced in the final analysis dataset (Table S4) and within treated users across the three possible treatment behaviors (Table S5).

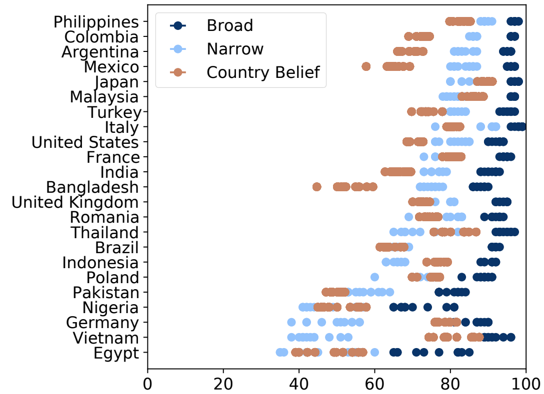
Table S2. Randomization Tests

	Vaccine	Masks	Distancing
Final	0.215	0.218	0.441
Robust	0.210	0.113	0.519

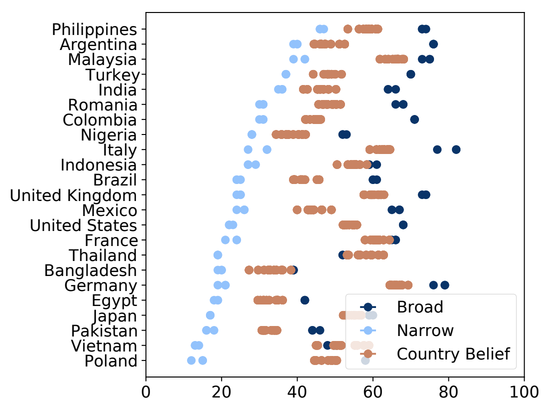
The p-values of a test that each behavior was shown the expected number of times. This reports the results of a joint test that each period share was equal to the expected. For waves 9-12, each behavior was shown 1/3 of the time and for waves 12 on the vaccine treatments were shown to 2/3 of respondents and the mask treatments were shown to 1/3 of respondents. This table cannot include the full dataset intent-to-treat analysis because the behavior randomization occurred when the treatment was shown.



(a) Vaccine Treatments



(b) Mask Treatments



(c) Distancing Treatments

Fig. S2. Treatment Variation

For each behavior (Vaccine, Masks, Distancing), we plot the information provided to subjects based on the broad and narrow definitions of compliance. The treatments were updated every two weeks as new waves of data were included. The points labeled “country belief” display the weighted average belief in a country of how many people out of 100 practice (or will accept, for vaccines) each behavior.

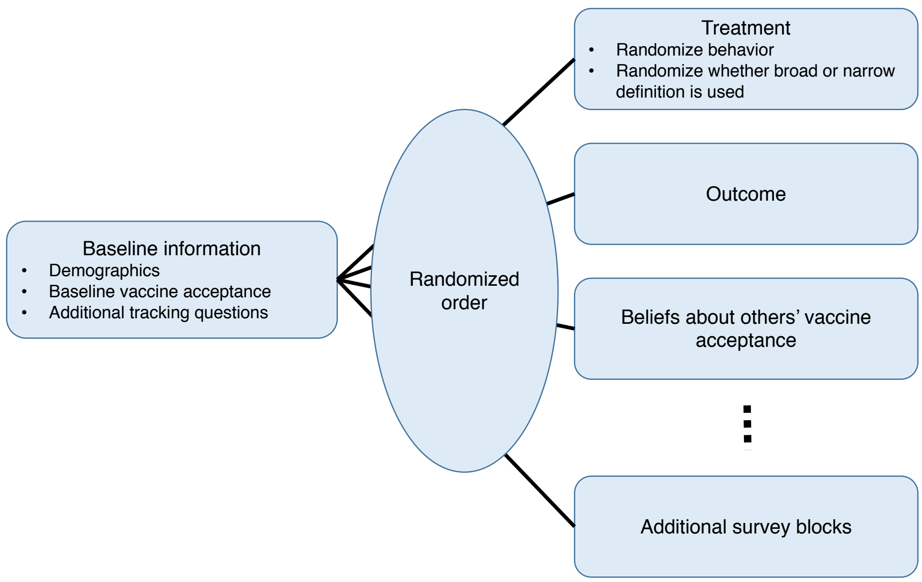
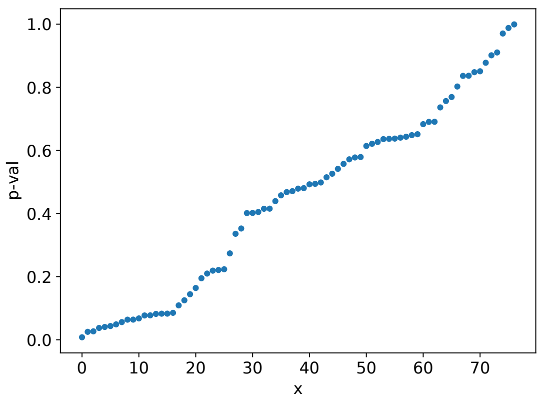
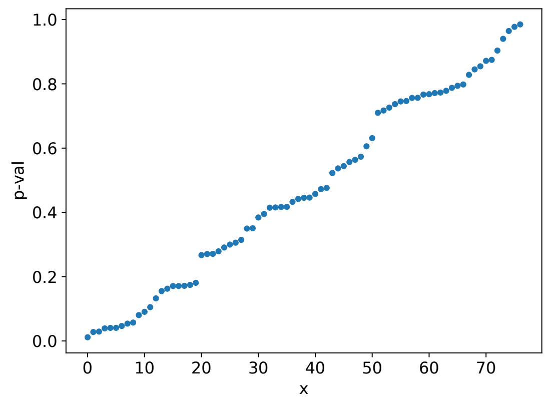


Fig. S3. Experiment Flow

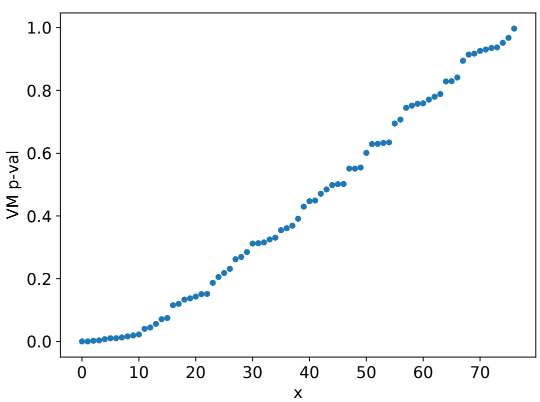
Illustration of the flow of a respondent through the survey. First, they are presented with tracking and demographic questions. They then enter a randomized portion where blocks are in random order. This includes the treatment, outcome, and many of the baseline covariates included in regressions for precision. Recall all covariates used in analysis are only used if they are pre-treatment and outcome.



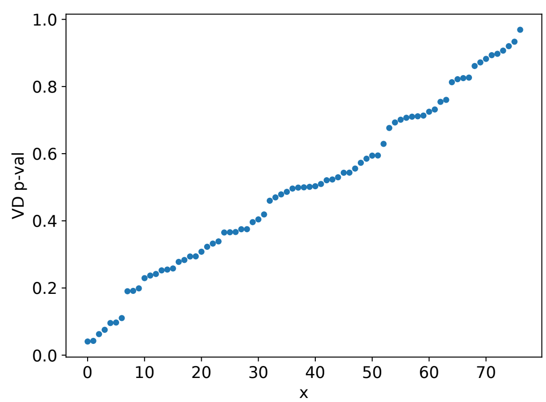
(a) Balance Tests: Intent-to-treat



(b) Balance Tests: Final Sample



(c) Balance Tests: Vaccine vs Mask Treatments



(d) Balance Tests: Vaccine vs Dist Treatments

Fig. S4. Balance Test p-Values

Ordered p-values for the balance tests described in Tables S3, S4, and S5 sorted in ascending order. All available pre-treatment covariates are included, which results in 76 tests. This includes roughly 40 covariates that are not presented in the tables for brevity. These are questions that permit multiple responses, including news media, sources, and trust, and a more detailed list of preventative measures taken.

