Tracking and Promoting Usage of a COVID-19 Contact Tracing App

This is a preprint of an article published in *Nature Human Behaviour*. The final authenticated version is available online at: <https://doi.org/10.1038/s41562-020-01044-x>.

Simon Munzert^{a*}, Peter Selb^b, Anita Gohdes^a,
Lukas Stoetzer^c, Will Lowe^a

^aHertie School,

Friedrichstr. 180, 10117 Berlin, Germany

^bDepartment of Politics and Public Administration, University of Konstanz,
Universitätsstr. 10, 78464 Konstanz, Germany

^cSCRIPTS, Humboldt University of Berlin,
Universitätsstr. 3b, 10117 Berlin, Germany

*To whom correspondence should be addressed; E-mail: munzert@hertie-school.org.
ORCID 0000-0003-4760-367X

Abstract Digital contact tracing apps have been introduced globally as an instrument to contain the COVID-19 pandemic. Yet, privacy by design impedes both the evaluation of these tools and the deployment of evidence-based interventions to stimulate uptake. We combine an on-line panel survey with mobile tracking data to measure the actual usage of Germany's official contact tracing app and reveal higher uptake rates among respondents with an increased risk of severe illness, but lower rates among those with heightened risk of exposure to COVID-19. Using a randomized intervention, we show that informative and motivational video messages have very limited effect on uptake. However, findings from a second intervention suggest that even small monetary incentives can strongly increase uptake and help make digital contact tracing a more effective tool.

As governments around the globe are seeking targeted exit strategies from lockdown measures to contain coronavirus disease 2019 (COVID-19), evidence of widespread pre-symptomatic transmission and short generation times (1, 2) calls into question traditional containment measures based on symptomatic surveillance. An early influential modeling study (3) suggested controlling the epidemic with digital contact tracing (DCT), that is, mobile apps that log and report encounters between infected persons and mobile users to prevent onward transmission. Apparently successful rollouts in Singapore and South Korea (4) have, at the time of writing (December 17th, 2020), encouraged more than 40 countries to introduce DCT apps (5).

Broad uptake in the population is considered key for DCT app effectiveness; an influential study suggests that a 60% adoption rate would be sufficient (3), although their simulations exclude supporting measures such as facial masks and social distancing (6). Minimizing uptake differentials across social groups is also important. Not only does unequal access to smartphone technology exacerbate existing inequities and raise ethical concerns (7), but the overall effectiveness of DCT apps depends on the users' contact network structures and mixing behavior (8, 9). For instance, app users might practice more social distancing and thus have fewer social contacts than nonusers (10), which would link app usage unfavorably to exposure and transmission risk.

In this paper we provide evidence on (a) selective uptake and usage of Germany's official Corona-Warn-App (henceforth, CWA, or simply the App) and (b) the effectiveness of interventions for reducing inequities in uptake and, most importantly, for boosting uptake rates in the population. We combine mobile tracking data with a three-wave online panel survey to evaluate interventions designed to stimulate App usage. Privacy-by-design principles such as data minimization and purpose limitation have guided the development of many DCT apps (11) but while privacy-preserving design may contribute to an app's public acceptance (12), it impedes evaluation of app usage. Mere download statistics are silent about actual usage and user profiles

(13) and while surveys can be used to measure usage patterns (14), reliance on self-reports of socially sensitive behavior can generate reporting biases (15, 16). Our study design evades these issues by linking temporally fine-grained behavioral data on app usage with a survey capturing individual measures of a rich set of subject characteristics.

Results

Our study is designed to both track and stimulate usage of the official German DCT app over a period of about 100 days (see Supplementary Discussion for background information). Fig. 1 illustrates the study setup. See the Methods section for a more detailed description of subject recruitment and panel design. Our initial survey sample included $n = 2,044$ individuals recruited from a commercial access panel using quotas to reflect the age, gender, and education distribution of the adult online population resident in Germany (see Supplementary Information). The provider also tracked the online behavior of a volunteer panel of $n = 1,132$ individuals, 649 of whom participated in the survey. Our study population therefore consists of three subgroups: a tracking-only group for which we have sparse but high-quality behavioral data ($n = 482$) and which we use as baseline control for the analysis of tracked app usage in the analyses, a survey-only group for which we have rich self-reports but no behavioral data ($n = 1,395$), and a survey-tracking group for which we have both self-reports and behavioral data ($n = 649$).

After completing an initial survey (July 30-August 11) about social characteristics, and COVID-19-related attitudes and behaviors, survey respondents were randomly assigned to one of two message interventions to stimulate app uptake, or a control group. Attitudinal and self-reported behavioral outcomes were surveyed again in a second panel wave (August 14-24) in which $n = 1,777$ or 87% of the initial respondents participated. Those who reported not to have the App installed in wave 2 ($n = 1,015$) were assigned to one of three treatment groups and offered a monetary incentive of 5, 2, or 1 Euro for agreeing to install the App, or to the

control group (no incentivization). A final panel wave with $n = 1,569$ respondents (77%) was fielded September 10-22 to measure additional outcomes. Mobile tracking data that are used in this study were collected from the date of the CWA's rollout (June 16) to September 22, and are unaffected by the dropout of respondents in later waves (panel attrition).

This study design offers gains in measurement accuracy and granularity at the expense of generalizability to our target population. However, compositional information about this population – smartphone users in Germany whose devices satisfy technical minimum standards (at least Android 6 and iOS 13.5) – is not available. Survey-only respondents are potentially more representative of the target population. Supplementary Information offers more detailed comparisons. It is important to note that our primary interest is App usage conditional on covariates, not in overall usage figures. Conditional distributions often travel between populations more easily, particularly if covariates are relevant for sample selection (17). For this reason, we did not use survey weights in the analyses but did model selection into the tracking sample and panel attrition to ensure these were independent of treatment-related variables (see Supplementary Tables 2–4).

Differential uptake. Fig. 2 displays App uptake, defined as opening the CWA at least once, across groups of survey-tracking respondents. In marked contrast to earlier survey work (14, 18, 19), uptake was more prevalent among older (50+yrs) than younger (18-49yrs) cohorts ($t = 2.88$, d.f. = 576.08, $P = 0.004$, 95% confidence interval (CI) 0.04, 0.19). Moreover, those with medical preconditions that increase the risk of severe illness from the coronavirus – usually older respondents – are more likely to use the App ($t = 2.70$, d.f. = 552, $P = 0.007$, 95% CI 0.03, 0.19). High levels of education ($t = 2.03$, d.f. = 183.61, $P = 0.044$, 95% CI 0.00, 0.23) and household income ($t = 2.74$, d.f. = 361.12, $P = 0.006$, 95% CI 0.04, 0.24) correlated positively with uptake, while gender ($t = 1.38$, d.f. = 620.23, $P = 0.169$, 95% CI

−0.02, 0.13) and parental status ($t = 0.18$, d.f. = 607.97, $P = 0.859$, 95% CI −0.07, 0.08) did not.

Respondents who reported following the guidelines to wear a mask, wash their hands frequently, and practice social distancing also tended to use the App at higher rates ($t = 4.07$, d.f. = 280.51, $P < 0.001$, 95% CI 0.09, 0.26). This suggests that app usage is unfavorably linked to compliance behavior with other NPIs such that those who already report to minimize risk by following established guidelines are also more likely to use the App.

Respondents who reported COVID-19 cases in their personal network, or live in areas with registered COVID-19 outbreaks ($t = 1.87$, d.f. = 158.03, $P = 0.063$, 95% CI −0.00, 0.20), reported using public transport ($t = 1.84$, d.f. = 219.46, $P = 0.068$, 95% CI −0.01, 0.18), visiting friends and family ($t = 1.88$, d.f. = 619.1, $P = 0.061$, 95% CI −0.00, 0.15), or restaurants and bars ($t = 0.87$, d.f. = 118.30, $P = 0.386$, 95% CI −0.06, 0.16) more than once in the week prior to the survey were also estimated more likely to use the App although these effects were not statistically significant at the pre-registered $P < 0.05$ level.

Uptake rates were significantly higher among respondents who trusted the national government ($t = 7.98$, d.f. = 413.38, $P < 0.001$, 95% CI 0.27, 0.44), the health care system ($t = 3.86$, d.f. = 118.36, $P < 0.001$, 95% CI 0.11, 0.33), and science in general ($t = 5.23$, d.f. = 133.46, $P < 0.001$, 95% CI 0.17, 0.38) in comparison to those with low levels of trust towards these institutions. These findings are in line with early evidence on acceptability and use of contact tracing apps in a cross-sectional study from France (18). Uptake was also higher among those who perceived COVID-19 as a threat either to themselves ($t = 2.52$, d.f. = 615.51, $P = 0.012$, 95% CI 0.02, 0.18) or to their friends and family ($t = 3.89$, d.f. = 499.54, $P < 0.001$, 95% CI 0.08, 0.23).

Finally, App usage was significantly higher among the digitally literate ($t = 2.45$, d.f. = 548.5, $P = 0.015$, 95% CI 0.02, 0.18) and those less concerned about data privacy ($t = 2.47$,

d.f. = 629, $P = 0.014$, 95% CI 0.02, 0.17).

