Correcting Misperceptions about Support for Social Distancing to Combat COVID-19

JAMES ALLEN IV
International Food Policy Research Institute

ARLETE MAHUMANE
National Institute of Health, Mozambique

JAMES RIDDELL IV
University of Michigan Medical School

TANYA ROSENBLAT and DEAN YANG
University of Michigan

HANG YU
Peking University

I. Introduction

Attitudes toward social distancing changed rapidly during the COVID-19 pandemic (Janzwood 2020). During this rapid change, people often underestimated support for social distancing in their communities. Early in the pandemic, 98% of our Mozambican sample thought that people should be social distancing but estimated that only 69% of others in the community felt similarly. This gap motivates a public health policy: simply inform people of high rates of community support for social distancing. What effect would such messaging have on social-distancing behavior?

Faustino Lessitala provided top-notch leadership and field management. Patricia Freitag, Ryan McWay, and Maggie Barnard provided excellent research assistance. Julie Esch, Laura Kaminski, and Lauren Tingwall's grant management was world class. We appreciate feedback from participants in Michigan's Health, History, Development, and Demography Seminar and the Special Online Conference on Experimental Insights from Behavioral Economics on COVID-19 organized by Johns Hopkins University and the London School of Economics. This work is supported by the Abdul Latif Jameel Poverty Action Lab Innovation in Government Initiative through a grant from the Effective Altruism Global Health and Development Fund (grant no. IGI-1366), a UK Foreign, Commonwealth and Development Office award through the Innovations for Poverty Action's Peace and Recovery Program (grant no. MIT0019-X9), the Michigan Institute for Teaching and Research in Economics Ulmer Fund (grant

Electronically published September 9, 2024

Economic Development and Cultural Change, volume 73, number 1, October 2024.

https://doi.org/10.1086/727192

In theory, the effect of such a "misperceptions-correction" intervention on social distancing is ambiguous: on the one hand, informing people that more neighbors support social distancing than expected encourages free riding and lowers the perceived benefits from social distancing. On the other hand, people should revise their belief about the seriousness of COVID-19 upward in order to rationalize the observed number of infections in their neighborhood despite the higher than expected social-distancing support. This perceived-infectiousness effect increases the perceived benefits from social distancing and dominates free riding in communities with high levels of infections.[1] Finally, the norm-adherence effect should induce people to follow whatever local social norm is set by their neighbors—in our case, this effect should always increase social distancing.

We implemented a randomized controlled trial to test the effect of informing people about high local support for social distancing. The treatment either revised beliefs upward or confirmed beliefs about high rates of support for social distancing. Abiding by COVID-19 protocols, we conducted all treatments and surveys by phone among 2,117 Mozambican households.

Our outcome variable is the extent to which a household engages in social distancing. Measurement of this behavior is challenging because of experimenter demand effects.[2] Yet most prior studies ask for self-reports about general social-distancing compliance. When we do so, 95% claim to observe government social-distancing recommendations. We therefore construct a novel measure of social distancing. First, we ask respondents to self-report several social-distancing actions. Second, we ask others in the community to report on the respondent's social distancing. We are aware of no prior study that makes use of reports of others on a respondent's social-distancing behavior. Incorporating

no. G024289), and the National Institute on Aging of the National Institutes of Health (award no. T32AG000221). Our protocols were reviewed and approved by institutional review boards (IRBs) at the University of Michigan (Health Sciences and Social and Behavioral Sciences IRB, approval no. HUM00113011) and the Mozambique Ministry of Health National Committee on Bioethics for Health (CNBS; reference no. 302/CNBS/20). The study was submitted to the American Economic Association RCT Registry on May 26, 2020 (registration ID no. AEARCTR-0005862: 10.1257/rct.5862). Replication data, code, and codebooks are provided through Dataverse at https://doi.org/10.7910/DVN/TMARZT. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the aforementioned institutions. Contact the corresponding author, James Allen IV, at j.allen@cgiar.org.

[1] Our model is related to the literature on decision-making under misspecified subjective models (Spiegler 2020). Agents hold incorrect assumptions on one model parameter (e.g., share of population social distancing), leading them to incorrect conclusions about other parameters (e.g., disease infectiousness).

[2] Jakubowski et al. (2021) find that self-reported mask wearing is overstated relative to measures based on observations of others.

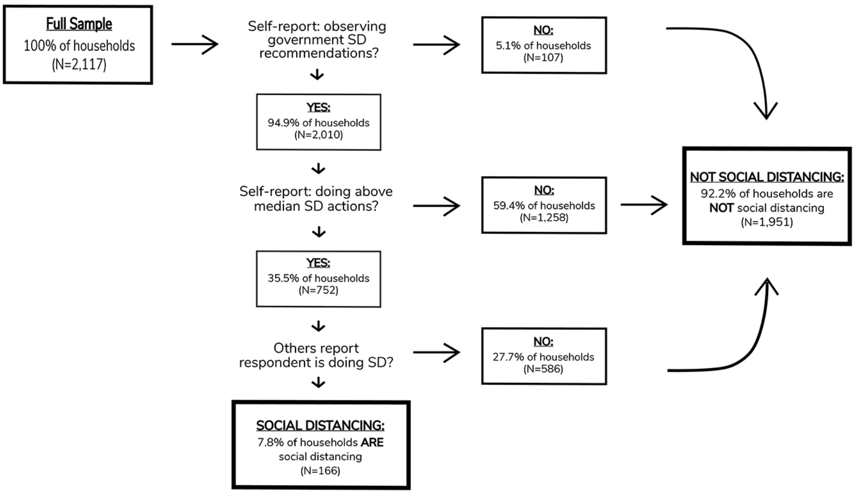

Figure 1. The social-distancing measure. Shown is the breakdown of the social-distancing measure at baseline. As prespecified, respondents were considered to be social distancing (SD) if they (1) self-reported "yes" to "In the past 14 days, have you observed the government's recommendations on social distancing?"; (2) self-reported doing more than the sample median number of "social-distancing actions" in the past 7 days; and (3) were considered to be social distancing according to leaders and other respondents in the community. Percentages reported are all shares of full sample ($N = 2{,}117$). See table 1 and section III.C of the main text for definitions of social-distancing questions.

self-reported actions and others' reports drops social distancing to a more discerning 8% (see fig. 1 and sec. III.C). Improved measurement leads the social-distancing rate to fall by an order of magnitude.