Table S3. Balance Tests: Intent-to-treat

	p-val	Control		Treated	
age	0.223	2.587	(0.002)	2.583	(0.002)
gender	0.401	1.441	(0.001)	1.440	(0.001)
education	0.468	2.781	(0.001)	2.779	(0.001)
own health	0.068	2.410	(0.002)	2.414	(0.002)
vaccine accept	0.848	1.491	(0.001)	1.491	(0.001)
knowledge existing treatments	0.210	0.218	(0.001)	0.219	(0.001)
info exposure past week	0.439	2.300	(0.001)	2.301	(0.001)
info exposure more less wanted	0.614	2.387	(0.002)	2.386	(0.002)
know positive case	0.405	1.281	(0.002)	1.279	(0.002)
prevention mask	0.999	3.607	(0.002)	3.607	(0.002)
prevention distancing	0.769	2.669	(0.003)	2.670	(0.003)
prevention hand washing	0.526	3.299	(0.002)	3.297	(0.002)
effect mask	0.641	2.983	(0.003)	2.981	(0.003)
effect hand washing	0.195	2.996	(0.003)	2.991	(0.003)
country management	0.082	1.831	(0.004)	1.822	(0.004)
community management	0.878	1.929	(0.003)	1.928	(0.003)
community action importance	0.498	3.354	(0.002)	3.352	(0.003)
community action norms	0.458	2.737	(0.003)	2.734	(0.003)
distancing importance	0.803	3.112	(0.003)	3.111	(0.003)
norms dist	0.221	49.008	(0.091)	49.162	(0.090)
norms masks	0.648	71.800	(0.086)	71.852	(0.086)
norms vaccine	0.837	61.939	(0.087)	61.917	(0.086)
risk community	0.164	2.541	(0.005)	2.531	(0.005)
risk infection	0.638	2.165	(0.005)	2.168	(0.005)
control infection	0.515	1.879	(0.006)	1.873	(0.006)
infection severity	0.083	1.272	(0.003)	1.264	(0.003)
employed 2020	0.542	0.725	(0.002)	0.727	(0.002)

Pre-treatment covariate means for all respondents who were eligible for treatment in both the treatment and control groups along with the p-value for the test of the null that the means are equal. For each covariate, only responses where the covariate is not missing and occurs before both treatment and control are included. To account for changes to the sampling frequencies, these p-values are from the coefficient on the intent-to-treat term in a regression of the covariate on treatment, period, and centered interactions between treatment and period. As we do not have weights for all respondents, this is an unweighted regression.

Table S4. Balance Tests: Final Dataset

	p-val	Control		Treated	
age	0.855	2.697	(0.003)	2.696	(0.003)
gender	0.757	1.442	(0.001)	1.441	(0.001)
education	0.446	2.826	(0.002)	2.826	(0.002)
own health	0.828	2.394	(0.002)	2.398	(0.002)
vaccine accept	0.710	1.510	(0.002)	1.510	(0.002)
knowledge existing treatments	0.417	0.213	(0.001)	0.211	(0.001)
info exposure past week	0.417	2.367	(0.002)	2.371	(0.002)
info exposure more less wanted	0.875	2.410	(0.002)	2.409	(0.002)
know positive case	0.029	1.329	(0.002)	1.325	(0.002)
prevention mask	0.181	3.640	(0.003)	3.643	(0.003)
prevention distancing	0.132	2.709	(0.004)	2.716	(0.004)
prevention hand washing	0.445	3.333	(0.003)	3.335	(0.003)
effect mask	0.155	2.996	(0.003)	2.990	(0.003)
effect hand washing	0.315	3.014	(0.003)	3.011	(0.003)
country management	0.537	1.796	(0.004)	1.782	(0.004)
community management	0.964	1.904	(0.004)	1.900	(0.004)
community action importance	0.747	3.371	(0.003)	3.369	(0.003)
community action norms	0.717	2.712	(0.004)	2.705	(0.004)
distancing importance	0.279	3.150	(0.003)	3.149	(0.003)
norms dist	0.028	49.517	(0.107)	49.797	(0.107)
norms masks	0.041	72.605	(0.101)	72.918	(0.101)
norms vaccine	0.871	62.591	(0.101)	62.572	(0.101)
risk community	0.163	2.564	(0.006)	2.544	(0.006)
risk infection	0.756	2.205	(0.006)	2.211	(0.006)
control infection	0.557	1.885	(0.007)	1.874	(0.007)
infection severity	0.011	1.269	(0.004)	1.257	(0.004)
employed 2020	0.415	0.725	(0.002)	0.728	(0.002)

Pre-treatment covariate means for all respondents who were eligible for treatment, completed the entire survey, and received a full survey completion weight in both the treatment and control groups along with the p-value for the test of the null that the means are equal. For each covariate, only responses where the covariate is not missing and occurs before both treatment and control are included. To account for changes to the sampling frequencies, these p-values are from the coefficient on the treatment term in a regression of the covariate on treatment, period, and centered interactions between treatment and period. This is a weighted regression using full completion survey weights.

Table S5. Balance Tests Between Treatments: Final Dataset

	VD p-val	VM p-val	Vaccine	Masks	Dist
age	0.543	0.011	2.716	2.684	2.611
gender	0.229	0.630	1.440	1.442	1.445
education	0.861	0.313	2.825	2.826	2.839
own health	0.278	0.629	2.399	2.401	2.386
vaccine accept	0.419	0.000	1.524	1.505	1.442
knowledge existing treatments	0.308	0.926	0.159	0.209	0.555
info exposure past week	0.294	0.633	2.375	2.369	2.354
info exposure more less wanted	0.339	0.361	2.420	2.404	2.355
know positive case	0.825	0.894	1.339	1.326	1.233
prevention mask	0.920	0.391	3.649	3.640	3.609
prevention distancing	0.479	0.551	2.723	2.711	2.687
prevention hand washing	0.500	0.134	3.339	3.331	3.326
effect mask	0.366	0.744	2.998	2.991	2.933
effect hand washing	0.897	0.232	3.016	3.003	3.013
country management	0.375	0.143	1.788	1.775	1.765
community management	0.544	0.013	1.912	1.885	1.880
community action importance	0.710	0.751	3.370	3.369	3.367
community action norms	0.503	0.695	2.711	2.704	2.679
distancing importance	0.882	0.841	3.150	3.148	3.149
norms dist	0.523	0.917	49.914	49.721	49.237
norms masks	0.693	0.447	73.143	72.896	71.308
norms vaccine	0.521	0.829	62.837	62.512	60.816
risk community	0.813	0.331	2.547	2.545	2.515
risk infection	0.460	0.771	2.218	2.208	2.179
control infection	0.242	0.498	1.880	1.872	1.849
infection severity	0.323	0.554	1.255	1.258	1.264
employed 2020	0.707	0.152	0.731	0.722	0.733

Pre-treatment covariate means for all respondents who were treated, completed the entire survey, and received a full survey completion weight along with the p-value for the test of the null that the means between treatment groups are equal. For each covariate, only responses where the covariate is not missing and occurs before both treatment and control are included. To account for changes to the sampling frequencies, these p-values are from the coefficient on the treatment behavior terms in a regression of the covariate on treatment behavior, period, and centered interactions between treatment behavior and period. This is a weighted regression using full completion survey weights.

S4. Analysis methods

The results presented in the main text and elaborated on in sections S5, S6, and S6.1 each use a similar pre-registered methodology that we briefly describe here. For the results in sections S5 and S6, we estimate the following linear regression for each behavior k

$$Y_{ik} = \delta_{0k} + \sum_{j \in J} \delta_{jk} D_{ik}^{j} + \gamma_k X_i + \sum_{j \in J} \eta_{jk} X_i D_{ik}^{j} + \varepsilon_{ik} \tag{S3}$$

where Y_{ik} is the outcome for individual i and behavior $k \in K = \{\text{vaccine, distancing, masks}\}$, D_{ik}^{j} is an indicator if individual i received treatment $j \in J = \{\text{Broad, Narrow}\}$ for behavior k, and X_i is a vector of centered covariates (50, 51). In the figures and tables, we report the δ_{jk}'s and suppress coefficients on covariates and interactions. All statistical inference uses heteroskedasticity-consistent Huber–White "sandwich" estimates of the variance–covariance matrix.