Promoting app usage. High levels of App usage are a precondition for it to be effective, yet, after six weeks of deployment, adoption levels plateaued. We investigate if and how app usage can be promoted by embedding two randomized experiments in the panel survey, in wave 1 with treatments designed to increase information and in wave 2 to increase incentives.

Wave 1: Information experiment: wave 1 participants were assigned to one of two information treatment conditions and a control condition. Treated respondents were shown a video of ~2 minutes on the App, that emphasized (a) app functionality, (b) data privacy, and (c) either the benefits of app usage for vulnerable populations (“pro-social message” condition) or for the respondent themselves (“self-interest message” condition).

Our selection of functionality and data privacy messages in wave 1 was designed to mimic existing government campaigns, address well known public concerns, and to build on established social science literature on motivations for individuals to comply with public health and safety initiatives. Previous evidence from public opinion surveys indicates that individuals who do not use the App worry about the App’s effectiveness and its data privacy (20), a finding we replicate (see Supplementary Fig. 19). To address these concerns, the videos shown to participants follow the government’s strategy of highlighting favorable arguments (see, e.g., here: <https://bit.ly/33MXnsX>) and address some of the most frequent public concerns, namely privacy, effectiveness, and general knowledge about the app’s functionality.

For the treatments, we build on existing research on pro-social (21) or self-interested (22) motivations as core factors underlying individual behavior. While the effectiveness of these types of messages has generally been mixed (23, 24), there is some evidence that pro-socially framed messages raise compliance with non-pharmaceutical interventions (25, 26). Recent evidence from a field experiment suggests that *both* pro-social and self-focused messaging frames

can stimulate individuals to seek additional information about COVID-19 (27). The pro-social and self-interest appeals were added as scrolling text on a filmed tablet screen at the end of video (see <https://youtu.be/dyhDdvrGEE> for the pro-social and <https://youtu.be/suOCvIW8R0> for the self-interest message). For more details, see Methods and Supplementary Information.

Wave 2: Incentives experiment: To investigate the costs and benefits of app uptake not linked to an information deficit, in wave 2 we randomly assigned self-reported non-users of the App ($n = 1,015$) to one of three incentivization conditions (or a control condition), offering them an equivalent of 1, 2, or 5 Euros if they would agree to install the App. Respondents who agreed were offered links to the app stores to facilitate compliance (see Supplementary Figs. 8–9).

We build on a large body of research in behavioral and health economics (28, 29) on the effectiveness of monetary incentives to study several outcomes of interest. Our first set of hypotheses relate to the effects of both interventions on app uptake. Here we use tracked measures for the tracking samples, and otherwise rely on reported and hybrid measures of uptake (see Supplementary Fig. 13). We further test for effects on app-related knowledge and attitudes, which could positively impact future uptake and compliance after app installation. For our second set of hypotheses we test whether the treatments spark further interest and motivate respondents to disseminate information within their networks. Both are potentially relevant factors to bolster overall uptake in the population. We tested this by encouraging participants in all treatment conditions to share a message via social media or email, and linked to more information about the App on the web (all tracked via tagged links in the survey). Finally, we asked respondents about actual and hypothetical behavior relevant for digital contact tracing to work. We ask whether they keep their smartphone’s Bluetooth functionality active, whether they would get tested and quarantine after receiving a risk warning from the App, and whether they would notify the App if tested positive.

Given the sample size, pre-treatment covariates, and repeated measures, the minimum de-

tectable effect for most outcomes on the basis of a test with 80% power for our pre-registered level of significance of 0.05 is small ($d < 0.20$) according to a standard definition of effect sizes (30). See Methods and Supplementary Information for further details on variable and model specifications.

Experimental results. To evaluate both the message and the incentivization experiment, we estimate average treatment effects using difference-in-means and saturated regressions (31) to identify the intent-to-treat (ITT) effects (see Methods and Supplementary Information; Supplementary Results for estimates of the Complier Average Causal Effect, CACE).

Information treatments. Both video treatments successfully disseminated factual knowledge about the App: In a pre-registered manipulation check in form of a battery of knowledge items asked directly after the video intervention, treated respondents were able to recall key information from the messages (Extended Data Fig. 1).

Pooling both treatments, the intervention increased knowledge about the App (standardized coefficient $B_{\text{message, pooled}}^{\text{ITT}} = 0.25$ standard deviations (SD); 95% confidence interval (CI) 0.16, 0.35; $P < 0.001$; Fig. 3 and Extended Data Fig. 2) and also induced positive attitudes towards it ($B_{\text{message, pooled}}^{\text{ITT}} = 0.11$ SD; 95% CI 0.05, 0.17; $P = 0.001$).

We do not find significant effects on uptake, whether measured using tracking data ($B_{\text{message, pooled}}^{\text{ITT}} = 0.01$ SD; 95% CI $-0.02, 0.03$; $P = 0.533$) or reported data ($B_{\text{message, pooled}}^{\text{ITT}} = 0.02$ SD; 95% CI $-0.02, 0.06$; $P = 0.282$). Only when using the hybrid measure do we estimate a small significant positive effect ($B_{\text{message, pooled}}^{\text{ITT}} = 0.03$ SD; 95% CI 0.00, 0.06; $P = 0.039$).

We found no evidence for a statistically significant influence of information treatment on sharing messages about the App ($B_{\text{message, pooled}}^{\text{ITT}} = -0.05$ SD; 95% CI $-0.14, 0.03$; $P = 0.222$), looking up additional information about the App ($B_{\text{message, pooled}}^{\text{ITT}} = 0.07$ SD; 95% CI $-0.04, 0.18$; $P = 0.208$) or activating Bluetooth ($B_{\text{message, pooled}}^{\text{ITT}} = 0.07$ SD; 95% CI $-0.03, 0.17$;

$P = 0.169$).

In addition, we find no clear evidence for statistically significant differences between pro-social and self-interest appeals (with the exception of positive attitudes ($B_{\text{pro-social vs. self-interest}}^{\text{ITT}} = -0.07$ SD; 95% CI $-0.14, -0.00$; $P = 0.045$) and activating Bluetooth ($B_{\text{pro-social vs. self-interest}}^{\text{ITT}} = 0.12$ SD; 95% CI $0.02, 0.22$; $P = 0.022$)). We found no evidence that participants who received the pro-social treatment were more likely to install the App (tracked) compared to respondents with self-interest treatment ($B_{\text{pro-social vs. self-interest}}^{\text{ITT}} = 0.00$ SD; 95% CI $-0.04, 0.05$; $P = 0.848$).

Incentives treatments. In the incentive experiment, we demonstrate that even small monetary incentives substantively and significantly increased App uptake among self-reported non-users of the App for all three outcome measures, tracked ($B_{\text{incentive, pooled}}^{\text{ITT}} = 1.05$ SD; 95% CI $0.71, 1.39$; $P < 0.001$; Fig. 3 and Extended Data Fig. 3), hybrid ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.84$ SD; 95% CI $0.66, 1.01$; $P < 0.001$) and reported uptake ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.84$ SD; 95% CI $0.64, 1.03$; $P < 0.001$). While installation rates in wave 3 were lower than rates of willingness to install the App directly after being incentivized (42%-50% subject agreement to install), there is still a 17 percentage point increase in tracked App uptake across treatments ($B_{\text{incentive, pooled}}^{\text{ITT}} = 17$ pp; 95% CI $12, 22$; $P < 0.001$).

We do not find significantly larger uptake rates in the higher-incentive groups compared to the lower-incentive groups ($B_{\text{EUR 5 vs. EUR 1}}^{\text{ITT}} = 0.54$ SD; 95% CI $-0.28, 1.37$; $P = 0.197$; Extended Data Fig. 3). Note, however, that although the differences in effect sizes between the EUR 1 and the EUR 5 incentive tier are estimated to be large ($B_{\text{EUR 1}}^{\text{ITT}} = 0.83$ SD; 95% CI $0.27, 1.39$; $P < 0.001$; Fig. 3) vs. ($B_{\text{EUR 5}}^{\text{ITT}} = 1.43$ SD; 95% CI $0.73, 2.13$; $P < 0.001$) we lack the power to distinguish these differences statistically. In contrast to our pre-registered hypothesis, we do not find significant differences in reported willingness to install the App after incentivization between the different monetary incentive tiers (EUR 1 vs. EUR 2: $t = 0.56$, d.f. = 518.83, $P = 0.579$, EUR 1 vs. EUR 5: $t = -1.34$, d.f. = 493.62, $P = 0.180$, EUR 2 vs.

EUR 5: $t = -1.90$, d.f. = 494.95, $P = 0.059$; Extended Data Fig. 4).

The pooled intervention had a moderate effect on knowledge about the App ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.16$ SD; 95% CI 0.03, 0.28; $P = 0.014$) and made it slightly more likely that respondents intend to share a message about the app ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.07$ SD; 95% CI 0.01, 0.12; $P = 0.019$), but the intervention did not significantly increase positive attitudes towards the App ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.09$ SD; 95% CI $-0.03, 0.21$; $P = 0.156$).