The average effect of the misperceptions-correction treatment in the full sample is small and not statistically significantly different from zero. However, as theory predicts, there is substantial treatment-effect heterogeneity: the treatment effect is statistically significantly more positive when local COVID-19 cases (per 100,000 population) are higher. In districts with few cases, the treatment effect is negative. In the district with the most COVID-19 cases, the treatment increases social distancing by 9.2 percentage points (statistically significant at the 5% level), a 70% increase over that district's control-group mean.

This pattern is consistent with the theoretical prediction that as infection rates rise, the perceived-infectiousness effect should increasingly dominate the free-riding effect of the misperceptions-correction treatment, leading the treatment effect to become more positive. We also test a further implication of the model: expectations of future infection rates should show similar treatment-effect heterogeneity. Empirical analyses confirm this prediction and provide additional support for the theoretical model.

Alongside the social-norm-correction treatment, we also randomly assigned a "leader-endorsement" treatment (an endorsement of social distancing by a community leader). The leader-endorsement treatment has a very small effect on social distancing that is not statistically significantly different from zero. We also find no treatment-effect heterogeneity for this treatment with respect to COVID-19 cases.

This paper contributes to understanding the effect of providing information about others' beliefs and attitudes (Benabou and Tirole 2011; Bicchieri and Dimant 2022). In health settings, Yu (2020) and Yang et al. (2021) find (in an overlapping Mozambican sample) that correction of overestimates of stigmatizing attitudes promoted HIV testing, though Banerjee, La Ferrara, and Orozco-Olvera (2019) find that informing Nigerian young adults of peers' attitudes on healthy sexual relationships did not change respondents' own attitudes.[3] Regarding social distancing, Martinez et al. (2021) show that respondents are influenced by others' social-distancing actions in hypothetical vignettes; however, no prior study has tested the effect of providing information about community support of social distancing on respondent behavior.

Our emphasis on interactions between free-riding and perceived-infectiousness effects is novel, but each effect has been studied separately. Free riding has been studied in the context of vaccination decisions (Hershey et al. 1994; Lau et al. 2019) and social distancing (Cato et al. 2020) and in similar Mozambican settings (Fafchamps, Vaz, and Vicente 2020). Perceived COVID-19 infection risk (e.g., due to vaccine anticipation; Andersson et al. 2021) has been shown to lower social-distancing intentions.

II. Theory

Our model focuses on the interaction between the free-riding and perceived-infectiousness effects for communities with low and high overall infection rates. We view norm adherence as a uniform effect that should always increase social distancing.

We consider a community where people have random pairwise meetings. People believe that a share x of the population supports social distancing and that the probability of becoming infected from unprotected meetings is α. People treat x as given but infer the infectiousness α from the current infection rate R in the community, which they can observe (we describe this inference below). The true infectiousness of the disease is $\hat{\alpha}$.

Importantly, people in the community have miscalibrated beliefs: the true share of the population supporting social distancing is $\hat{x}$ (we are interested in

[3] In other contexts, correcting misperceptions of community support or approval (i.e., the injunctive norm) has also been shown to change energy consumption (Schultz et al. 2007), female labor force participation (Bursztyn, González, and Yanagizawa-Drott 2020), donations to charities addressing climate change (Falk et al. 2021), and recycling-program participation (Fuhrmann-Riebel et al. 2023).

the case $\hat{x} > x$). People infer the true infectiousness $\hat{\alpha}$ of the disease only if they are correctly calibrated ($\hat{x} = x$).

Individual effort. A supporter engages in preventative effort e and assumes that other supporters choose effort e^* (in equilibrium we have $e = e^*$). Nonsupporters choose effort $e = 0$.

When someone supporting social distancing meets another person, she escapes exposure with probability

$$A(e, e_{\text{other}}) = \sqrt{e + e_{\text{other}}} = \begin{cases} \sqrt{e + e^*} & \text{if other person is supporter,} \\ \sqrt{e} & \text{if other person is nonsupporter.} \end{cases} \tag{1}$$

Hence, the marginal benefit of effort decreases both with own effort e and with the other person's effort e^*.[4]

The expected probability of escaping exposure is therefore

$$\bar{A}(e, e^*) = (1 - x)\sqrt{e} + x\sqrt{e + e^*}. \tag{2}$$

An agent becomes exposed with probability $1 - \bar{A}(e, e_{\text{other}})$. If exposed, she gets infected with probability α and suffers disutility $-C$ from infection.[5] If she is not exposed, then she does not get infected. Her baseline utility from no infection equals $\bar{U}$. The cost of preventative effort is e. Hence, her total utility equals

$$\bar{U} - \alpha(1 - \bar{A}(e, e_{\text{other}}))C - e. \tag{3}$$

The agent chooses e to maximize her utility, giving us the following first-order condition:

$$\frac{\alpha C}{2\sqrt{e}}\left[1 - x\left(1 - \frac{1}{\sqrt{1 + (e^*/e)}}\right)\right] = 1. \tag{4}$$

In equilibrium it has to be the case that the population effort e^* equals e. Hence, we can characterize equilibrium effort as

[4] We assume the other person's effort is unobservable. This is consistent with our finding that respondents underestimate the extent of social distancing.

[5] For simplicity, we assume that infectiousness does not vary with the agent's type (supporter or nonsupporter). Otherwise, we would need to keep track of two levels of infectiousness. The qualitative results would not change.

$$e = \left(\frac{\alpha C}{2}\left[1 - x\left(1 - \frac{1}{\sqrt{2}}\right)\right]\right)^2. \tag{5}$$

This demonstrates the free-riding effect: increasing the share x of supporters decreases effort because the marginal utility from own effort decreases. Also, effort increases if the disease is more infectious (higher α) and if illness is costlier (higher C).

Infection rate. People observe the current infection rate in the community. Infections come from two sources: nonsupporters become sick at rate $\alpha(1 - x\sqrt{e})$ while supporters become sick at rate $\alpha(1 - \bar{A}(e, e))$. Hence, people in the community assume that the current infection rate is generated by the following process:

$$\begin{aligned} R &= \alpha\left[\underbrace{(1 - x)(1 - x\sqrt{e})}_{\text{nonsupporters}} + \underbrace{x(1 - \sqrt{e}(1 + (\sqrt{2} - 1)x))}_{\text{supporters}}\right] \\ &= \alpha\left[1 - \sqrt{e}\underbrace{2x\left(1 - x\left(1 - \frac{1}{\sqrt{2}}\right)\right)}_{=G(x)}\right]. \end{aligned} \tag{6}$$

However, the true process determining current infections is actually

$$R = \hat{\alpha}\left[1 - \sqrt{e}G(\hat{x})\right]. \tag{7}$$

In other words, the true infection process is driven by the same social-distancing effort of supporters but different infectiousness $\hat{\alpha}$ and different $\hat{x}$.