In section S6.1, we estimate a similar regression. As our analysis of heterogeneity focuses on the vaccine treatment, we will suppress the behavior index k. Here covariates are centered at their weighted group-specific means.

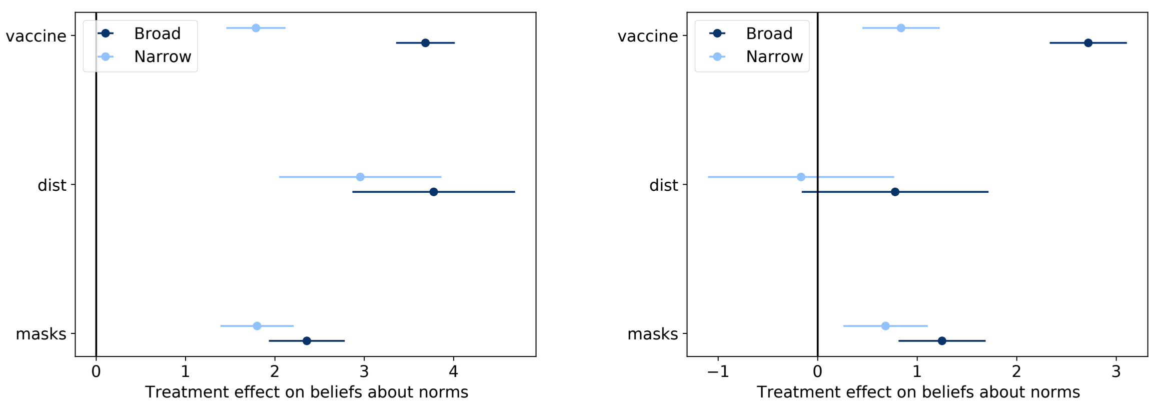
$$Y_i = \sum_{b \in B} 1[b_i = b] \left(\delta_0^b + \sum_{j \in J} \delta_j^b D_{ij}^b + \gamma_k X_i + \sum_{j \in J} \eta_j^b X_i D_{ij}^b \right) + \varepsilon_{ik} \tag{S4}$$

S4.1. Mixed-effects model. In the main text and Figure 3c, we report results from a linear mixed-effects model with coefficients that vary by country. This model is also described in our preregistered analysis plan. Note that the coefficients for the overall (across-country) treatments effects in this model differ slightly from the estimates from the model; that is, the "Average" points in Figure 3b and 3c do not match exactly. As noted in our analysis plan, "sandwich" standard errors are not readily available here, so reported 95% confidence intervals are obtained by estimating the standard errors via a bootstrap.

S5. Effects on beliefs about descriptive norms

Figure S5 present evidence that the treatments do update beliefs about the descriptive norms of survey respondents. The figures plot coefficients on treatment from a regression of survey norms on treatment status, including centered covariates and interactions as described in the pre-analysis plan. In this analysis, treated respondents are those who receive the treatment before the question eliciting beliefs about norms. This will not agree, in general, with the treatment status for the main analysis given the randomized question order in the survey. The covariates included in this analysis are pre-treatment and outcome relative to this treatment definition.

Figure S5a compares treatment and control respondents and Figure S5b conditions on treated individuals and then uses individuals who received an information treatment for a different behavior as control. The coefficients plotted in Figure S5b are smaller than in Figure S5a, which indicates that normative information on other behaviors may induce an update in beliefs on the focal behavior.



(a) Treatment vs. Control

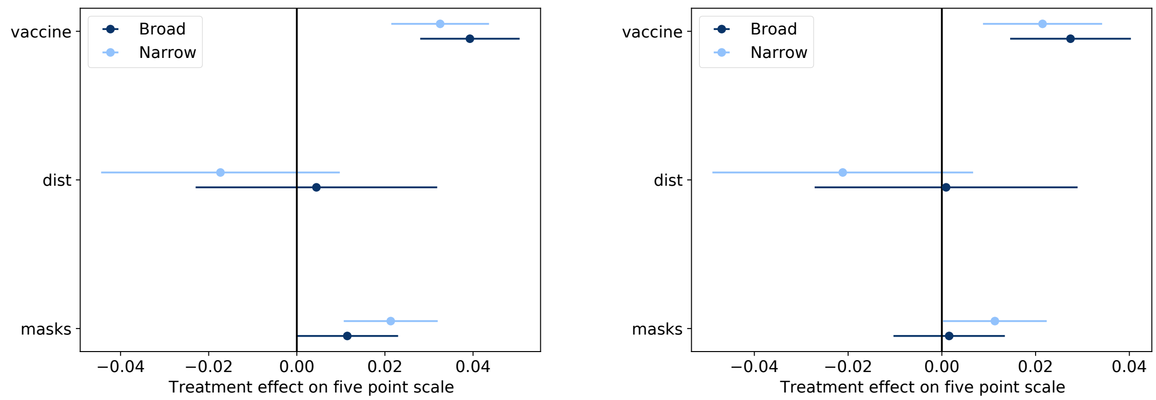
(b) Treatment vs. Other Behavior Treatment

Fig. S5. Effects on beliefs about descriptive norms

	(1)	(2)	(3)	(4)	(5)	(6)
Broad Treatment	1.250***	0.778	2.719***	2.357***	3.777***	3.685***
	(0.223)	(0.479)	(0.198)	(0.217)	(0.464)	(0.168)
Narrow Treatment	0.682***	-0.167	0.838***	1.801***	2.955***	1.788***
	(0.217)	(0.477)	(0.198)	(0.209)	(0.463)	(0.169)
Control: Other Treatment	X	X	X			
Behavior	masks	dist	vaccine	masks	dist	vaccine
Number Controls	149458	34002	91217	229715	52724	225615
Number Treated	75125	17354	130389	75125	17354	130389
Observations	224,583	51,356	221,606	304,840	70,078	356,004
R^2	0.170	0.145	0.206	0.182	0.157	0.210
Adjusted R^2	0.170	0.143	0.206	0.182	0.156	0.210
Residual Std. Error	25.040	27.046	24.162	25.445	27.085	24.566
F Statistic	130.581***	46.184***	191.547***	201.909***	64.198***	329.570***

*p<0.1; **p<0.05; ***p<0.01

Table S6. Effects on beliefs about descriptive norms, for primary and alternative definitions of the control group



(a) Treatment vs. Control

(b) Treatment vs. Other Behavior Treatment

Fig. S6. Treatment effects with primary and alternative definition of the control group

S6. Effects on intentions

Figure S6 displays regression coefficients for the primary analysis, where the intention to partake in the outcome behavior is regressed on treatment, centered covariates, and their interactions. As discussed in the pre-analysis plan, we use both the randomized timing of information treatments and the randomized focal behavior of the intervention. Figure S6a uses respondents who receive the information after the outcome is measured as the control group and Figure S6b uses individuals who receive the information treatment for a different behavior as the control group. The results are largely consistent and suggest that the information treatment significantly increases reported vaccine acceptance, while effects for distancing and masks are smaller and not statistically distinguishable from zero.