Contrary to our pre-registered expectations, we find no evidence that subjects with high levels of social responsibility or high levels of self-interest were differentially affected by the “pro-social” and “self-interest” messages (Extended Data Fig. 5). For the incentive treatments, we find some evidence for a negative effect of age on uptake (hybrid and reported measure), indicating that young respondents are more responsive to incentives (see Supplementary Tables 42–43 and 51–52).

App usage over time. Fig. 4 displays App uptake rates in different treatment groups over the study period. The trajectory of observed App adoption in the tracking-only group (black curve) as well as the survey-tracking group (red and orange curves) track the official population download figures (see Supplementary Discussion), although they level out at a higher level than the target population. At the beginning of wave 1, 35% (95%-CI [30,39]) subjects in the tracking-only group and 41% (95%-CI [37,45]) subjects in the survey-tracking group had installed the app, which represents a significant difference ($t = 2.18$, d.f. = 1053, $P = 0.030$, 95% CI 0.01, 0.12). The upper panel of Fig. 3 shows no evidence for uptake effects of the information treatments: Over the course of the first wave during which the information treatment was delivered, the uptake-rate increased by 2% to 43% installed (95%-CI [38,48]) in the treatment group and by 2% to 42% installed (95%-CI [35,49]) in the control group.

The uptake increase during second wave is fully explained by the incentive experiment,

where 75% of self-reported non-users were offered incentives to install the App (solid blue curve; lower panel of Fig. 3) and 25% of self-reported nonusers were offered no incentive (dashed blue curve). Financial incentives increased uptake from 8% (95%-CI [4,12]) prior to wave 2 to 20% (95%-CI [14,16]) by the end of wave 2, while in the control group uptake increased from 9% (95%-CI [3,16]) to 11% (95%-CI [4,18]).

At the end of our study, 50% (95%-CI [46,54]) of the survey-tracking subjects (vs. 37% (95%-CI [32,41]) in the tracking-only group) had adopted the App (see Supplementary Information for a more detailed discussion of these differences).

Discussion

In the face of surging cases and an exhaustion of traditional contact tracing capacity(32), adoption rates of COVID-19 contact tracing apps are stagnating. In this article, we provide evidence on uptake and usage of one of the most popular COVID-19 contact tracing apps, Germany's "Corona-Warn-App" and explore experimentally how to make these apps work more effectively by increasing their usage in the population.

This is the first study to link survey information on socio-demographic, behavioral and motivational characteristics with app usage data captured by passive mobile tracking software, allowing us to describe individual-level correlates of observed app usage. Our study design offers key advantages to existing approaches, which likely suffer from reporting bias (see Supplementary Table 19 for illustration with our sample) and attrition.

Our results suggest skepticism about the current effectiveness of contact tracing apps and optimism about the potential to increase it by expanding the user base through incentives. On the one hand, the observed uptake patterns indicate a suboptimal distribution across population strata. High adoption rates among those with high risk of virus exposure and transmission are crucial to maximize the effectiveness of contact tracing. We find significantly lower uptake rates

among those who exhibit lower levels of compliance with other NPI measures and – somewhat surprisingly – among younger age groups. In Supplementary Information we provide additional evidence that the positive relationship between age and app use is likely not just a consequence of selection bias. Nevertheless, our experimental findings show that differential uptake, and low uptake rates more generally, are not predetermined: minimal monetary incentives have a strong impact on app uptake across incentive tiers of 17 percentage points on average. Given the context the experiment was embedded in, this is not a trivial finding. The success of the App in terms of download statistics before our survey (14m downloads at the beginning of the study period and 18.4m by the end (33)) and the fact that uptake in the sample was already higher than in the population made it a challenging case for stimulating uptake.

In a setting where high-profile campaigns and media attention had created considerable levels of awareness of the App in the first place, providing additional information and appealing to the common good or even personal advantage did little to convince people to use the App. Our findings do not imply that we would expect information and arguments to also be unsuccessful in other contexts (25, 26, 34, 35). Rather, they indicate that monetary incentives can mobilize additional compliance when information and arguments fail or have plateaued.

It is important to be aware of the limitations of our study before deriving implications for decision-making (36). Our sample represents, by design, a favorable case for DCT, and issues of realism of experimental stimuli and sample representativeness remain (37, 38). The incentives were delivered in a controlled setting and the set of treatments was constrained. Members of a commercial panel may also be used to receiving direct financial incentives in return for their time. Recent evidence also suggests that commercial survey panel participants report lower levels of pro-social attitudes than population-based survey samples (39), which may contribute towards explaining that we did not find evidence for the pro-social message treatment in our experiment to be effective in boosting uptake. Furthermore, the effects of different stimuli might

depend on context that we did not manipulate in our setup, including information levels and App uptake rates in the population (see Supplementary Discussion for additional information on tracing app usage estimates in other countries). Finally, broad adoption is only one of several criteria that contribute to the overall effectiveness of DCT. We do not address technical issues (40, 41), which might pose an additional barrier to effective DCT, or downstream consequences of low or high uptake rates for classical contact tracing and testing infrastructures (42).

With these limitations in mind, our findings do offer guidance for expanding app usage in an efficient and effective manner (43). After significant investments in advertising and a broad media coverage (44), awareness-raising campaigns may have already reached their limits. Now may be the time to provide more concrete incentives. Expensive interventions such as fully subsidized smartphones (45) are probably neither feasible nor necessary to expand the user base. Financial incentives could be provided in other ways, e.g., by offering free credit on mobile app stores. An important limitation of the current digital contact tracing method used in Germany and many other countries is that it is only available to individuals who have access to a smartphone with an operating system that is compatible with the App's software, thereby possibly exacerbating existing digital and health inequalities (7). Other countries, such as Singapore, are distributing free Tokens capable of exchanging Bluetooth signals, thereby mimicking the functionality of the App (see <https://www.tracetogether.gov.sg/common/token/>).

Our study highlights the potential of passive tracking data to inform public health decision-making beyond existing applications exploiting mobile phone data (46–48). Data collection and impact evaluations such as ours can build on existing survey and tracking panel infrastructure in an increasing number of countries and markets. Notably, in the case of DCT apps the costs needed to establish such digital observatories (in the case of this study, about € 30,000 covering expenses for survey and tracking data, plus human labor costs) are dominated by software development, infrastructure maintenance, and marketing costs. Building informed consent into

a commercial passive tracking panel allows us to use digital data without compromising privacy rights (49), and we recommend this evaluation design to other researchers.

Methods

Consent and ethics. The study was pre-registered prior to data collection on July 29, 2020 at <https://osf.io/6jstp/>. It was approved by the Dean’s office at the Hertie School, which serves in lieu of an ethics committee. Our research complies with GDPR requirements and all relevant ethical regulations, as documented in the DFG Code of Conduct Guidelines for Safeguarding Good Research Practice. The data were collected, shared with the researchers, and published in an anonymous and time-aggregated form with permission of the provider and with informed consent of the participants.

Combining passive behavioral data with survey responses to investigate usage behavior of a contact tracing app, which is designed not to share any usage data in conjunction with auxiliary data about the user with third parties, calls for particular caution. First, following standard practice in survey research, respondents remain anonymous to the researchers. Information that is potentially compromising (such as a combination of zip code with sociodemographic data) is not published as replication data; instead, we only publish as replication data derivative information on COVID-19 case rates in respondents’ local areas that cannot be uniquely linked to their zip code area. Second, individuals who join the passive tracking panel are informed about the type of data collected, and are kept anonymous. They are also informed that the software can be removed or temporarily suspended at any time.

Data. *Subject recruitment and panel design.* In partnership with the online survey firm Repondi, panelists were invited to participate in a study on attitudes in the context of the COVID-19 pandemic with several survey waves. The sample sizes were constrained by both the research

budget and, in the case of the survey-plus-tracking sample, the overall size of the passive tracking panel operated by the provider. The survey design and programming was implemented on our end, while Respondi rewarded respondents with their “mingle points” system. The value of points awarded was nominal as a small compensation for time taken to fill out the survey, and does not exert undue influence (participants of the regular (tracking) panel received the equivalent of EUR 1 (EUR 2) for the initial 20-minute survey and EUR 0.50 (EUR 1) for each of the 10-minute follow-up surveys. Participation was voluntary. Participants had to be at least 18 years old and reside in Germany. Panelists were selected according to their gender (2 groups), age (5 groups) and education (3 groups) to approximate the marginal distributions of the 2019 Best for Planning study (50). Participants from the tracking panel were recruited from the subset of panelists who had the tracking software installed on a mobile device.