A. Basic Equilibrium

Supporters initially assume that the disease has low infectiousness, and they adjust their estimate of α upward until the current infection rate R stabilizes.

> PROPOSITION 1. In equilibrium, effort level e, the current infection rate R, and the assumed infectiousness α satisfy equations (5), (6), and (7). Moreover, $\hat{\alpha} > \alpha$ if $\hat{x} > x$.

In equilibrium, both the assumed infection process (eq. [6]) and the real infection rate (eq. [7]) must produce observed infection rate R. For the second part, note that $G(x)$ is increasing in $x \in [0, 1]$; hence, $\hat{x} > x$ implies $\hat{\alpha} > \alpha$ to generate the same infection rate R.

B. *Treatment Effect*

We now consider the effect of our treatment informing people that the population share supporting social distancing is really $\hat{x} > x$.

Proposition 1 implies that if supporters are informed that the true population share supporting social distancing is $\hat{x} > x$, they must infer higher disease infectiousness than they initially assumed (because their estimated disease infectiousness immediately jumps from α to true $\hat{\alpha}$). This is the perceived-infectiousness effect.

Supporters of social distancing will adjust their effort level to a new level $\hat{e}$, but there are two countervailing effects:

1. Holding assumed infectiousness α constant, the free-riding effect decreases effort.
2. The perceived-infectiousness effect increases effort, because the agent now believes the disease is more infectious than initially thought (perceived α increases), increasing the gain from social distancing.

Intuitively, the perceived-infectiousness effect varies monotonically with R: when infections are low, the effort of supporters is low, and both supporters and nonsupporters get infected at similar rates. Hence, agents revise the estimate of infectiousness α only slightly upward in response to the treatment. On the other hand, when infections are high, the effort of supporters is high, and the upward revision will be larger.

The following theorem makes this intuition precise. Instead of doing comparative statics on R (which is determined in equilibrium), we state the comparative-statics results in terms of the infectiousness $\hat{\alpha}$ (for given x and $\hat{x}$). Note that R increases with $\hat{\alpha}$.

> THEOREM 1. Assume an agent is informed that a share $\tilde{x} > x$ of the population supports social distancing. Then there is a threshold $\hat{\alpha}^*$ such that for any $\hat{\alpha} < \hat{\alpha}^*$, the free-riding effect dominates and equilibrium effort decreases, and for $\hat{\alpha} > \hat{\alpha}^*$ the perceived-infectiousness effect dominates and the equilibrium effort increases.
>
> *Proof.* See section A of the online appendix for the proof.

The interplay between the free-riding and perceived-infectiousness effects also yields analogous predictions regarding a central belief about COVID-19: the future infection rate. In the end-line survey, we ask respondents to estimate this. The expected future rate differs from the current infection rate R, because this study occurs at a point when infection rates are clearly evolving. The misperceptions-correction treatment changes respondent beliefs about social-distancing support and about infectiousness and therefore should change expected future infection rates. Recall that nonsupporters are always infected with higher probability than supporters. The higher the infectiousness parameter $\hat{\alpha}$, the higher should be future infection rates for both groups. When $\hat{\alpha}$ is currently small, the perceived-infectiousness effect is small. Simultaneously, the treatment corrects beliefs about the share of social-distancing supporters upward, which should reduce estimates of future infection rates because supporters have lower infection rates. Thus, the expected future infection rate decreases when $\hat{\alpha}$ is currently small. In contrast, when $\hat{\alpha}$ is currently large, the treatment leads to a large increase in perceived infectiousness, implying that the disease will infect higher shares of both supporters and nonsupporters. This will tend to increase expected future infection rates.

To summarize, the misperceptions-correction treatment effect on the expected future infection rate should show heterogeneity similar to that described in theorem 1. The treatment effect on the expected future infection rate is strictly negative if the current local infection rate (R, which moves monotonically with $\hat{\alpha}$) is small enough. The treatment effect on the expected future infection rate increases with the current infection rate and can become positive if current infection rates are sufficiently high.

In our empirical analyses, we test these predictions regarding heterogeneity in the misperceptions-correction treatment effect.

III. Sample and Data

A. Data

We implemented three rounds of surveys by phone in July–November 2020: a prebaseline survey, a baseline survey, and an end-line survey (for a study timeline, see fig. A.2; figs. A.1–A.3 are in the online appendix). Respondents were drawn across 76 communities in central Mozambique from a sample of a prior study (Yang et al. 2021) that focused on HIV-vulnerable households—a policy-relevant sample especially vulnerable to COVID-19.[6] To avoid risk of spreading COVID-19 via in-person interaction with study participants, we also limited the sample to those households with phones. Thus both HIV vulnerability and phone ownership are two relevant factors to bear in mind when considering

[6] The American Economic Association RCT Registry for Yang et al. (2021): https://doi.org/10.1257/rct.3990-5.1. In that prior study, we run a randomized evaluation of a bundled community-level HIV/AIDS program whose main component was home visits by case-care workers to promote HIV testing to HIV-vulnerable households, e.g., those with HIV-positive or other chronically ill members, orphaned children, or a grandparent as the household head. In this study, we use community-stratified randomization and regress with community fixed effects to rule out the influence of this prior intervention on our results.

the external validity of the results. We surveyed one adult per household. Appendix section B provides details on the COVID-19 context, study communities, and the study timeline.

Between the prebaseline survey and the baseline survey, we randomly assigned households to treatments and registered a preanalysis plan (PAP). The baseline survey was immediately followed by over-the-phone treatment implementation. There was a minimum of 3.0 weeks and average of 6.3 weeks between baseline and end-line surveys for all respondents. Baseline and end-line surveys occurred when COVID-19 cases were rising rapidly.

The end-line sample size is 2,117 respondents, following a sample size of 2,226 at baseline. The retention rate between baseline and end line is 95.1% overall, at least 94.4% in each of the seven districts surveyed, and balanced across treatment conditions. We also surveyed 145 community opinion leaders over the 76 study communities—at least one, an average of 2.11, and at most 4 per community—for inputs to the primary outcome and treatments as described below.