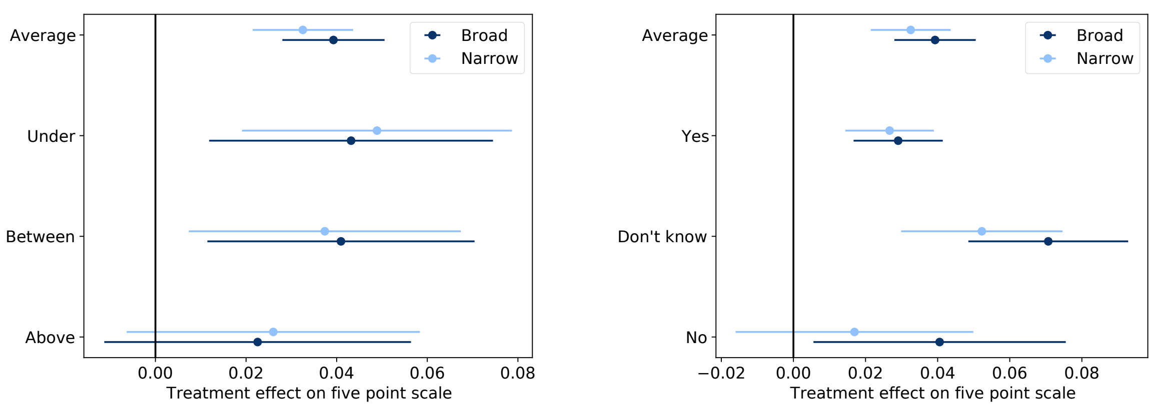
Table S8 presents results from the same analysis after transforming the outcome variable into binary indicators. This allows us to understand across which thresholds the treatment has induced people to cross. The coefficients indicate that the treatment is inducing people to report they will at least probably take the vaccine and definitely take the vaccine. Similar regressions restricted to those who report they don't know if they will take the vaccine at baseline are presented in Table S9. Among this group, there is a larger effect and it is concentrated in moving people to say they will "probably" take the vaccine.

S6.1. Heterogeneous treatment effects. Figure S7 plots regression coefficients for estimates of heterogeneous treatment effects in equation 4 across different dimensions. We see the positive effects of our treatment concentrated in those with lower baseline beliefs about norms (Figure S7a) and in those who are unsure if they will accept a vaccine (Figure S7b). Estimates are also reported in Tables S11 and S12.

	(1)	(2)	(3)	(4)	(5)	(6)
Broad Treatment	0.002	0.001	0.027***	0.011*	0.004	0.039***
	(0.006)	(0.014)	(0.007)	(0.006)	(0.014)	(0.006)
Narrow Treatment	0.011**	-0.021	0.021***	0.021***	-0.017	0.033***
	(0.006)	(0.014)	(0.006)	(0.005)	(0.014)	(0.006)
Control: Other Treatment	X	X	X			
Behavior	masks	dist	vaccine	masks	dist	vaccine
Number Controls	159962	42237	98940	242238	64323	233076
Number Treated	80847	21296	132517	80847	21296	132517
Observations	240,809	63,533	231,457	323,085	85,619	365,593
R^2	0.248	0.234	0.618	0.251	0.244	0.610
Adjusted R^2	0.248	0.233	0.617	0.251	0.243	0.610
Residual Std. Error	0.692	0.855	0.805	0.699	0.856	0.812
F Statistic	143.880***	84.485***	989.712***	204.798***	113.443***	1513.455***

*p<0.1; **p<0.05; ***p<0.01

Table S7. Treatment effects with primary and alternative definition of the control group



(a) Baseline Belief Partition

(b) Vaccine Acceptance

Fig. S7. Heterogeneous Treatment Effects

	$>$ No, definitely not	$>$ Probably not	$>$ Unsure	$>$ Probably
Intercept	0.916***	0.846***	0.673***	0.487***
	(0.001)	(0.001)	(0.001)	(0.001)
Narrow Treatment	0.000	0.005***	0.011***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
Broad Treatment	0.000	0.005***	0.016***	0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	365,593	365,593	365,593	365,593
R^2	0.296	0.493	0.560	0.459
Adjusted R^2	0.296	0.493	0.560	0.459
Residual Std. Error	0.232	0.257	0.310	0.368
F Statistic	152.334***	577.685***	1647.634***	1508.106***

*p$<$0.1; **p$<$0.05; ***p$<$0.01

Estimates of equation 3 with binary outcomes. The outcome variable for each column is an indicator equal to one if the respondent reported a value higher than the column name. For example, in the column "$>$ Probably not" the outcome Y_i equals one if the respondent answered "Unsure", "Probably", or "Yes, definitely".

Table S8. Distributional treatment effects

	$>$ No, definitely not	$>$ Probably not	$>$ Unsure	$>$ Probably
Intercept	0.968***	0.900***	0.295***	0.046***
	(0.001)	(0.002)	(0.003)	(0.002)
Narrow Treatment	0.003	0.005	0.025***	0.018***
	(0.003)	(0.004)	(0.007)	(0.004)
Broad Treatment	0.001	0.006	0.049***	0.015***
	(0.003)	(0.004)	(0.007)	(0.004)
Observations	69,497	69,497	69,497	69,497
R^2	0.111	0.079	0.091	0.063
Adjusted R^2	0.109	0.077	0.089	0.061
Residual Std. Error	0.167	0.290	0.448	0.218
F Statistic	6.337***	8.683***	24.288***	8.521***

*p$<$0.1; **p$<$0.05; ***p$<$0.01

Estimates of equation 3 with binary outcomes on sample of respondents who say they don't know if they will take a vaccine at baseline. The outcome variable for each column is an indicator equal to one if the respondent reported a value higher than the column name. For example, in the column "$>$ Probably not" the outcome Y_i equals one if the respondent answered "Unsure", "Probably", or "Yes, definitely".

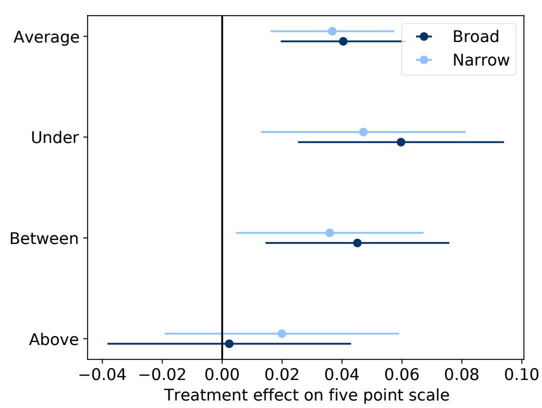
Table S9. Distributional treatment effects for "Don't know" respondents

	$>$ No, definitely not	$>$ Probably not	$>$ Unsure	$>$ Probably
Intercept	0.876***	0.772***	0.549***	0.358***
	(0.002)	(0.003)	(0.003)	(0.003)
Narrow Treatment	0.001	0.012**	0.019***	0.017***
	(0.005)	(0.005)	(0.006)	(0.007)
Broad Treatment	-0.001	0.005	0.019***	0.020***
	(0.005)	(0.005)	(0.006)	(0.006)
Observations	48,699	48,699	48,699	48,699
R^2	0.357	0.534	0.580	0.472
Adjusted R^2	0.355	0.532	0.578	0.471
Residual Std. Error	0.266	0.287	0.324	0.352
F Statistic	43.277***	176.441***	374.368***	175.635***

*p<0.1; **p<0.05; ***p<0.01

Estimates of equation 3 with binary outcomes on sample of respondents who have a baseline beliefs about how many people in their community will take a vaccine under the narrow treatment number. The outcome variable for each column is an indicator equal to one if the respondent reported a value higher than the column name. For example, in the column "$>$ Probably not" the outcome Y_i equals one if the respondent answered "Unsure", "Probably", or "Yes, definitely".

Table S10. Distributional treatment effects for "Under" respondents



Heterogeneous treatment effects based on baseline beliefs about how many people in their community will accept a vaccine. We remove respondents who say they believe 0, 50, or 100 percent of people in their community will accept a vaccine to mitigate measurement error due to a bias towards round numbers.

Fig. S8. Mitigating Round Number Bias: Baseline Belief Partition

	Average	Above	Between	Under
Broad Treatment	0.039***	0.023	0.041***	0.043***
	(0.006)	(0.017)	(0.015)	(0.016)
Narrow Treatment	0.033***	0.026	0.037**	0.049***
	(0.006)	(0.017)	(0.015)	(0.015)
Observations	365,593	30,731	34,008	48,699
R^2	0.610	0.394	0.625	0.649
Adjusted R^2	0.610	0.391	0.623	0.648
Residual Std. Error	0.812	0.799	0.692	0.826
F Statistic	1513.455***	43.505***	166.301***	329.346***

*p<0.1; **p<0.05; ***p<0.01

The joint test that the broad and narrow coefficients are equal across groups has a p-value of 0.807, and the two-sided test that the broad (narrow) treatment effects in the Under and Above groups are equal has a p-value of 0.38 (0.31).