Mobile tracking data. Respondi’s tracking panel uses the Wakoopa software, which collects data on web visits and mobile app use on all devices registered by the participant. The collected data is sent via a secure connection to a cloud-based environment. Respondents provided informed consent and are given the ability to pause or halt data sharing at any time. We purchased a subset of the tracking data, covering only usage information on the Corona-Warn-App (timestamp, connection, duration) and device metadata device type (OS type, manufacturer, model). The tracking data we use covers the time period between June 16 (the launch date of the Corona-Warn-App) and September 21. For the 1,132 users tracked in the survey-tracking or the tracking-only sample, we identify 494 users who installed and used the App at least once. Overall, the software logged 16,266 instances of interaction with the App, with a median active usage duration of 18 seconds per interaction with the App (i.e., having the App opened). Due to a technical mistake by the provider, tracking data on CWA use were only available for users of mobile devices with Android, but not Apple, operating systems. Consequently, we excluded respondents from the survey-tracking and tracking-only panel that had an Apple device regis-

tered ($n = 129$ in total; already subtracted from the figures reported above). See Supplementary Information for analyses of additional tracking data by device type.

Message treatment. Following the pre-analysis plan, we implemented the following treatment conditions: (1) Treatment group 1 (“Pro-social message”) receive an explainer video of 2:05 minutes length on the contract tracing app with an emphasis on (a) privacy-related aspects and (b) the benefits of app usage for vulnerable populations. (2) Treatment group 2 (“Self-interest message”) received an explainer video of 2:05 minutes on the contract tracing app with an emphasis on (a) privacy-related aspects (identical to the Treatment 1) and (b) the benefits of app usage for the respondent him- or herself. (3) The control group received no video. Due to lack of time, we did not pilot the videos and only ran internal technical quality checks to adjust the order of the content, design of the scrolling text, and overall message length. The first part of the video message on functionality and privacy was based on a CWA explainer video published by the German edition of the PC World on June 16, 2020 (see <https://youtu.be/I3C9BrC9I-8>).

Incentivization treatment. Following the pre-analysis plan, we implemented the following treatment conditions for those who reported to not having installed the app: (1) The control group received no incentivization and no encouragement to install the app. (2) The treatment groups were incentivized to install the app, but without requiring further commitments. The incentivization varied across three different levels, 100, 200 and 500 Mingle points, the platform-specific currency used by the survey provide, which is equivalent to EUR 1, EUR 2, and EUR 5 (approximately USD 1.30, 2.60, and 5.80, respectively). Subjects were then requested to state their agreement or non-agreement. Agreement automatically qualified for the monetary payoff, which was dealt with by the survey provider. The payment was not conditional on further compliance measures. The provider was informed about respondent agreement at regular intervals while the survey was running. Once a respondent agreed to install the app, they were shown a page with links to the Apple App Store and the Google Play store, along with further

information about the app and a statement clarifying that the research was not affiliated with the government or other authorities. To facilitate installation for those respondents who did not fill out the survey on their mobile phones, they could request to see QR codes, which were then displayed on a follow-up page, and use them together with their phone to access the installation page.

Measurement. *App uptake.* We use both a survey-based measure and the tracking data to measure app uptake. The survey-based measure is the basis for the outcome “reported app uptake”. The tracking-based measure is the basis for the outcome “tracked app uptake”. Furthermore, we construct a “hybrid app uptake” measure to allow us to pool the different samples to analyze effects on app uptake (see Supplementary Fig. 13). To arrive at the survey-based measure, two items are used. First, and in wave 1 only, respondents are asked whether they have access to a smartphone. If the answer is positive, they receive a follow-up question asking them whether they or someone else has installed or not installed the official Corona-Warn-App on their smartphone. The three possible answers are “App installed”, “App not installed”, and “App installed, but uninstalled since then”. In waves 2 and 3, the separate smartphone access question is replaced with an additional response option “I don’t have access to a smartphone”.

Bluetooth usage. Bluetooth usage could not be directly observed in the tracking data. Therefore, we use a survey item in all three waves, asking the respondents who report to have the app installed how often they use the Corona Warning App by activating Bluetooth communication on their smartphone. The five answer categories “Always”, “Mostly”, “Sometimes”, “Rarely”, and “Never” are recoded into numeric scores (Never = 0 to Always = 4).

App knowledge. We used a battery of five right-or-wrong items to measure factual knowledge on the app (see Extended Data Fig. 1). To avoid learning effects, these questions were not asked before treatment in wave 1. The outcome variable is calculated as the total number of

correct answers divided by total number of items. The knowledge battery was run in all waves. The data from wave 1 are exclusively used for the manipulation check and the data from wave 2 is used to estimate the effect of the message treatments on app knowledge reported in the main text.

App attitudes. We measure attitudes towards the app by conducting a principal components analysis (PCA) of their responses to the following set of items: “I don’t think that the app is of any use in fighting the pandemic”, “I think it makes sense that the app is used as an instrument for tracking infections”, “I do not feel well enough informed about the app”, and “I am concerned about the privacy of the app”. We run PCA in the set of responses in wave 1 to derive an attitude index by taking the first principal component, and then predict the attitude index in waves 2 and 3 by using the same set of weights computed in wave 1.

App message sharing behavior and app information lookup. At the end of the wave 1 and wave 2 questionnaires, respondents were shown an information screen that featured links to external pages with additional information about the app as well as links to the Corona Warning App in Apple’s App Store and Google’s Play Store (see Supplementary Fig. 14). Furthermore, additional links offered ways to share information about the app via social media (Facebook, Twitter, WhatsApp) and email. These links were internally tagged, which allowed us to track clicks on any of them. We used this to measure sharing behavior, which we coded 1 if a respondent clicked on any of the Facebook, Twitter, WhatsApp or Email share buttons on the information screen, and otherwise 0. Analogously, we created a measure of information lookup, which we coded 1 if a respondent clicked on any of the information or app store links, otherwise 0.

Hypothetical app usage. First, we asked respondents how sure they were that they would have themselves tested and quarantined for COVID-19 when the app informed them about a recent risk encounter. The seven answer categories (1 = “Certainly not get tested/quarantined” to 7 = “Certainly get tested/quarantined”) were recoded into numeric scores. Second, we asked

how sure they were that they would report a positive test result in the app. The seven answer categories (1 = “Certainly not report” to 7 = “Certainly report”) were recoded into numeric scores.

Covariates. *Age* is measured in five categories (18-29 years, 30-39 years, 40-49 years, 50-59 years, 60+ years). For some descriptives we use age as a continuous variable to facilitate interpretation. *Gender* is measured using a dummy variable (male = 0, female = 1). *Education* is measured in three categories (low = [did not finish school (yet), or finished school but holds no qualification to pursue education to satisfy university entrance requirements], intermediate = [finished school with qualification to pursue further education to satisfy university entrance requirements], high = [finished school achieving university entrance requirements, and/or holds university degree and/or post-graduate degree]). *Number of children* is measured in four categories (No children, 1 child, 2 children, 3 children, 4 or more children) and for some descriptives coded binary as children yes/no to facilitate interpretation). *Household income* is measured in six categories (“Up to EUR 500”, “Between EUR 500 and 1,500”, “Between EUR 1,500 and 3,000”, “Between EUR 3,000 and 5,000”, “Between EUR 5,000 and 10,000”, “More than EUR 10,000”). In the main models, these are collapsed into three categories (“Up to EUR 1,499 EUR”, “EUR 1,500 to 2,999”, “EUR 3,000 and more”). *Living in region with high number of cases* is measured by matching respondents’ zip code to district data published by the RKI as of August 25, 2020. Zip code areas with a mean number of cases larger than the 90% percentile of all zip code areas (here: more than 484 cases per 100k inhabitants) are coded as high-incidence regions. Also using zip code data, we classify the respondent as *living in an urban region* when the zip code belongs to a town or city. *Employment status: working* is measured by classifying the respondent’s employment status as working if he/she reports to be working full time or part time, attending a school or university, or being in a trainee-ship or apprenticeship scheme. *Trust in government, scientists, and the healthcare system* is measured using an item that asks

respondents how much trust they have in the respective institutions. The five answer categories, ranging from “Not trust at all” to “Complete trust” are recoded into numeric scores (“Not trust at all” = 0 to “Complete trust” = 4). *Pre-existing health conditions* are measured by asking respondents whether they “have any pre-existing conditions that increase the risk of a severe course of COVID-19 (e.g., high blood pressure, obesity, diabetes, COPD)” and encoding their answers to yes/no. *Incidences of COVID-19 infections in personal environment* are measured by asking respondents whether they know of anyone (including yourself, family, friends, acquaintances, colleagues, or neighbors) who has been infected by COVID-19. Answers “Yes, 1-3 people” and “Yes, more than 3 people” are encoded as 1, “No” as 0. *Threat perception of COVID-19 for themselves and family/friends* is measured by asking respondents how concerned they are about consequences of the coronavirus for their themselves and family and friends separately. The four-point scale (“Not at all concerned”, “Not too concerned”, “Somewhat concerned”, “Very concerned”) is recoded into numeric scores (“Not at all concerned” = 0 to “Very concerned” = 4). *Data privacy concerns* are measured by conducting a principal components analysis (PCA) of responses to a set of three items. We run PCA with two components on the set of responses in wave 1 to derive a data privacy concerns index by taking the first principal component (see Supplementary Table 15). *Social responsibility* is measured by conducting a principal components analysis (PCA) of responses to a set of three items. We run PCA with one component on the set of responses in wave 1 to derive a social responsibility index by taking the first principal component (see Supplementary Table 16). *Self-interest* is measured by conducting a principal components analysis (PCA) of responses to a set of three items. We run PCA with one component on the set of responses in wave 1 to derive a self-interest index by taking the first principal component (see Supplementary Table 17). *Digital literacy* is measured by conducting a principal components analysis (PCA) of responses to a set of four items. We run PCA with one component on the set of responses in wave 1 to derive a digital literacy index by taking the first