B. Measuring Misperceptions

We measure both true and perceived support for social distancing as follows. First, to measure actual community support for social distancing, we asked respondents “Do you support the practice of social distancing to prevent the spread of coronavirus? (Yes, No, Don’t Know, Refuse to Answer),” which captures the respondent’s first-order belief of the injunctive norm for social distancing. We then calculated the fraction of “Yes” responses across the sample and within each community.[7] Directly after, to measure perceived community support, we asked respondents “For every 10 households in your community, how many do you think support the practice of social distancing to prevent the spread of coronavirus? (integer 0–10),” capturing the respondent’s second-order beliefs of the injunctive norm for social distancing within his or her community. The difference between the true and perceived community support for social distancing is the respondent’s misperception of the social norm.

Three possible concerns with our measure of perceived support for social distancing include the role of uncertainty, the restricted scale, and bias from experimenter demand effects. First, a possible concern is that unawareness and uncertainty around new social norms and others’ beliefs—plausible at the start of the pandemic—may lead respondents away from the extreme points of the answer scale. However, in appendix section C, we present a cumulative

[7] The fraction was calculated by dividing the number of “Yes” responses by the number of all responses (i.e., Yes, No, Don’t Know, Refuse to Answer).

distribution of our perceived community support measure across survey rounds that shows respondents readily utilized the extreme ends of the scale, with 8% and 35% of the sample at prebaseline reporting perceived community support of 0% and 100%, respectively, and 51% of the sample at baseline reporting 100%. Second, despite more common use of a 0–100 scale when measuring perceived norms (e.g., Falk et al. 2021; Fuhrmann-Riebel et al. 2023), we simplified our scale to an 11-point 0–10 scale due to past difficulties eliciting "percentage" measures in this context, repeated feedback from our field team and local partners that a 0–100 scale was too complex, and the inability to use a "slider" mechanism over the phone. Given the high concentration of perceived community support at 100% at baseline, the restricted scale may attenuate the treatment effect of the misperceptions correction on perceived community support given that there is "little room to improve" for many respondents in the sample. Third, experimenter demand effects may have led respondents to report higher shares of perceived support for social distancing in order to make their communities look favorable. Such action would lead to an upward-biased estimate of true perceptions of community support and, in turn, an underestimate of the misperception of the social norm, which would also lead to an attenuation of the treatment effect for the misperceptions-correction intervention.[8] We ask the reader to bear in mind these possible limitations when interpreting the results.

C. *Primary Outcome*

The primary outcome is an indicator that the respondent practiced social distancing, as prespecified in our PAP. It is constructed from self-reports of social distancing and from others' reports of the respondent's social distancing. The outcome is equal to 1 if the respondent is practicing social distancing according to both self-reports and reports of others, and 0 otherwise.

Respondents are social distancing according to their self-report if both of the following are true: (1) they answer "yes" to "In the past 14 days, have you observed the government's recommendations on social distancing?"; and (2) they report doing more than the sample median number of "social-distancing actions" in the past 7 days. A list of eight social-distancing actions and their corresponding summary statistics is presented in appendix section D. At prebaseline and baseline, respondents were asked about a randomly selected four

[8] See sec. IV.A for a description of the misperceptions-correction treatment. If upward-biased estimates of perceived support remain less than or equal to true community support, then the misperceptions correction is implemented and may still boost respondents' true perception of community support in a way not captured by our measure; however, if the bias leads to overestimating true community support, then respondents will become ineligible for the misperceptions-correction treatment thereby attenuating the treatment effect (but not biasing upward).

social-distancing actions and, with a sample median of three for both surveys, had to report doing all four actions to be considered to be social distancing. At end line, respondents were asked about all eight social-distancing actions and, with a sample median of six, had to report doing seven or eight actions to be considered to be social distancing.[9]

To collect others' reports on a respondent's social distancing, study participants were asked about their social interactions with 10 other community study participants. These 10 others were identified from social network data and geographic proximity. Additionally, community leaders were also asked about social interactions with all study participants in their respective community.[10] At baseline, the average respondent household was known by 0.98 community leaders and 3.21 neighboring survey respondents. The reports of others were collected at baseline and end line.

In collecting the reports of others, we asked others whether they had seen anyone from the respondent household in the past 14 days.[11] If so, we then asked: (1) Did he/she come closer than 1.5 meters to you or others not of his/her household at any point in the past 14 days?; (2) Did he/she shake hands, try to shake hands, or touch you or others not of his/her household in the past 14 days?; and (3) In general, did he/she appear to be observing the government's recommendations on social distancing (avoid large gatherings and keep at least 1.5 meters distance from people not of his/her household)? Respondents are considered to be social distancing according to others if all others responded "no," "no," and "yes" (respectively) to these three questions, reported having not seen the respondent in the past 14 days, or reported not knowing the respondent.[12]

Figure 1 displays how these questions lead to the social distancing outcome. First, 95% of respondents say "yes" to the self-report on general social distancing. When considering self-reports above the sample median number of social-distancing actions, the social-distancing rate falls to 36%. Finally, incorporating

[9] While this threshold was prespecified, results are robust to alternate definitions of this component (see app. sec. G.3), e.g., a threshold of six, or dropping social-distancing actions no. 4 and no. 6 for which respondents might misinterpret and answer "No" if not showing symptoms.

[10] The average community leader was asked about 33.90 households (standard deviation = 22.10, minimum = 2, second highest = 99, maximum = 228—a special case where one individual was the traditional leader across multiple communities). To mitigate survey fatigue, leaders were told up front of the number and offered a stepwise incentive that increased for each additional set of 25 study households.

[11] As is common in this context, households were identified by the name of the household head and a list of other known household members.

[12] At baseline, 90.55% of respondent households were known by some other respondent or community leader.

the reports of others reduces the rate further to 8%. Limited overlap between self-reports and reports of others on social distancing suggests that each is providing different sets of information. We suspect that self-reports likely overreport social distancing because of experimenter demand bias, whereas others' reports are likely less biased by experimenter demand and rather over-report because of recall bias or lack of observation (as respondents not known or not seen in the past 14 days were not assumed to violate social-distancing behavior).[13] Together, we believe the combined measure is a novel improvement from simple self-reports, though we leave comparison of both measurement methods to observed behavior as an avenue for future work. Incorporating additional information into the social-distancing measure—by using self-reports of specific social-distancing behaviors and reports of others—leads to substantially lower social-distancing rates.