Table S11. Heterogeneous Treatment Effects: Baseline Beliefs

	Average	No	Don't Know	Yes
Broad Treatment	0.039***	0.041**	0.071***	0.029***
	(0.006)	(0.018)	(0.011)	(0.006)
Narrow Treatment	0.033***	0.017	0.052***	0.027***
	(0.006)	(0.017)	(0.011)	(0.006)
Observations	365,593	53,119	69,497	239,822
R^2	0.610	0.211	0.114	0.119
Adjusted R^2	0.610	0.209	0.112	0.118
Residual Std. Error	0.812	0.994	0.743	0.725
F Statistic	1513.455***	50.166***	19.873***	43.186***

*p<0.1; **p<0.05; ***p<0.01

The joint test that the broad and narrow coefficients are equal across groups has a p-value of 0.012, and the two-sided test that the broad (narrow) treatment effects in the Yes and Don't know groups are equal has a p-value of <0.01 (0.05). The two sided test that the broad (narrow) treatment effects in the Don't know and No groups are equal has a p-value of 0.15 (0.08).

Table S12. Heterogeneous Treatment Effects: Baseline Vaccine Acceptance

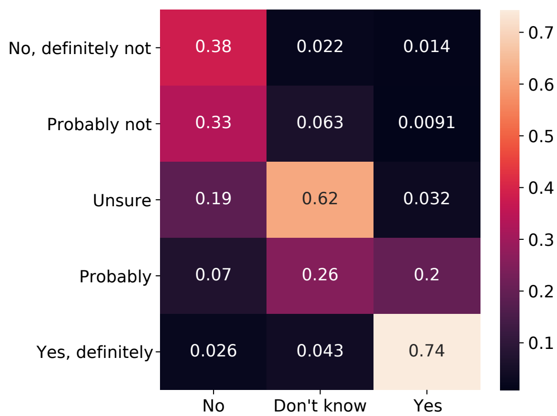


Fig. S9. Correlation of Baseline Vaccine Acceptance and Outcome (Detailed) Vaccine Acceptance

Heatmap showing relationship between baseline vaccine acceptance question (x-axis) and the outcome vaccine acceptance question (y-axis) for the control users. Each cell shows the probability of an outcome response conditional on the baseline response and each column sums to one.

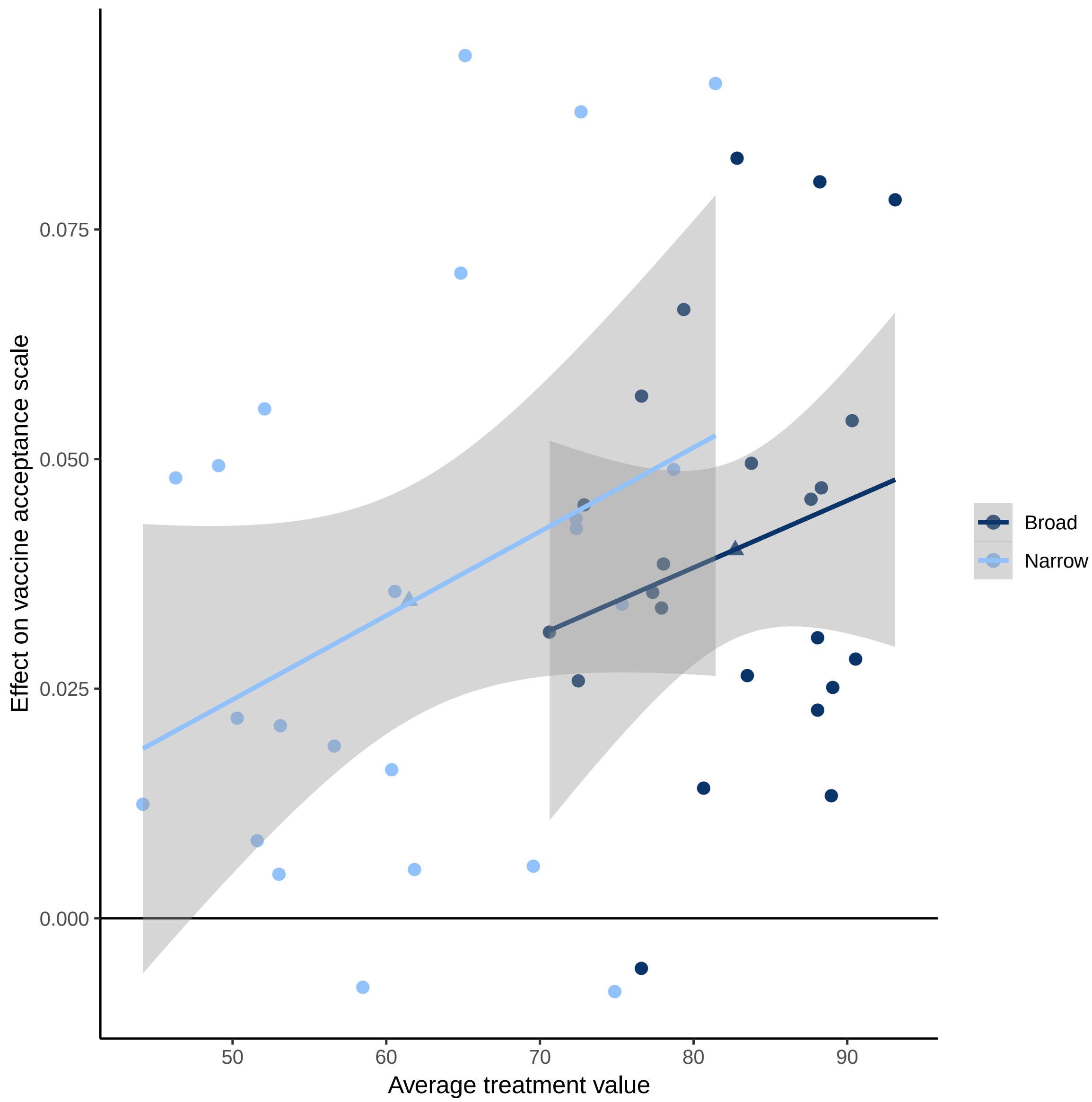
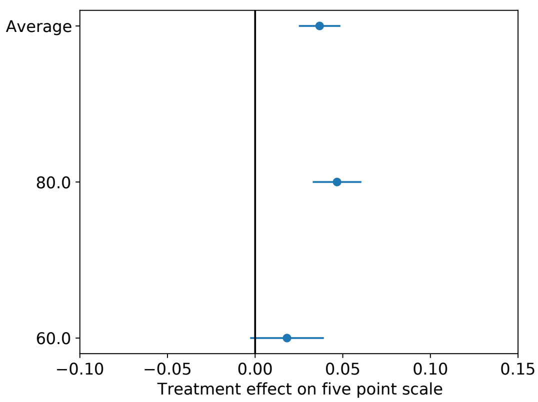
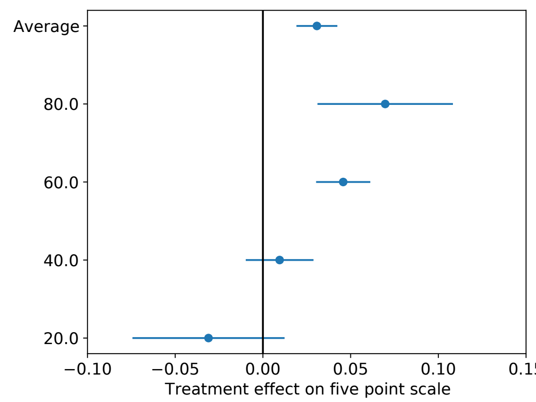


Fig. S10. Explaining Country-Level Heterogeneity with Variation in Treatment Values

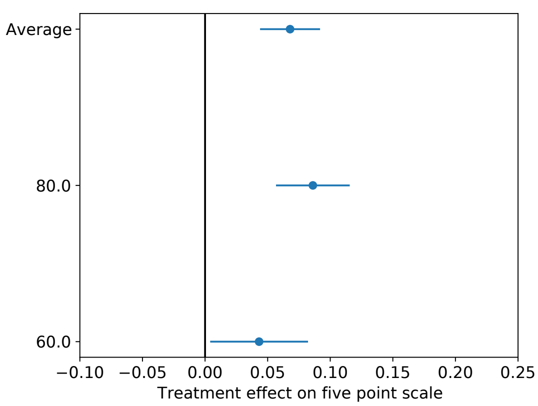
A scatter plot of country-level treatment effect estimates and the average normative information treatment shown over the course of the experiment. Each point represents a country-level treatment effect while the triangle represents the grand-mean of treatment effects and the average information shown across all countries in the experiment (weighted by the number of responses).