principal component (see Supplementary Table 18). *NPI compliance* is measured by asking respondents to what extent they comply with behavioral recommendations to contain the pandemic. The four-point scale (how concerned they are about consequences of the coronavirus for themselves and their family and friends separately. The four-point scale (“I almost always follow them”, “I try to adhere to them, but often I do not succeed (e.g., for professional reasons)”, “I barely adhere to them”, “I am not aware of these recommendations”) is recoded into a binary indicator of compliance (“I almost always follow them” = 1, otherwise = 0). *Risk behavior* is measured using a battery of items, asking respondents how often in the last seven days they have (a) used public transport, (b) visited a restaurant, café or bar, and (c) met with friends, relatives or acquaintances in person. For each item, the four-point scale (“Never”, “Once this week”, “Several times this week”, “Every day”) was broken down into “Low” (Never, Once this week) and “High” (Several times this week, Every day) risk behavior.

Analysis. *Modelling approach.* To evaluate both the message and the incentivization experiment, we estimate average treatment effects using difference-in-means and saturated regressions (31) to identify the ITT, and the instrumental-variables framework (51) to estimate the Complier Average Causal Effect (CACE). For all covariate-adjusted regressions, we use a lasso procedure to select covariates. The estimates reported in the main text are the covariate-adjusted ITT estimates. All other estimates along with results from minimum detectable effect size calculation are reported in Supplementary Results. Two-tailed statistical tests were used in all analyses and we did not correct the p-values for multiple hypotheses. Respondents were, due to the nature of our experimental manipulations, aware of which treatment they were receiving, although not that it was a treatment. We performed model-specific list-wise deletion of observations with missing values.

Sample usage. For survey-based outcome measures, we pool the survey-only and the survey-

plus-tracking samples. For app usage as outcome, we run three types of analyses. Under the first type, we restrict our analyses to the survey-plus-tracking sample and report both unadjusted and covariate-adjusted estimates. Under the second type, we pool the survey-plus-tracking sample and the tracking-only sample, which provides additional control units, and report both unadjusted and covariate-adjusted estimates. In the latter case, the set of possible covariates reduces to those that are also available for members of the tracking-only panel. Under the third type, we pool all sub-samples and generate a hybrid outcome measure that uses reported app usage in the survey-only sample and the tracked app usage in the tracking sample, if available.

Data availability

Anonymized versions of the datasets used in this study are shared with permission of the data provider and in compliance with GDPR requirements on GitHub at: <https://github.com/simonmunzert/covid19-app-experiment>. Information that would have made subjects directly identifiable was deliberately withheld by the survey provider. In addition, we removed information that could potentially be used to indirectly identify subjects (including zip code and matched numeric COVID case data).

Code availability

The computer code for this study is open source and available on GitHub at: <https://github.com/simonmunzert/covid19-app-experiment>.

References

1. Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, M. Wang, Presumed asymptomatic carrier transmission of COVID-19. *Journal of the American Medical Association* **323**, 1406–1407 (2020).

2. X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* **26**, 672–675 (2020).
3. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368** (2020).
4. S. Whitelaw, M. A. Mamas, E. Topol, H. G. Van Spall, Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health* (2020).
5. P. H. O’Neill, T. Ryan-Mosley, B. Johnson, “A flood of coronavirus apps are tracking us. Now it’s time to keep track of them.”, 2020, (<https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/>).
6. R. Hinch, W. J. Probert, A. Nurtay, M. Kendall, C. Wymatt, M. Hall, K. Lythgoe, A. B. Cruz, L. Zhao, A. Stewart, *et al.*, OpenABM-Covid19 - An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. Preprint at *medRxiv* <https://www.medrxiv.org/content/10.1101/2020.09.16.20195925v1> (2020).
7. M. J. Parker, C. Fraser, L. Abeler-Dörner, D. Bonsall, Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *Journal of Medical Ethics* **46**, 427–431 (2020).
8. R. Canetti, A. Trachtenberg, M. Varia, Anonymous Collocation Discovery: Harnessing Privacy to Tame the Coronavirus. Preprint at *arXiv* <http://arxiv.org/abs/2003.13670> (2020).
9. A. M. Dorélien, A. Simon, S. Hagge, K. T. Call, E. Enns, S. Kulasingam, Minnesota Social Contacts and Mixing Patterns Survey with Implications for Modelling of Infectious Disease Transmission and Control. *Survey Practice* **13** (Aug. 6, 2020).
10. V. von Wyl, S. Bonhoeffer, E. Bugnion, M. A. Puhan, M. Salathé, T. Stadler, C. Troncoso, E. Vayena, N. Low, A research agenda for digital proximity tracing apps. *Swiss medical weekly* **150**, w20324 (2020).
11. N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, S. K. Jha, A survey of covid-19 contact tracing apps. *IEEE Access* **8**, 134577–134601 (2020).
12. S. Trang, M. Trenz, W. H. Weiger, M. Tarafdar, C. M. Cheung, One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. *European Journal of Information Systems*, 1–14 (2020).
13. J. H. Reelfs, O. Hohlfeld, I. Poesse, Corona-Warn-App: Tracing the Start of the Official COVID-19 Exposure Notification App for Germany. Preprint at *arXiv* <http://arxiv.org/abs/2008.07370> (2020).

14. S. Altmann, L. Milsom, H. Zillesen, R. Blasone, F. Gerdon, R. Bach, F. Kreuter, D. Nosenzo, S. Toussaert, J. Abeler, Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. *JMIR mHealth and uHealth* **8**, e19857 (2020).
15. S. Munzert, P. Selb, Can we directly survey adherence to non-pharmaceutical interventions? Evidence from a list experiment conducted in Germany during the early Corona pandemic. *Survey Research Methods* **14**, 205–209 (2020).
16. R. Tourangeau, T. Yan, Sensitive questions in surveys. *Psychological Bulletin* **133**, 859 (2007).
17. E. Bareinboim, J. Pearl, Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* **113**, 7345–7352 (2016).
18. M. Guillon, P. Kergall, Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. *Public Health* **188**, 21–31 (2020).
19. R. Thorneloe, T. Epton, W. Fynn, M. Daly, N. Stanulewicz, A. Kassianos, G. W. Shorter, S.-J. Moll, M. Campbell, S. Sodergren, *et al.*, Scoping Review of Mobile Phone App Uptake and Engagement to Inform Digital Contact Tracing Tools for COVID-19. Preprint at *PsyArXiv* <https://psyarxiv.com/qe9b6/> (2020).
20. A.-K. Sonnenberg, *Neue Corona-Warn-App: Zweifel am Nutzen ist häufigster Grund für eine Nicht-Nutzung [New Corona Warn App: Skepticism about effectiveness is most frequent reason for non-usage]*, 2020, (<https://yougov.de/news/2020/06/22/neue-corona-warn-app-zweifel-am-nutzen-ist-haefigs/>).
21. C. Engel, Dictator games: A meta study. *Experimental economics* **14**, 583–610 (2011).
22. J. Persky, The ethology of homo economicus. *Journal of Economic Perspectives* **9**, 221–231 (1995).
23. A. Bellato, Psychological factors underlying adherence to COVID-19 regulations: A commentary on how to promote compliance through mass media and limit the risk of a second wave. *Social Sciences & Humanities Open* **2**, 100062 (2020).
24. C. Bonell, S. Michie, S. Reicher, R. West, L. Bear, L. Yardley, V. Curtis, R. Amlôt, G. J. Rubin, Harnessing behavioural science in public health campaigns to maintain ‘social distancing’ in response to the COVID-19 pandemic: key principles. *J Epidemiol Community Health* **74**, 617–619 (8 2020).
25. C. van der Linden, J. Savoie, Does Collective Interest or Self-Interest Motivate Mask Usage as a Preventive Measure Against COVID-19? *Canadian Journal of Political Science* **53**, 391–397 (2020).
26. S. Pfattheicher, L. Nockur, R. Böhm, C. Sassenrath, M. B. Petersen, The Emotional Path to Action: Empathy Promotes Physical Distancing and Wearing of Face Masks During the COVID-19 Pandemic. *Psychological Science* **31**, 1363–1373 (2020).