IV. Research Design

A. Treatments

We implemented a randomized controlled trial to estimate the effects of two treatments on social distancing: (1) misperceptions correction and (2) leader endorsement.[14] Before the baseline survey, we randomly assigned 30% of households that completed the prebaseline survey to one of the two treatments, 30% of households that completed the prebaseline survey to the other treatment, and the remaining 40% of households that completed the prebaseline survey to a control group. Sample sizes by treatment condition were as follows: misperceptions correction ($N = 628$, 29.7% of sample), leader endorsement ($N = 637$, 30.1% of sample), and control group ($N = 852$, 40.3% of sample). Treatment scripts are located in appendix section E.

For the misperceptions-correction treatment, we used the following data: (1) respondents' own support for social distancing from the prebaseline survey, from which we estimated the true share of community support for social distancing (as the fraction of respondents expressing support within the community), and (2) respondents' perceived share of community support for social distancing at baseline (reported as an integer out of 10). Immediately after completing the baseline survey, treated individuals underestimating the share were told the true share that supported social distancing, rounded to an

[13] For example, complete lack of observation by others was true for 9% of the sample (see n. 12).

[14] These two treatments were registered in a PAP uploaded to the American Economic Association RCT Registry (registration ID no. AEARCTR-0005862: https://doi.org/10.1257/rct.5862) prior to the start of the intervention at baseline. Previously, our American Economic Association RCT pretrial profile had also included a third social-distancing treatment arm proposing to provide individuals with information on the private and public value of social distancing; however, we cut this treatment to improve power for the remaining treatments prior to registering the PAP.

integer out of 10.[15] Treated individuals correctly estimating the share were also told that they were correct. In practice, 92.4% of treated respondents received this treatment, 53.2% of whom underestimated community support for social distancing and 46.8% of whom correctly estimated it. The small minority overestimating the share was not provided additional information.[16]

For the leader-endorsement treatment, we identified and surveyed community opinion leaders prior to the baseline survey and requested their permission to tell others in their community that they "support social distancing, are practicing social distancing, and encourage others to do the same." Then, in this treatment, we reported this endorsement to respondents, mentioning the community leader(s) by name.[17]

Attrition between baseline and end line is low (4.9%). In appendix section F, we show that attrition and key baseline variables are balanced across treatment conditions. Further, at end line, 97.9% recall receiving the baseline survey, and, of those, 99.4% report trusting the COVID-19 information we provided.[18]

B. *Regressions*

A prespecified ordinary least squares regression equation provides treatment-effect estimates:[19]

$$Y_{ijd} = \beta_0 + \beta_1 T1_{ijd} + \beta_2 T2_{ijd} + \eta B_{ijd} + \boldsymbol{\delta}_{ijd}^{\text{others}} + \boldsymbol{\delta}_{ijd}^{\text{leaders}} + \gamma_{jd} + \varepsilon_{ijd}, \tag{8}$$

where Y_{ijd} is the social-distancing indicator for respondent i in community j and district d; $T1_{ijd}$ and $T2_{ijd}$ are indicator variables for the misperceptions-correction and leader-endorsement treatment groups, respectively; B_{ijd} is the baseline value of the dependent variable; $\boldsymbol{\delta}_{ijd}^{\text{others}}$ is a vector of dummy variables for the number of other respondents (from 0 to 8) who report knowing the respondent's household; $\boldsymbol{\delta}_{ijd}^{\text{leaders}}$ is a vector of indicators for the number of community leaders (from 0 to 4) who report knowing the respondent's household;[20] γ_{jd}

[15] In 63 out of 76 communities (82.9%), the number we convey to respondents is 10 out of 10, and in 13 communities (17.1%), the number is 9 out of 10.

[16] While respondents were not incentivized to truthfully guess community support (for scalability), true beliefs can still be updated for all except those who overestimated true community support with an upward-biased guess; however, the latter case should only attenuate our treatment effect and not bias it upward.

[17] Communities had at least one and an average of 2.09 endorsements from community leaders (standard deviation = 0.94, maximum = 4).

[18] Trust may have arisen from multiple in-person household surveys since 2017 (see Yang et al. 2021).

[19] Appendix sec. G.1 shows that all conclusions are robust to logit and probit specifications.

[20] As prespecified, we cap $\boldsymbol{\delta}_{ijd}^{\text{others}}$ at the first integer that covers more than 90% of the sample and $\boldsymbol{\delta}_{ijd}^{\text{leaders}}$ at the maximum number of leaders found in any community.

is community fixed effects; and ε_{ijd} is a mean-zero error term. We report robust standard errors.[21]

Coefficients β_1 and β_2 represent the intent-to-treat effects of the misperceptions-correction and leader-endorsement treatments (respectively) on social distancing.

We modify equation (8) to estimate heterogeneity in treatment effects with respect to local COVID-19 caseloads:

$$\begin{aligned} Y_{ijd} = {} & \beta_0 + \beta_1 T1_{ijd} + \beta_2 T2_{ijd} + \beta_3 (T1_{ijd} \times \mathrm{COVID}_d) \\ & + \beta_4 (T2_{ijd} \times \mathrm{COVID}_d) + \eta B_{ijd} + \boldsymbol{\delta}_{ijd}^{\mathrm{others}} + \boldsymbol{\delta}_{ijd}^{\mathrm{leaders}} + \boldsymbol{\gamma}_{jd} + \varepsilon_{ijd}. \end{aligned} \tag{9}$$

Equation (9) adds interactions between treatment indicators and the cumulative number of district-level COVID-19 cases per 100,000 population at the start of the end-line survey.[22] Coefficients β_1 and β_2 in equation (9) now represent the effects of the treatments in districts where COVID-19 cases are zero (slightly out of sample), and β_3 and β_4 represent the change in the respective treatment effect for a 1-unit increase in district-level COVID-19 cases per 100,000 population.

C. *Hypotheses*

We prespecified the hypothesis that each treatment (β_1 and β_2 in eq. [8]) would have positive effects. Subject-matter experts (surveyed without knowing results) concurred with this expectation.[23] The mean expert predictions were that the misperceptions-correction and leader-endorsement treatments would increase social distancing by 5.23 and 5.56 percentage points, respectively.

We also test the hypotheses that the effect of the misperceptions-correction treatment on social distancing and on the expected future infection rate will be greater in areas with a higher current COVID-19 infection rate (β_3 in eq. [9] will be positive). We did not prespecify these hypotheses but advance them on the basis of our theoretical model.

V. Results

A. *Pretreatment Descriptives*

Table 1 presents pretreatment summary statistics of social-distancing support, perceptions, and behavior in the first 6 months of the COVID-19 pandemic.