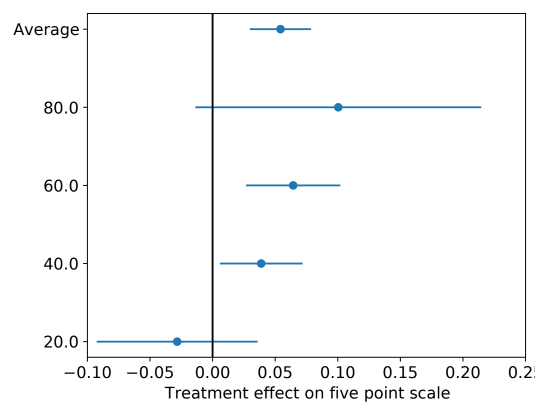
(a) Full Sample: Broad Treatment



(b) Full Sample: Narrow Treatment



(c) Baseline Vaccine Unsure: Broad Treatment



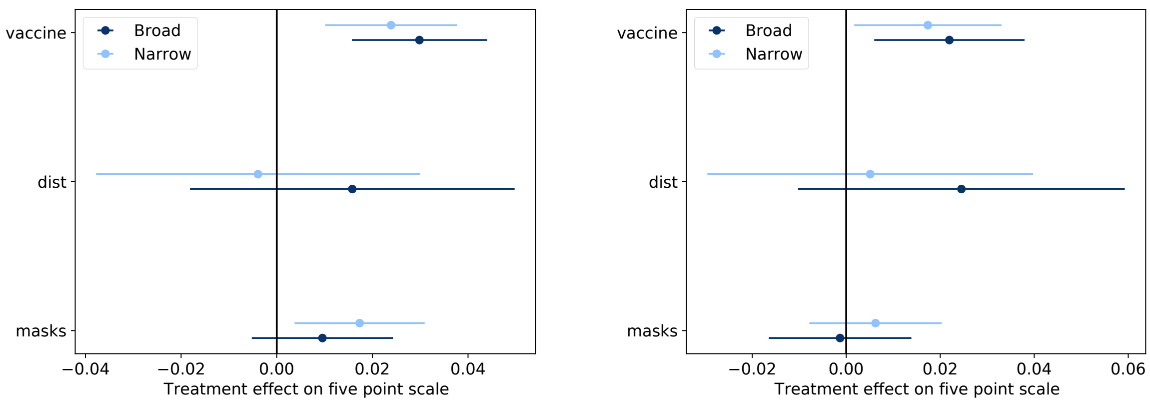
(d) Baseline Vaccine Unsure: Narrow Treatment

Fig. S11. Treatment Effect by Treatment Number

The coefficients reported in these figures are from a regression of vaccine acceptance measured on a five point scale on indicators of the treatment number the individual was shown (if treated) grouped into bins of width 20 percentage points. We also include covariates and their interactions with treatment as in the main analysis. Figures S11a and S11b report the coefficients from the regression on the full sample of individuals of separate regressions for the broad and narrow treatments. Figures S11c and S11d report the coefficients from the same regression on the subset of individuals who reported "Don't know" to the baseline vaccine acceptance question.

S6.2. Robustness checks. One concern with survey experiments is that results could reflect researcher demand effects, where subjects respond how they think the researchers would want them to respond. While we cannot rule this out completely, we do not believe this is driving our results(52, 53). We may be less worried about researcher demand effects in this survey as it has a more general advertised purpose and it covers several topics, so normative information is not particularly prominent (Figure S1). Furthermore, unlike other sampling frames with many sophisticated study participants (e.g., Amazon Mechanical Turk), respondents are recruited from a broader population (Facebook users). In addition, we observe null results for observable behaviors such as distancing and mask wearing, which we would expect to suffer from the same researcher demand effects as vaccine acceptance. We also compare the outcome of subjects who receive the vaccine norm treatment to those receiving the treatment providing information about masks and distancing and find the treatment effect persists. Moreover, we may expect researcher demand effects to be smaller when the information treatment and the outcome are not immediately adjacent. In all cases, for the vaccine acceptance outcome, there is always at least one intervening screen of questions (the future mask-wearing and distancing intentions questions). Furthermore, they are often separated by more than this. We consider a subset of respondents where the treatment and the outcome are separated by at least one “block” of questions between them. Results of this analysis are presented in Figure S12 and Table S13. The treatment effect estimates on this smaller sample are less precise, but both positive. For vaccines, the treatment effects muted somewhat as the p-values that the treatment effect is equal across this smaller sample and the broader sample are 0.02 and 0.03 for the broad and narrow treatments, respectively. Moreover, Table S14 shows even with the larger gap between treatment and outcome the information is still moving a relatively large share of people who are unsure or more negative to at least probably accepting the vaccine.

Figure S13 plots the distribution of the number of screens between treated and control. In Figure S13a, we plot the distribution for the entire sample and in Figure S13b we plot the distribution for the subset of those with at least one block between treatment and control. For this group there are at least three pages between the treatment and outcome



(a) Treatment vs. Control

(b) Treatment vs. Other Behavior Treatment

Fig. S12. Robustness to Researcher Demand Effects

Main analysis conducted on the subset of individuals with at least one “block” of questions separating treatment and outcome.

	(1)	(2)	(3)	(4)	(5)	(6)
Broad Treatment	-0.001	0.025	0.022***	0.010	0.016	0.030***
	(0.008)	(0.018)	(0.008)	(0.008)	(0.017)	(0.007)
Narrow Treatment	0.006	0.005	0.017**	0.017**	-0.004	0.024***
	(0.007)	(0.018)	(0.008)	(0.007)	(0.017)	(0.007)
Control: Other Treatment	X	X	X			
Behavior	masks	dist	vaccine	masks	dist	vaccine
Number Controls	105973	27441	65464	160843	41734	154676
Number Treated	53773	13784	88139	53773	13784	88139
Observations	159,746	41,225	153,603	214,616	55,518	242,815
R^2	0.222	0.209	0.616	0.226	0.218	0.605
Adjusted R^2	0.221	0.207	0.615	0.226	0.216	0.605
Residual Std. Error	0.707	0.865	0.811	0.715	0.870	0.819
F Statistic	89.371***	47.748***	652.743***	123.872***	60.798***	980.206***

*p<0.1; **p<0.05; ***p<0.01

Estimates of equation 3 on the restricted sample when outcome and treatment are separated by at least one additional block of questions.