27. S. Banker, J. Park, Evaluating prosocial COVID-19 messaging frames: Evidence from a field study on Facebook. *Judgment and Decision Making* **15**, 1037–1043 (2020).
28. E. Kamenica, Behavioral economics and psychology of incentives. *Annu. Rev. Econ.* **4**, 427–452 (2012).
29. E. L. Giles, S. Robalino, E. McColl, F. F. Sniehotta, J. Adams, The Effectiveness of Financial Incentives for Health Behaviour Change: Systematic Review and Meta-Analysis. *PLOS ONE* **9**, 1–16, (<https://doi.org/10.1371/journal.pone.0090347>) (Mar. 2014).
30. S. S. Sawilowsky, New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* **8**, 26 (2009).
31. W. Lin, Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique. *Annals of Applied Statistics* **7**, 295–318 (2013).
32. M. Linden, J. Dehning, S. B. Mohr, J. Mohring, M. Meyer-Hermann, I. Pigeot, A. Schöbel, V. Priesemann, The foreshadow of a second wave: An analysis of current COVID-19 fatalities in Germany. Preprint at *arXiv* <http://arxiv.org/abs/2010.05850> (2020).
33. Robert-Koch-Institut, “Kennzahlen zur Corona-Warn-App”, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Kennzahlen.pdf, 22 September 2020.
34. L. French Bourgeois, A. Harell, L. B. Stephenson, To Follow or Not to Follow: Social Norms and Civic Duty during a Pandemic. *Canadian Journal of Political Science* **53**, 273–278 (2020).
35. K. Murphy, H. Williamson, E. Sargeant, M. McCarthy, Why people comply with COVID-19 social distancing restrictions: Self-interest or duty? *Australian & New Zealand Journal of Criminology* **53**, 477–496 (2020).
36. H. IJzerman, N. A. Lewis, A. K. Przybylski, N. Weinstein, L. DeBruine, S. J. Ritchie, S. Vazire, P. S. Forscher, R. D. Morey, J. D. Ivory, F. Anvari, Use caution when applying behavioural science to policy. *Nature Human Behaviour* **4**, 1092–1094 (2020).
37. R. Baker, J. M. Brick, N. A. Bates, M. Battaglia, M. P. Couper, J. A. Dever, K. J. Gile, R. Tourangeau, Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology* **1**, 90–143 (2013).
38. N. Keiding, T. A. Louis, Web-based enrollment and other types of self-selection in surveys and studies: consequences for generalizability. *Annual Review of Statistics and Its Application* **5**, 25–47 (2018).
39. E. Hargittai, A. Shaw, Comparing internet experiences and prosociality in amazon mechanical turk and population-based survey samples. *Socius* **6**, 2378023119889834 (2020).
40. J. Klingwort, R. Schnell, Critical Limitations of Digital Epidemiology. *Survey Research Methods* **14**, 95–101 (2020).

41. L. Maccari, V. Cagno, Do we need a Contact Tracing App? Preprint at *arXiv* <http://arxiv.org/abs/2005.10187> (2020).
42. J. Almagor, S. Picascia, Can the app contain the spread? An agent-based model of COVID-19 and the effectiveness of smartphone-based contact tracing. Preprint at *arXiv* <http://arxiv.org/abs/2008.07336> (2020).
43. C. Farronato, M. Iansiti, M. Bartosiak, S. Denicolai, L. Ferretti, R. Fontana, “How to get people to actually use contact-tracing apps”, *Harvard Business Review Digital Articles*, July 15, 2020.
44. Frankfurter Allgemeine Zeitung, *Bundesregierung verrät Kosten: So viel kostet den Staat die Corona-Warn-App [Federal government reveals costs: This is how much the Corona warning app costs the state]*, 2020, (<https://www.faz.net/aktuell/wirtschaft/digitec/so-viel-kostet-den-staat-die-corona-warn-app-16999459.html>).
45. M. Loi, How to fairly incentivise digital contact tracing. *Journal of Medical Ethics*, (doi: 10.1136/medethics-2020-106388) (2020).
46. N. Oliver, B. Lepri, H. Sterly, R. Lambiotte, S. Deletaille, M. De Nadai, E. Letouzé, A. A. Salah, R. Benjamins, C. Cattuto, V. Colizza, N. de Cordes, S. P. Fraiberger, T. Koebe, S. Lehmann, J. Murillo, A. Pentland, P. N. Pham, F. Pivetta, J. Saramäki, S. V. Scarpino, M. Tizzoni, S. Verhulst, P. Vinck, Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* **6** (2020).
47. C. O. Buckee, S. Balsari, J. Chan, M. Crosas, F. Dominici, U. Gasser, Y. H. Grad, B. Grenfell, M. E. Halloran, M. U. G. Kraemer, M. Lipsitch, C. J. E. Metcalf, L. A. Meyers, T. A. Perkins, M. Santillana, S. V. Scarpino, C. Viboud, A. Wesolowski, A. Schroeder, Aggregated mobility data could help fight COVID-19. *Science* **368**, ed. by J. Sills, 145–146 (2020).
48. Y. Zhou, R. Xu, D. Hu, Y. Yue, Q. Li, J. Xia, Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. *The Lancet Digital Health* **2**, e417–e424 (2020).
49. M. Ienca, E. Vayena, On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature Medicine* **26**, 463–464 (2020).
50. Best for Planning, “Berichtsband 2019”, GIK Media, 2019.
51. A. S. Gerber, D. P. Green, *Field Experiments: Design, Analysis, and Interpretation* (W. W. Norton, 2012).

Acknowledgments This research was generously funded by a grant from the Hertie Faculty Activity Fund. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Sebastian Ramirez-Ruiz for excel-

lent research assistance, and Dayna Sadow for proofreading the manuscript. Furthermore, we thank Basak Çali, Andy Guess, Anselm Hager, Christian Traxler, and Julian Wucherpfennig for invaluable feedback. Finally, we are grateful to Holger Nowak, Katharina Storz, and Myrto Papoutsis from ResponDi AG for their support.

Author contributions S.M., P.S., A.G., L.S. and W.L. developed the study concept and study design. S.M. and L.S. programmed the survey. Data preparation and acquisition were performed by S.M. and L.S. S.M., P.S. and L.S. completed the data analysis and interpretation. S.M., P.S., A.G., L.S. and W.L. drafted the manuscript. All authors approved the final version of the manuscript for submission.

Competing interests The authors declare no competing interests.

Figure legends

Figure 1: Overview of the study design. Illustration of app tracking data, sample types, survey waves, and deployment of message and incentivization treatments.

Figure 2: Pre-survey tracked App uptake by subgroups. Uptake rates as observed among survey-tracking panel members ($n = 649$) on July 28, 2020. Uptake rate is reported for each sociodemographic stratum, risk status and behavior, and attitude type. Group means show 83% confidence intervals so non-overlapping intervals indicate a significant difference at $p < .05$. Overall sample uptake rate (0.41) is indicated by black line.

Figure 3: Effect of message and incentive treatments on uptake, knowledge, attitudes, and behavior. Each plot shows standardized ITT estimates with 95% confidence intervals from fully saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. The video message sample comprises $n = 2,044/1,356/1,337$ respondents for the estimation of the pooled/pro-social/self-interest treatment effects respectively. The incentive sample comprises $n = 1,015/513/516/494$ respondents for the estimation of the pooled/EUR 1/EUR 2/EUR 5 treatment effects respectively.

Figure 4: App adoption rates over time for message and incentive groups. The upper plot shows estimated rates and 83% confidence intervals for treated, control, and baseline ‘tracking only’ respondents. The lower plot shows estimated rates and 83% confidence intervals for treated and control respondents in the incentive group. Note that this group is limited to those who never adopted the App or uninstalled it by wave 2. Non-overlapping intervals indicate a significant difference at $p < .05$.

Extended Data

Extended Data Fig. 1: Results of manipulation check of message experiment. Proportions of knowledge items correctly answered, by treatment group ($N_{\text{Pro-social}} = 707$, $N_{\text{Self-interest}} = 690$, $N_{\text{Control}} = 647$). Two-sample t-test statistics: $T_{\text{Pro-social vs. Control}}^{\text{Item}(1)} = 6.09$, d.f. = 1268.4, $P < 0.001$, 95% CI 0.10, 0.19, $T_{\text{Pro-social vs. Control}}^{\text{Item}(2)} = 8.07$, d.f. = 1252, $P < 0.001$, 95% CI 0.15, 0.25, $T_{\text{Pro-social vs. Control}}^{\text{Item}(3)} = 12.10$, d.f. = 1348.70, $P < 0.001$, 95% CI 0.26, 0.36, $T_{\text{Pro-social vs. Control}}^{\text{Item}(4)} = 5.89$, d.f. = 1337.20, $P < 0.001$, 95% CI 0.09, 0.18, $T_{\text{Pro-social vs. Control}}^{\text{Item}(5)} = 1.25$, d.f. = 1349.8, $P = 0.211$, 95% CI -0.01, 0.06, $T_{\text{Pro-social vs. Self-interest}}^{\text{Item}(1)} = -0.04$, d.f. = 1392.3, $P = 0.970$, 95% CI -0.04, 0.04, $T_{\text{Pro-social vs. Self-interest}}^{\text{Item}(2)} = 1.53$, d.f. = 1384.1, $P = 0.125$, 95% CI -0.01, 0.08, $T_{\text{Pro-social vs. Self-interest}}^{\text{Item}(3)} = 0.99$, d.f. = 1389.6, $P = 0.322$, 95% CI -0.03, 0.08, $T_{\text{Pro-social vs. Self-interest}}^{\text{Item}(4)} = -12.40$, d.f. = 1386.8, $P < 0.001$, 95% CI -0.36, -0.27, $T_{\text{Pro-social vs. Self-interest}}^{\text{Item}(5)} = 2.41$, d.f. = 1386.8, $P = 0.016$, 95% CI 0.01, 0.08, $T_{\text{Self-interest vs. Control}}^{\text{Item}(1)} = 6.10$, d.f. = 1265.5, $P < 0.001$, 95% CI 0.10, 0.20, $T_{\text{Self-interest vs. Control}}^{\text{Item}(2)} = 6.50$, d.f. = 1280.9, $P < 0.001$, 95% CI 0.11, 0.21, $T_{\text{Self-interest vs. Control}}^{\text{Item}(3)} = 10.95$, d.f. = 1329.1, $P < 0.001$, 95% CI 0.23, 0.33, $T_{\text{Self-interest vs. Control}}^{\text{Item}(4)} = 18.94$, d.f. = 1299.1, $P < 0.001$, 95% CI 0.40, 0.50, $T_{\text{Self-interest vs. Control}}^{\text{Item}(5)} = -1.11$, d.f. = 1306.6, $P = 0.266$, 95% CI -0.05, 0.01.