[21] Appendix sec. G.2 shows that clustering standard errors by the 76 communities or 7 districts has minimal effect on standard errors and does not affect whether any coefficients are statistically significant at conventional levels.

[22] The main effect of COVID_d is absorbed by γ_{jd}.

[23] Predictions by 71 individuals provided at https://socialscienceprediction.org/ (survey closing date January 2, 2021).

TABLE 1

SUMMARY STATISTICS OF PRETREATMENT SOCIAL-DISTANCING MEASURES

	Prebaseline			Baseline			
	Observations	Mean	Standard Deviation	Observations	Mean	Standard Deviation	*t*-Test *p*-Value
1. Respondent supports social distancing	2,117	.976[a]	.153	2,117	.989[b]	.104	.001
2. Perceived share of community supporting social distancing	2,109	.689[a]	.313	2,114	.800[b]	.262	.000
3. Primary social-distancing indicator: if rows 4 and 7				2,117	.078	.269	
→4. Self-report social-distancing indicator: if rows 5 and 6	2,117	.383	.486	2,117	.355	.479	.045
→5. Self-report: followed government rules past 14 days	2,117	.952	.214	2,117	.949	.219	.692
→6. Self-report: social-distancing behaviors above median	2,117	.396	.489	2,117	.361	.481	.012
→7. Others' social-distancing indicator: if rows 8 and 9				2,117	.232	.422	
→8. Other households' report of social distancing				2,117	.378	.485	
→9. Leaders' report of social distancing				2,117	.519	.500	

Note. Prebaseline data collected from July 10 to August 16, 2020, and baseline data collected from August 26 to October 4, 2020. Variables are as follows. Row 1: indicator equal to 1 if respondent answers "yes" to supporting "the practice of social distancing to prevent the spread of coronavirus," and 0 otherwise. Row 2: perceived share of households (asked as "for every 10 households") in a community that supports social distancing. Row 3: indicator for social distancing equal to 1 if respondent is social distancing according to self (row 4) and others' reports (row 7), and 0 otherwise. Row 4: indicator for social distancing according to self if respondent answered "yes" to observing the government's recommendations on social distancing in the past 14 days (row 5) and reports doing more than the sample median number of social-distancing behaviors (row 6), and 0 otherwise. Row 7: indicator for social distancing according to others if all other respondents (row 8) and community leaders (row 9) reported not knowing the respondent household, not seeing the respondent household in the past 14 days, or—if seen—that the respondent household (i) did not come closer than 1.5 meters to others outside the household; (ii) did not shake hands, try to shake hands, or touch others outside the household; and (iii) appeared to be observing the government's recommendations on social distancing, and 0 otherwise. All variables have a minimum of 0 and a maximum of 1. The *t*-test column displays the *p*-value of a paired *t*-test on the difference between prebaseline and baseline measures (where prebaseline data are available).

[a] Paired *t*-tests comparing reported and perceived support for social distancing at prebaseline ($p = .000$).

[b] Paired *t*-tests comparing reported and perceived support for social distancing at baseline ($p = .000$).

First, we document a large and statistically significant gap between actual and perceived support for social distancing. At both prebaseline and baseline, more than 97% of respondents support social distancing; however, respondents underestimate the community share expressing such support, on average estimating 69% in a prebaseline survey and 80% at baseline. Second, we observe a large and statistically significant 11 percentage point increase in the perceived share of community support between pretreatment survey rounds, consistent with the idea that misperceptions for new public health behaviors are most prevalent at the start of the public health crisis and then diminish over time as social networks share information. Third, despite increases in reported and perceived support for social distancing, we see small decreases in self-reported social-distancing behavior; in the theoretical model, this behavior is predicted where the current local infection rate is low, as was indeed the case for all study communities prior to the end-line survey.[24]

B. Average Treatment Effects

In column 1 of table 2, we present regression estimates for our primary outcome.[25] Both treatment coefficients are small in magnitude, and neither is statistically significantly different from zero. These findings diverge from expert predictions of treatment effects. We strongly reject the null that our $T1$ and $T2$ treatment-effect estimates are equal to the positive mean expert predictions (p-value $< .001$ in each case).

However, we find the misperceptions correction has a positive effect on measures of perceived community support for social distancing. Analyses presented in appendix section C (not prespecified) show that the treatment effect is concentrated on the lower end of the distribution, having a significant positive effect on a respondent perceiving that at least 50% of households in his or her community support social distancing.

C. Treatment-Effect Heterogeneity

In column 2 of table 2, we present regression estimates of treatment-effect heterogeneity (eq. [9]) with respect to the local infection rate, measured as COVID-19 cases per 100,000 population in the respondent's district.

The misperceptions-correction treatment effect is heterogeneous with respect to local COVID-19 cases. The coefficient on the interaction term with $T1_{ijd}$ is positive and statistically significant at the 1% level. The coefficient on the $T1_{ijd}$ main effect is the predicted effect of misperceptions correction in a district with zero cases (slightly out of sample) and suggests that the misperceptions correction would reduce social distancing by 3.4 percentage points in such a location (statistically significant at the 5% level).

[24] See fig. A.2 to see relatively low levels of new COVID-19 cases in Mozambique during the prebaseline and baseline surveys relative to the end-line survey.
[25] The complete set of prespecified analyses are presented in app. sec. H.

TABLE 2

TREATMENT EFFECTS ON SOCIAL DISTANCING AND EXPECTED COVID-19 ILLNESSES

	Primary Social Distancing Indicator		Perceived Share of People in Community That Will Get Sick from COVID-19	
	(1)	(2)	(3)	(4)
*T*1: Misperceptions correction	.0042	−.0466**	.0418	−.1936**
	(.0140)	(.0191)	(.0322)	(.0944)
*T*2: Leader endorsement	−.0054	−.0258	−.0209	−.0598
	(.0137)	(.0198)	(.0308)	(.0944)
*T*1 × District COVID-19 cases		.0030***		.0073**
		(.0011)		(.0029)
*T*2 × District COVID-19 cases		.0012		.0013
		(.0010)		(.0029)
Observations	2,117	2,117	812	812
R^2	.158	.163	.146	.152
Control mean dependent variable	.0857	.0857	.3590	.3590
Control standard deviation dependent variable	.2801	.2801	.3685	.3685

Note. Dependent variable in cols. 1–2 is defined in table 1. Dependent variable in cols. 3–4 is the expected future infection rate: "For every 10 people in your community, how many do you think would get sick from coronavirus?" (converted to share from 0 to 1). "*T*1: Misperceptions correction" is equal to 1 if respondent was randomly assigned to the misperceptions-correction treatment, and 0 otherwise. "*T*2: Leader endorsement" is equal to 1 if respondent was randomly assigned to the leader-endorsement treatment, and 0 otherwise. "*T*1 × District COVID-19 cases" and "*T*2 × District COVID-19 cases" are the respective treatment indicators interacted with district-level cumulative COVID-19 cases per 100,000 population at the start of the end-line survey (see app. sec. B.3, table A.1, col. 2; tables A.1–A.14 are in the online appendix). All regressions control for a baseline measure of the dependent variable, a vector of indicators for number of community leaders knowing the respondent at baseline (0 through 4), and a vector of indicators for number of other respondents knowing the respondent at baseline (0 through 8). All regressions also include community fixed effects. Robust standard errors are in parentheses.