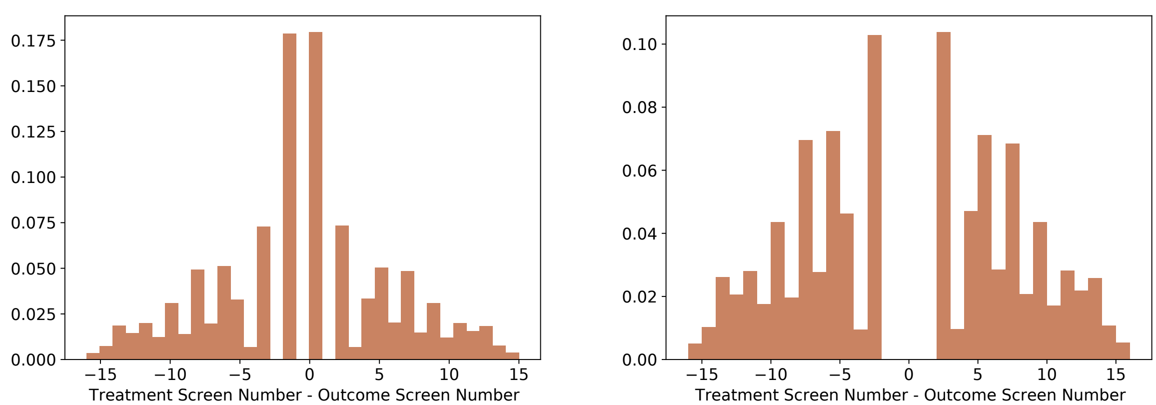
Table S13. Robustness to Greater Separation of Treatment and Outcome

	> No, definitely not	> Probably not	> Unsure	> Probably
Intercept	0.916***	0.846***	0.672***	0.486***
	(0.001)	(0.001)	(0.001)	(0.002)
Narrow Treatment	-0.000	0.002	0.007**	0.015***
	(0.002)	(0.002)	(0.003)	(0.003)
Broad Treatment	-0.000	0.002	0.013***	0.015***
	(0.002)	(0.002)	(0.003)	(0.003)
Observations	242,815	242,815	242,815	242,815
R^2	0.291	0.490	0.559	0.457
Adjusted R^2	0.291	0.490	0.559	0.457
Residual Std. Error	0.234	0.258	0.312	0.369
F Statistic	100.054***	368.874***	1071.256***	998.540***

*p<0.1; **p<0.05; ***p<0.01

Estimates of equation 3 on the restricted sample when outcome and treatment are separated by at least one additional block of questions. The outcome variable in this analysis are binary indicators if the outcome was at least a certain response as in table S8.

Table S14. Robustness to Greater Separation of Treatment and Outcome: Distributional Treatment Effects



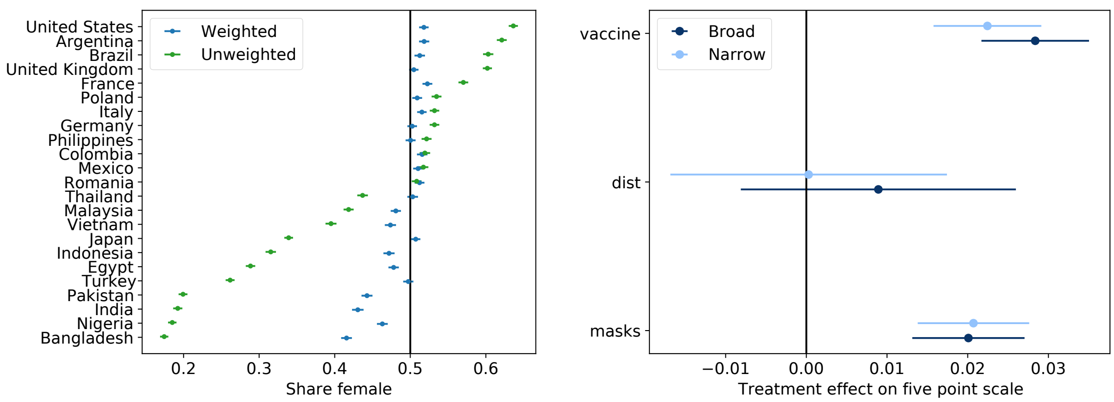
(a) Number of Screens Between Treatment and Outcome

(b) Number of Screens Between Treatment and Outcome in Robustness Check Sample

Fig. S13. Separation of Treatment and Outcome

(a) Histogram of the number of screens between treatment and outcome. Negative numbers represent treated respondents and positive numbers are control respondents. The distribution is not smooth as the randomized order is at the block level, and blocks have varying number of screens (pages) within them. (b) The same histogram, but for the set of respondents with at least one block between treatment and outcome.

S6.3. Unweighted estimates. All analyses presented in the paper take advantage of survey weights that adjust the survey for sampling and non-response bias (33). This is to make the analysis as representative as possible for the countries we survey. To motivate the use of weights, consider Figure S14a which plots the estimated share of females in the population. The unweighted estimates in green contain substantial bias, and the weighted estimators largely (although not completely) reduce this bias. Formally, non-response weighting assumes data are missing at random (conditional on covariates used for weighting, respondents are a random sample of those sampled) (54). While this is a strong assumption, we find it more plausible than the assumption required for an unweighted analysis that assumes the sample is a random sample from the target population, which we can confidently reject (Figure S14a). As a robustness check, however, we run the analysis using unweighted estimators and find the treatment effects are robust to the use of weights (Figure S14b).



(a) Weighted and unweighted female share

(b) Main effects without weights

Fig. S14. Survey Weights

(a) Weighted and unweighted estimates of the share of females. We expect this to be roughly 0.5, and weights greatly reduce the bias in the unweighted estimates.(32, 33) (b) Main treatment effects using unweighted estimators.

S7. Norm–intention correlations

In the main text, Figure 1 (inset) shows the association between beliefs about descriptive norms and intentions to accept a COVID-19 vaccine. Figure S15 disaggregates this information by country. As in the main text, this is a purely observational association but is computed on the main experimental sample (i.e., starting in late October).

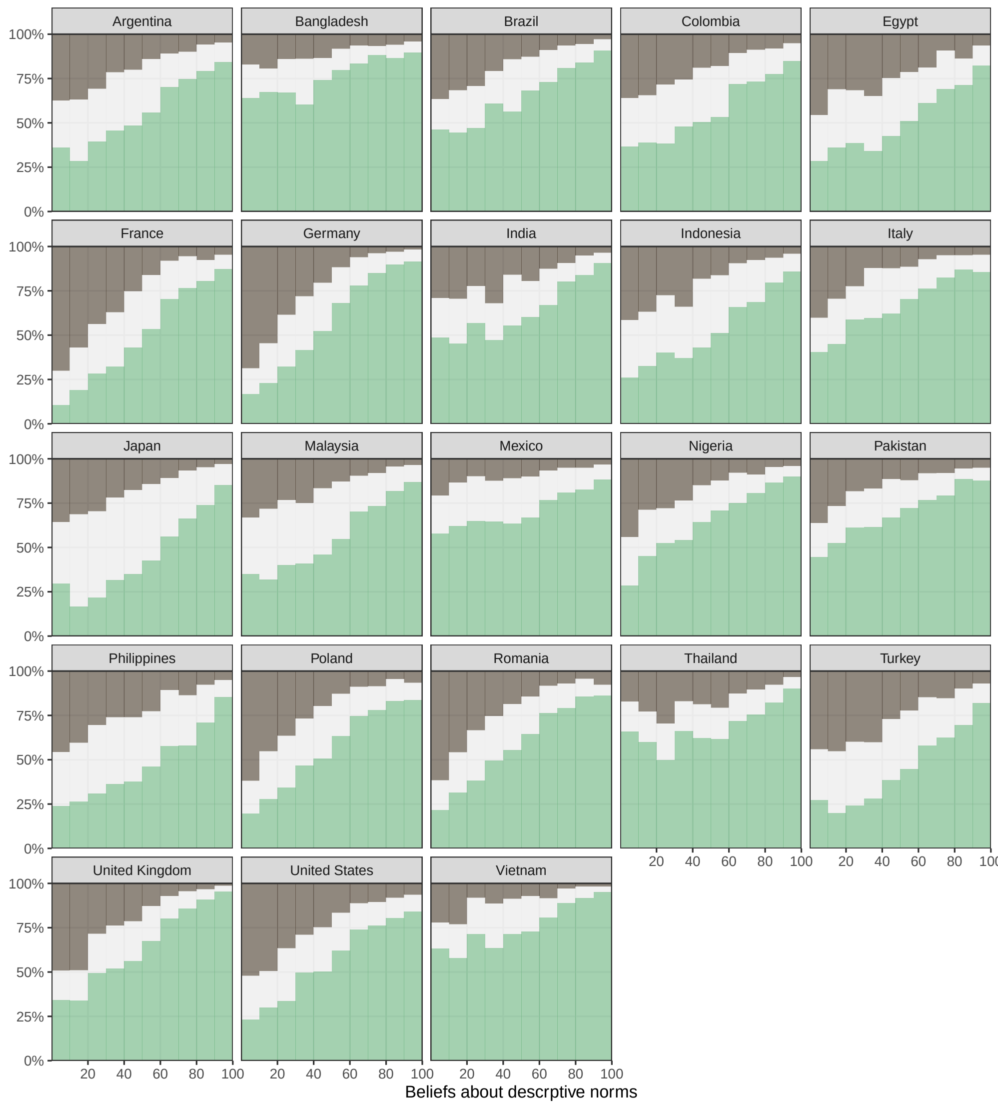


Fig. S15. People who believe a larger fraction of their community will accept a vaccine are on average more likely to say they will accept a vaccine and this is true within the 23 included countries. The vertical axis shows the percentage of respondents who replied Yes (green), Don't know (gray) and No (brown) to whether they will accept a COVID-19 vaccine.