Outcome	Pooled vs. control			Pro-social vs. control		
	B	95% CI	P	B	95% CI	P
Uptake (tracked)	0.01	[-0.02, 0.03]	0.533	0.01	[-0.03, 0.05]	0.576
Uptake (hybrid)	0.03	[0.00, 0.06]	0.039	0.02	[-0.01, 0.06]	0.177
Uptake (reported)	0.02	[-0.02, 0.06]	0.282	0.01	[-0.04, 0.06]	0.640
Knowledge	0.25	[0.16, 0.35]	<0.001	0.21	[0.10, 0.32]	<0.001
Positive attitudes	0.11	[0.05, 0.17]	0.001	0.07	[-0.00, 0.14]	0.054
Message sharing (tracked)	-0.05	[-0.14, 0.03]	0.222	-0.07	[-0.15, 0.02]	0.111
Information lookup (tracked)	0.07	[-0.04, 0.18]	0.208	0.10	[-0.04, 0.24]	0.164
Bluetooth active (reported)	0.07	[-0.03, 0.17]	0.169	0.13	[0.02, 0.24]	0.018
Test/quarantine if given risk warning (reported)	0.13	[-0.02, 0.28]	0.081	0.10	[-0.06, 0.26]	0.218
Notify app if tested positive (reported)	0.07	[-0.11, 0.25]	0.435	0.04	[-0.16, 0.25]	0.680

Outcome	Self-interest vs. control			Pro-social vs. self-interest		
	B	95% CI	P	B	95% CI	P
Uptake (tracked)	0.01	[-0.03, 0.04]	0.701	0.00	[-0.04, 0.05]	0.848
Uptake (hybrid)	0.04	[-0.00, 0.08]	0.060	-0.01	[-0.06, 0.03]	0.538
Uptake (reported)	0.04	[-0.02, 0.09]	0.174	-0.02	[-0.07, 0.03]	0.357
Knowledge	0.30	[0.19, 0.41]	<0.001	-0.07	[-0.18, 0.03]	0.178
Positive attitudes	0.14	[0.07, 0.21]	<0.001	-0.07	[-0.14, -0.00]	0.045
Message sharing (tracked)	-0.04	[-0.14, 0.06]	0.469	-0.03	[-0.10, 0.04]	0.391
Information lookup (tracked)	0.04	[-0.08, 0.16]	0.534	0.06	[-0.09, 0.21]	0.428
Bluetooth active (reported)	0.01	[-0.11, 0.13]	0.883	0.12	[0.02, 0.22]	0.022
Test/quarantine if given risk warning (reported)	0.16	[-0.00, 0.32]	0.057	-0.06	[-0.20, 0.09]	0.435
Notify app if tested positive (reported)	0.10	[-0.09, 0.29]	0.308	-0.06	[-0.23, 0.12]	0.520

Extended Data Fig. 2: Results of the main adjusted ITT models in terms of the effects of the message experiment. Standardized ITT estimates from fully saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. 95% confidence intervals and p-values from two-tailed tests reported. The video message sample comprises $N = 2,044/1,356/1,337$ respondents for the pooled/pro-social/self-interest treatment groups respectively.

Outcome	Pooled vs. control			EUR 1 vs. control		
	B	95% CI	P	B	95% CI	P
Uptake (tracked)	1.05	[0.71, 1.39]	<0.001	0.83	[0.27, 1.39]	0.004
Uptake (hybrid)	0.84	[0.66, 1.01]	<0.001	0.80	[0.52, 1.08]	<0.001
Uptake (reported)	0.84	[0.64, 1.03]	<0.001	0.89	[0.59, 1.18]	<0.001
Knowledge	0.16	[0.03, 0.28]	0.014	0.14	[-0.01, 0.30]	0.070
Positive attitudes	0.09	[-0.03, 0.21]	0.156	0.20	[0.05, 0.36]	0.011
Message sharing (tracked)	0.07	[0.01, 0.12]	0.019	0.05	[-0.02, 0.13]	0.157

Outcome	EUR 2 vs. control			EUR 5 vs. control		
	B	95% CI	P	B	95% CI	P
Uptake (tracked)	0.98	[0.47, 1.50]	<0.001	1.43	[0.73, 2.13]	<0.001
Uptake (hybrid)	0.66	[0.40, 0.93]	<0.001	1.09	[0.77, 1.41]	<0.001
Uptake (reported)	0.68	[0.41, 0.95]	<0.001	0.95	[0.65, 1.25]	<0.001
Knowledge	0.15	[-0.02, 0.31]	0.076	0.22	[0.07, 0.38]	0.005
Positive attitudes	-0.05	[-0.19, 0.10]	0.523	0.11	[-0.04, 0.25]	0.138
Message sharing (tracked)	0.08	[-0.04, 0.20]	0.180	0.06	[-0.02, 0.14]	0.157

Outcome	EUR 1 vs. EUR 2			EUR 1 vs. EUR 5		
	B	95% CI	P	B	95% CI	P
Uptake (tracked)	0.13	[-0.58, 0.84]	0.725	0.54	[-0.28, 1.37]	0.197
Uptake (hybrid)	-0.14	[-0.51, 0.22]	0.443	0.26	[-0.15, 0.66]	0.210
Uptake (reported)	-0.21	[-0.56, 0.15]	0.252	0.07	[-0.31, 0.44]	0.721
Knowledge	-0.04	[-0.20, 0.11]	0.589	0.09	[-0.07, 0.24]	0.277
Positive attitudes	-0.25	[-0.40, -0.10]	0.001	-0.09	[-0.24, 0.06]	0.263
Message sharing (tracked)	0.03	[-0.11, 0.17]	0.713	0.00	[-0.11, 0.12]	0.939

Outcome	EUR 2 vs. EUR 5		
	B	95% CI	P
Uptake (tracked)	0.52	[-0.39, 1.42]	0.264
Uptake (hybrid)	0.41	[0.01, 0.81]	0.047
Uptake (reported)	0.27	[-0.09, 0.64]	0.136
Knowledge	0.12	[-0.04, 0.27]	0.133
Positive attitudes	0.15	[0.02, 0.29]	0.026
Message sharing (tracked)	-0.02	[-0.17, 0.12]	0.763

Extended Data Fig. 3: Results of the main adjusted ITT models in terms of the effects of the incentive experiment. Standardized ITT estimates from saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. 95% confidence intervals and p-values from two-tailed tests reported. The incentive sample comprises $N = 1,015/513/516/494$ respondents for the pooled/EUR 1/EUR 2/EUR 5 treatment groups respectively.

Extended Data Fig. 4: Agreement rates of app installation incentivization by monetary incentive. ($N_{\text{EUR } 1} = 259$, $N_{\text{EUR } 2} = 262$, $N_{\text{EUR } 5} = 240$). Two-sample t-test statistics: $T_{\text{EUR } 1 \text{ vs. EUR } 2} = 0.56$, d.f. = 518.83, $P = 0.579$, 95% CI $-0.06, 0.11$, $T_{\text{EUR } 1 \text{ vs. EUR } 5} = -1.34$, d.f. = 493.62, $P = 0.180$, 95% CI $-0.15, 0.03$, $T_{\text{EUR } 2 \text{ vs. EUR } 5} = -1.90$, d.f. = 494.95, $P = 0.059$, 95% CI $-0.17, 0.01$.