** $p < .05$.

*** $p < .01$.

Figure 2 displays this treatment effect heterogeneity. We plot district-specific treatment effects (estimating eq. [8] separately in each of seven districts) on the *y*-axis (with 95% confidence intervals) against district case counts on the *x*-axis. In the six districts with the lowest case counts, coefficients are negative. By contrast, in Chimoio, the district with the most cases (39.08/100,000) that also accounts for one-quarter of the sample, we estimate a large positive effect: 9.2 percentage points—a 70% increase over that district's control group (statistically significant at the 5% level).

This heterogeneous treatment effect holds up to various robustness checks (presented in app. sec. G). First, we run logit and probit specifications of the primary results. Second, we cluster standard errors by community and district. Third, we vary the threshold by which self-reported "social-distancing actions" were incorporated in the social-distancing indicator. Fourth, we test four

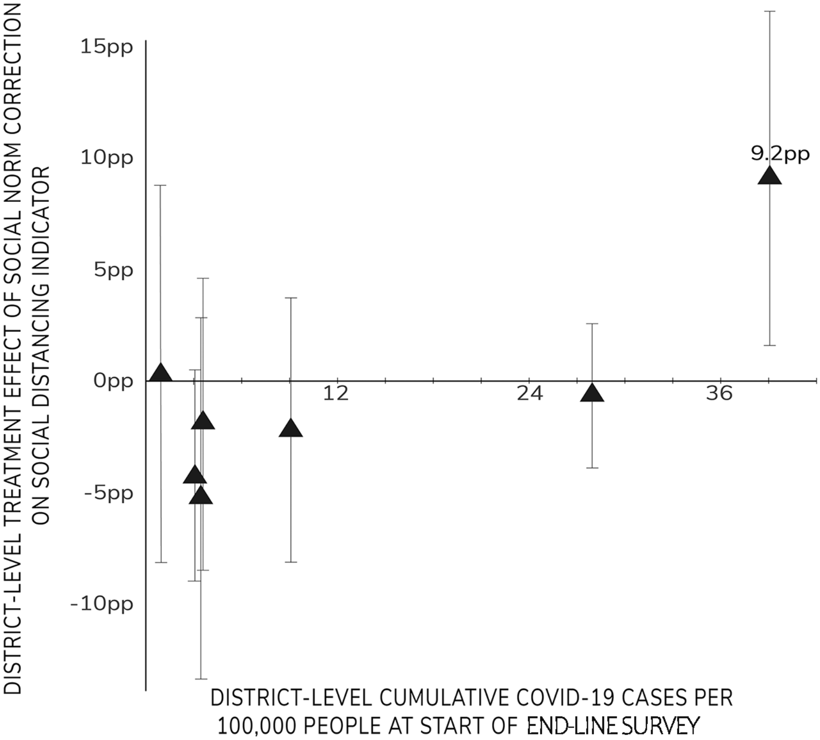

Figure 2. District-level misperceptions-correction treatment effects by COVID-19 cases. Misperceptions-correction treatment effects (triangles) estimated separately for each of seven districts (with 95% confidence intervals). District-level treatment effects plotted on the vertical axis against district-level cumulative COVID-19 caseloads at start of end-line survey (per 100,000 population) on the horizontal axis.

alternative measures of the local COVID-19 infection rate, including the simple case count (not per capita) and high case count indicators, to show that the treatment-effect heterogeneity is not unique to our preferred measure. Fifth, we exclude the top COVID-19 and largest-sampled district of Chimoio to verify that it alone is not driving our results. In all cases, we find that our primary results are very similar.

By contrast, the leader-endorsement treatment effect is not heterogeneous with respect to local caseloads. The coefficient on the corresponding interaction term in column 2 of table 2 is small in magnitude and not statistically significantly different from zero. Some reasons why this treatment may not be effective, even when COVID-19 cases are high, include limited familiarity of leaders among all community members or limited confidence that the leader's endorsement reflected true beliefs rather than political "lip service." Coupled with findings from Banerjee et al. (2019) on gossips spreading information, the result suggests that network-central individuals may be effective at transmitting information but not necessarily because their opinions have a dominating influence on community members' beliefs.

The interplay between the free-riding and perceived-infectiousness effects is the distinctive feature of our theoretical model. When the perceived-infectiousness effect is large enough, it overcomes the countervailing free-riding effect, and the misperceptions-correction treatment leads to more social distancing. An additional implication of the theory is that the treatment should have similar heterogeneous effects on the expected future infection rate.

We conduct this additional test of the theory, examining treatment effects on the expected future infection rate.[26] In columns 3 and 4 of table 2, the outcome is the share of the community the respondent thinks will get sick from COVID-19 (responses were integers out of 10; we divide by 10 to yield a 0–1 scale). In column 3, we estimate average treatment effects. Each coefficient is small in magnitude and not statistically significantly different from zero.

In column 4, we estimate heterogeneity in treatment effects with respect to local cases and find the same pattern as in column 2. The misperceptions correction decreases the expected future infection rate in districts with no cases, and this effect becomes more positive as current cases rise (the $T1_{ijd}$ main effect and interaction term coefficients are both statistically significant at the 5% level).

These treatment-effect heterogeneity findings in columns 2 and 4 of table 2 jointly support the theoretical model. When current infection rates are low, the misperceptions-correction treatment does not change perceived infectiousness much but leads to realizations that social-distancing support is higher than previously thought. People therefore reduce estimates of the future infection rate and also reduce their own social distancing (choosing to free ride). By contrast, when current infection rates are high, the treatment causes larger increases in perceived infectiousness. Notwithstanding an increase in the share of social-distancing supporters, people increase their estimate of the future infection rate and increase their social distancing.