S8. Country-level validation of survey data

At the time of this writing, vaccinations are not readily available to the vast majority of people in the sample and so it is difficult to compare intentions with actual take-up due to supply constraints. In an attempt to quantify the correspondence between our survey responses and other measures of corresponding behavior, we run an auxiliary analysis comparing self-reported receipt of a COVID-19 vaccine from the survey with country-level uptake (COVID-19 vaccine data retrieved from https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations). The estimated vaccinated share of adults from the survey measure is highly predictive of the cross-country variation in actual vaccination shares, explaining over 80% of the cross-sectional variation (Table S15).

Intercept	-1.565
	(1.379)
Survey Share Vacc	0.794***
	(0.158)
Observations	22
R^2	0.829
Adjusted R^2	0.820
Residual Std. Error	4.488
F Statistic	25.183***

*p<0.1; **p<0.05; ***p<0.01

Regression of true vaccination share on estimated share vaccinated based on the survey measure. There are only 22 observations rather than 23 because data on the true share vaccinated are unavailable in Egypt. Adjustment for attenuation bias due to measurement error in the survey results in nearly identical results.

Table S15. Survey Data Predictive of Cross-Sectional Vaccine Take-up

S9. Intention to behavior correlation

A limitation of this experiment is that we measure self-reported vaccination intentions rather than eventual vaccination. When the experiment was first fielded, vaccinations were not available to the public and this remains true for many countries studied throughout the experiment making it difficult or not possible to measure the effect on actual COVID-19 vaccine uptake. To provide some evidence that survey intentions are predictive of vaccination behavior we conducted a supplemental survey in two waves in the United States, where vaccines have become widely available. First, from April 2, 2021 to May 1, 2021 we asked an online panel in the United States from CloudResearch if they had been vaccinated and their vaccination intentions. We then followed up from May 18, 2021 to June 1, 2021 to ask those who had not been vaccinated at baseline the same question. There were 1,350 respondents who completed both the baseline and endline survey. We then predict endline vaccination status with baseline vaccination intentions and this is plotted inf Figure S16. Our vaccination intentions measure is quite predictive of future vaccination status, with over half of those responding “Yes, definitely” having received at least one dose of a vaccine two months later. We also ask for the approximate date of when they received their vaccine and plot the distribution of acceptance over time in Figure S17, and it is clear that those with stronger intentions to receive a vaccine not only receive the vaccine at higher rates but also do so more quickly.

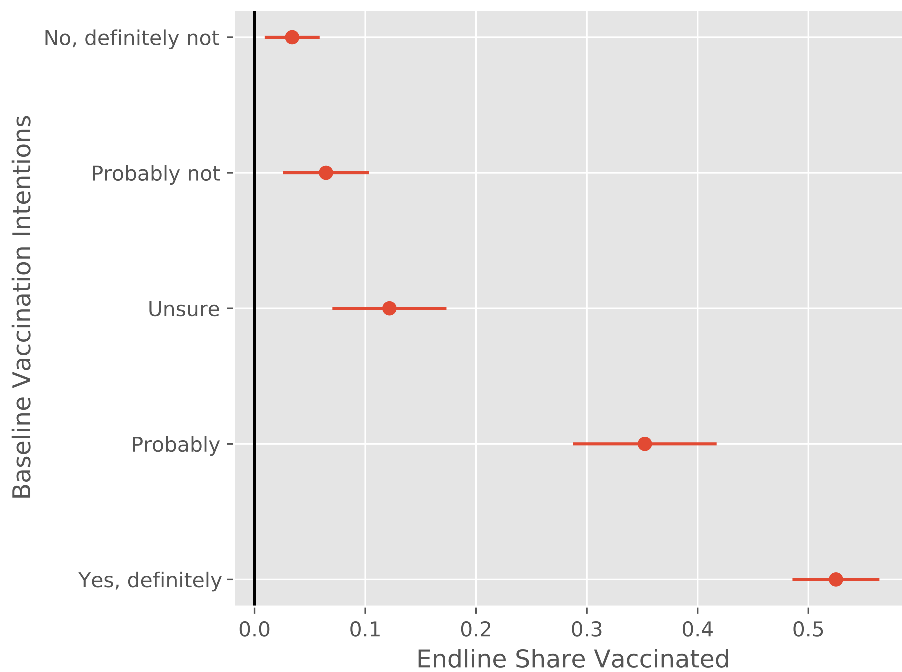


Fig. S16. Vaccination Intentions Predict Future Behavior

Coefficients from regression of endline self-reported vaccination status on baseline vaccination intentions.

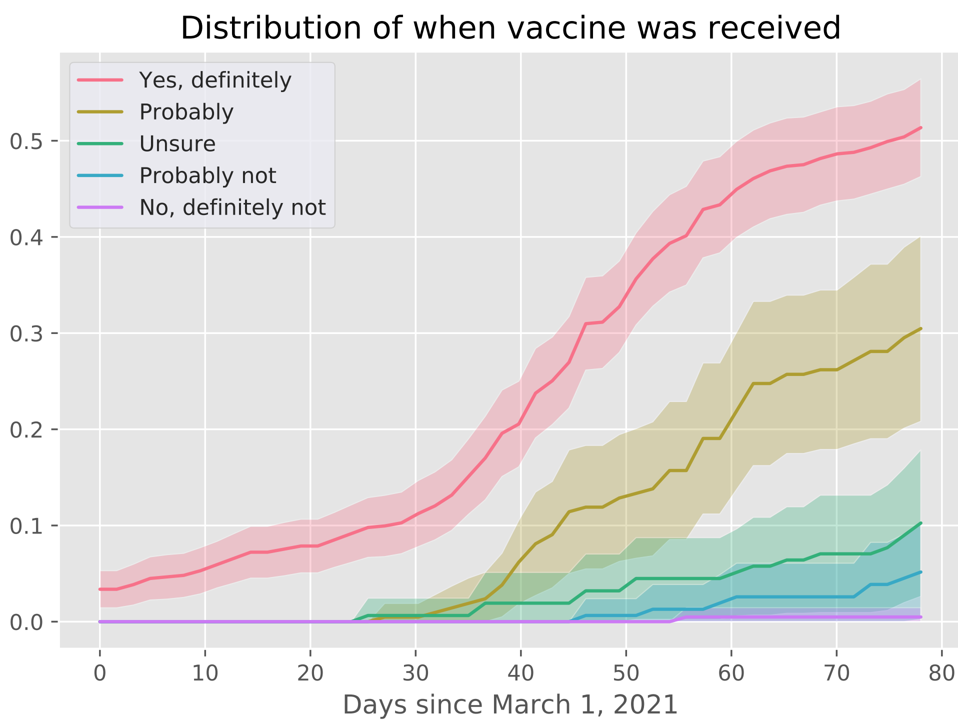


Fig. S17. Vaccination Intentions and Take-up Over Time

Distribution regression of the date that someone received their COVID-19 vaccine by baseline intention group. Those who are unvaccinated at endline are coded as 1,000 so the line plots the share who have received at least one dose over time. There is some mismeasurement, as some respondents reported not having received a vaccine in the baseline survey (April), while saying they received their first dose in March during the endline survey. We plot bootstrapped 95% confidence intervals.