VI. Conclusion

Support for social distancing increased rapidly during the COVID-19 pandemic. If people are unaware of the extent to which others' beliefs on social distancing have changed, would revealing true high rates of such support lead

[26] The question is, "For every 10 people in your community, how many do you think would get sick from coronavirus?" Sample sizes in these regressions are smaller. We implemented this question midway through the end-line survey after finding preliminary evidence suggesting the need to explore mechanisms behind treatment-effect heterogeneity.

to more social distancing? In theory, the effect of providing such information is ambiguous: it could reduce social distancing if free-riding effects dominate but could have a positive effect on social distancing if perceived-infectiousness effects dominate. Perceived-infectiousness effects are more likely to dominate when the current local infection rate is higher.

We implemented a randomized controlled trial to test the effect of a “misperceptions-correction” treatment, which revealed high community support for social distancing. The treatment effect on social distancing exhibits the spatial heterogeneity predicted by theory: negative in areas with low infection rates (reflecting the dominance of free-riding effects) and more positive in areas with higher rates (as perceived-infectiousness effects become increasingly prominent). In the area with the most cases, amounting to one-quarter of our sample, the treatment effect is positive and large in magnitude. The treatment effect on the expected future infection rate shows similar heterogeneity, confirming an additional theoretical prediction.

Our results suggest that when local infection rates are high, health policies shifting perceptions of community support for social distancing upward could help promote social-distancing behavior. Future research is needed to confirm the external validity of these findings and determine how the results translate to other contexts. For example, in cities, looser social networks among neighbors might lead to larger misperceptions of community support while population-dense housing might further activate the perceived-infectiousness effect; alternatively, in communities with lower baseline support for social distancing, a misperceptions-correction treatment may be less motivating but may also potentially “gain more ground” among those with the lowest support who also underestimate the social norm. These findings may also help predict the effects of analogous public health messaging that communicates community support for preventive measures against other infectious diseases.

References

Andersson, Ola, Pol Campos-Mercade, Armando Meier, and Erik Wengstrom. 2021. “Anticipation of COVID-19 Vaccines Reduces Social Distancing.” Unpublished manuscript. https://doi.org/10.2139/ssrn.3765329.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2019. “Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials.” *Review of Economic Studies* 86, no. 6:2453–90. https://doi.org/10.1093/restud/rdz008.

Banerjee, Abhijit, Eliana La Ferrara, and Victor H. Orozco-Olvera. 2019. “The Entertaining Way to Behavioral Change: Fighting HIV with MTV.” NBER Working Paper no. 26096 (July), National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w26096.

Benabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126, no. 2:805–55. https://doi.org/10.1093/qje/qjr002.

Bicchieri, Cristina, and Eugen Dimant. 2022. "Nudging with Care: The Risks and Benefits of Social Information." *Public Choice* 191:443–64. https://doi.org/10.1007/s11127-019-00684-6.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020. "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." *American Economic Review* 110, no. 10:2997–3029.

Cato, Susumu, Takashi Iida, Kenji Ishida, Asei Ito, Kenneth Mori McElwain, and Masahiro Shoji. 2020. "Social Distancing as a Public Good Under the COVID-19 Pandemic." *Public Health* 188:51–53. https://doi.org/10.1016/j.puhe.2020.08.005.

Fafchamps, Marcel, Ana Vaz, and Pedro C Vicente. 2020. "Voting and Peer Effects: Experimental Evidence from Mozambique." *Economic Development and Cultural Change* 68, no. 2:567–605.

Falk, Armin, Peter Andre, Teodora Boneva, and Felix Chopra. 2021. "Fighting Climate Change: The Role of Norms, Preferences, and Moral Values." CESifo Working Paper no. 9175, CESifo, Munich.

Fuhrmann-Riebel, Hanna, Ben D'Exelle, Kristian López Vargas, Sebastian Tonke, and Arjan Verschoor. 2023. "Correcting Misperceptions about Trends and Norms to Address Weak Collective Action." Unpublished manuscript. https://doi.org/10.2139/ssrn.4311583.

Hershey, John C., David A. Asch, Thi Thumasathit, Jacqueline Meszaros, and Victor W. Waters. 1994. "The Roles of Altruism, Free Riding, and Bandwagoning in Vaccination Decisions." *Organizational Behavior and Human Decision Processes* 59, no. 2:177–87. https://doi.org/10.1006/obhd.1994.1055.

Jakubowski, Aleksandra, Dennis Egger, Carolyne Nekesa, Layna Lowe, Michael Walker, and Edward Miguel. 2021. "Self-Reported vs. Directly Observed Face Mask Use in Kenya." *JAMA Network Open* 4, no. 7:e2118830. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2782566.

Janzwood, Scott. 2020. "The Social Distancing Norm Cascade: The Role of Belief Systems in Accelerating Normative Change during the COVID-19 Pandemic." CI Technical Paper no. 2020-1, Cascade Institute, Victoria, BC. https://cascadeinstitute.org/wp-content/uploads/2020/05/Janzwood-Social-Distancing-Norm-Cascade-0514.pdf.

Lau, Krystal, Marisa Miraldo, Matteo M. Galizzi, and Katharina Hauck. 2019. "Social Norms and Freeriding in Influenza Vaccine Decisions in the UK: An Online Experiment." *Lancet* 394, no. 2:65. https://doi.org/10.1016/S0140-6736(19)32862-4.

Martinez, Deborah, Cristina Parilli, Carlos Scartascini, and Alberto Simpser. 2021. "Let's (Not) Get Together! The Role of Social Norms on Social Distancing during COVID-19." *PLoS ONE* 16, no. 3:e0247454. https://doi.org/10.1371/journal.pone.0247454.

Schultz, Wesley P., Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive

Power of Social Norms." *Psychological Science* 18, no. 5:429–34. https://doi.org/10.1111/j.1467-9280.2007.01917.x.

Spiegler, Ran. 2020. "Behavioral Implications of Causal Misperceptions." *Annual Review of Economics* 12:81–106. https://doi.org/10.1146/annurev-economics-072219-111921.

Yang, Dean, James Allen IV, Arlete Mahumane, James Riddell IV, and Hang Yu. 2021. "Knowledge, Stigma and HIV Testing: An Analysis of a Widespread HIV/AIDS Program." NBER Working Paper no. 28716 (April), National Bureau of Economic Research, Cambridge, MA. https://doi.org/10.3386/w28716.

Yu, Hang. 2020. "Three Essays in Development Economics." PhD diss., Department of Economics, University of Michigan.