Outcome	Pro-social message X Social responsibility score			Self-interest message X Self-interest score		
	B	95% CI	P	B	95% CI	P
Uptake (tracked)	0.01	[-0.04, 0.05]	0.720	-0.01	[-0.06, 0.03]	0.541
Uptake (hybrid)	-0.02	[-0.07, 0.02]	0.360	0.01	[-0.04, 0.05]	0.750
Uptake (reported)	-0.01	[-0.07, 0.04]	0.660	0.02	[-0.03, 0.07]	0.448
Knowledge	-0.05	[-0.16, 0.07]	0.440	0.02	[-0.09, 0.14]	0.684
Positive attitudes	-0.08	[-0.14, -0.01]	0.030	0.04	[-0.04, 0.11]	0.343
Message sharing (tracked)	-0.03	[-0.18, 0.13]	0.720	0.00	[-0.10, 0.11]	0.960
Information lookup (tracked)	0.03	[-0.10, 0.15]	0.670	-0.03	[-0.19, 0.12]	0.666
Bluetooth active (reported)	0.05	[-0.06, 0.16]	0.380	0.10	[-0.01, 0.20]	0.066
Tested after warning	-0.07	[-0.30, 0.16]	0.550	-0.01	[-0.20, 0.19]	0.954
Reporting positive test	0.01	[-0.18, 0.20]	0.920	-0.02	[-0.23, 0.18]	0.816

Extended Data Fig. 5: Results of pre-registered effect heterogeneity tests for the message treatments. Standardized ITT estimates from fully saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. 95% confidence intervals and p-values from two-tailed tests reported. Effects reported with regards to the interaction effect between treatment indicator and moderator. Coefficients are computed in the covariate-adjusted ITT model. The pre-registered hypotheses were as follows: (1) Subjects with high levels of social responsibility will show higher treatment effects in the pro-social treatment than subjects with low levels of social responsibility. (2): Subjects with high levels of self-interest will show higher treatment effects in the self-interest treatment than subjects with low levels of self-interest. The video message sample comprises $N = 2,044/1,356/1,337$ respondents for the pooled/pro-social/self-interest treatment groups respectively.

Tracking and Promoting Usage of a COVID-19 Contact Tracing App

Supplementary Information

Contents

1	Supplementary Discussion	3
1.1	Background: The German COVID-19 Contact Tracing App	3
2	Supplementary Methods	6
2.1	Panel characteristics	6
2.1.1	Sample differences	6
2.1.2	Panel attrition	8
2.1.3	Selection bias	8
2.1.4	Tracked device type	10
2.2	Message experiment setup and implementation	15
2.2.1	Treatment delivery	15
2.2.2	Covariate balance	15
2.2.3	Compliance	15
2.2.4	Manipulation check	19
2.3	Incentivization experiment setup and implementation	21
2.3.1	Treatment delivery	21
2.3.2	Covariate balance	22
2.3.3	Compliance	23
2.4	Variable specifications	28
2.4.1	App uptake	28
2.4.2	Probability to install app	29
2.4.3	Bluetooth usage	29
2.4.4	App knowledge	29
2.4.5	App attitudes	29
2.4.6	App message sharing behavior and app information lookup	30
2.4.7	Hypothetical app usage	30
2.4.8	Covariates	31
2.5	Survey Questionnaire	35
2.5.1	Sociodemographics	35
2.5.2	Attitudes and opinions	39
2.5.3	Risk status and behavior	44
2.5.4	Tracing app usage and attitudes	47
2.6	Deviations with respect to pre-analysis plan	59

3	Supplementary Results	61
3.1	App usage	63
3.2	Message experiment	70
3.2.1	Replicated main results with W3 outcomes & conditional on passing the manipulation checks	70
3.2.2	Hypothesis tests (pre-registered)	73
3.2.3	Hypothesis tests (not pre-registered)	82
3.2.4	Effect heterogeneity	85
3.3	Incentivization experiment	86
3.3.1	Hypothesis tests (pre-registered)	86
3.3.2	Effect heterogeneity	91

1 Supplementary Discussion

1.1 Background: The German COVID-19 Contact Tracing App

The Corona-Warn-App (original brand name) was developed by SAP and Deutsche Telekom on behalf of the German government. It uses Bluetooth low energy technology to exchange tokens with app users that are within 1.5 meters of each other for a period of at least 10 minutes. Any app user who tests positive for COVID-19 can share this information voluntarily with the app. Other app users are then notified that they have had contact with an infected person within a 14 day period, and whether that contact is rated as high or low risk. The calculation is performed locally on the user's device using the exposure logging data, and takes into account the duration and distance of encounters with individuals who were tested positive for COVID-19. Testing is available for users on a voluntary basis. About half of all testing labs can directly share the test results with App users (1).

The app is intended to supplement human contact tracing efforts. It was introduced on June 16, 2020, and has registered more than 18 million downloads in the first 100 days since its publication. As of 22 September 2020, the app has been downloaded 18.6 million times (2). During this time period, the Robert Koch Institute reported that more than 4,300 individuals have received codes to share their positive test results with the app (2).

While no officially comparable statistics have been published, the uptake of the Corona-Warn-App in Germany has been relatively high when compared to the country's population size

Supplementary Figure 1: Downloads of the German “Corona-Warn-App” (Sources: (3)).

(approx 83m). For example, India's (approx 1.3b) contact tracing app was downloaded more than 100 million times in the first 40 days (4). In France (approx 66m inhabitants) the official StopCovid app was released on June 2, 2020 and was downloaded 2.3 million times in the first three months (5), a considerably lower number than in Germany. Supplementary Figure 2 provides an overview of COVID-19 contact tracing app implementations worldwide, as of November 2020 (6, 7). Supplementary Figure 3 provides estimates of the total number of users and the user/population ratio of COVID-19 contact tracing apps as reported by the MIT Covid Tracing Tracker project (6). Most user numbers reflect download statistics, which likely provide an inflated estimate of the total number of users. Furthermore, as discussed in the paper the population figure does not represent the optimal baseline to estimate penetration rates because it is uninformative about actual smartphone ownership and usage.

Eleven EU countries are currently planning to or have facilitated data transfer between their official contact tracing apps. These include Germany, Italy, Ireland, Austria, Czech Republic, Denmark, Estonia, Latvia, the Netherlands, Poland, and Spain (8). Users will have to update their app and agree to the EU data exchange. The data transfer will only include anonymized data, which means that no inferences will be possible regarding individuals. Thus far, there is little evidence that private businesses are requesting customers to install the contact tracing app before entering their premises. One such case made headlines in Germany, when the owners of a small campground made the installation of the app obligatory for their guests (9).

Supplementary Figure 2: COVID-19 contact tracing app implementations worldwide, as of November 2020.

Supplementary Figure 3: Estimated total number of users and user/population ratio of COVID-19 contact tracing apps, by country and app. The Norwegian app Smittestopp was discontinued in September 2020.).

2 Supplementary Methods

2.1 Panel characteristics

Supplementary Table 1 reports sample means (as of wave 1) on the covariates that were used as quota criteria. We used the 2019 Best for Planning (b4p '19) study (10) to identify target quota. The Best for Planning study is a face-to-face survey targeting Germany's German-speaking population aged 14 or older. It is based on more than 30k respondents who were sampled on the basis of a multi-stage probability sample. The survey-only sample exhibits some minor differences with regards to education (the lowest category is slightly underrepresented, while the middle category is slightly overrepresented). The discrepancies are noticeably larger for the survey-plus-tracking sample. Here, males are slightly overrepresented, as are participants in the age groups 30-39, 40-49, and 50-59 year olds. People aged 60 or older are underrepresented. Furthermore, the sample is biased towards people with a higher educational background, with the two upper educational strata being overrepresented and the lowest category underrepresented.

Supplementary Table 1: Descriptive statistics of respondent characteristics (quotas), by sample

Respondent characteristic	b4p '19	Sample		
		Pooled	Survey-only	Survey-tracking
Gender				
—Male	49	49	47	53
—Female	51	51	52	47
Age				
—18-29	18	18	15	16
—30-39	16	19	18	20
—40-49	17	20	18	20
—50-59	21	22	21	27
—60+	28	21	28	17
Education				
—Lower secondary school	33	22	26	16
—Middle school	32	38	37	40
—High school	35	40	37	44

2.1.1 Sample differences

Next, we explore potential differences between the various samples. Apart from potential sociodemographic discrepancies, it seems plausible that respondents who are willing to provide data to a polling firm via passive tracking software are less concerned about data and privacy issues than the population as a whole and more data-literate. As some of the main reasons for

Supplementary Table 2: Model of selection into tracking panel and into survey among tracking panelists

	Tracking panel	Tracking survey
Intercept	0.57 (0.14)	0.39 (0.05)
Age: 30-39yrs	0.00 (0.04)	0.13 (0.04)
Age: 40-49yrs	-0.01 (0.04)	0.14 (0.04)
Age: 50-59yrs	0.05 (0.04)	0.26 (0.04)
Age: 60+yrs	-0.09 (0.04)	0.35 (0.05)
Gender: Female	-0.02 (0.02)	0.02 (0.03)
Education: Medium	0.09 (0.03)	0.01 (0.04)
Education: High	0.09 (0.03)	0.04 (0.04)
Control access to personal information	0.00 (0.01)	
Government protects my privacy	0.00 (0.01)	
Concerned about companies accessing data	-0.04 (0.01)	
Data Literacy	0.04 (0.01)	
R ²	0.04	0.05
Adj. R ²	0.03	0.05
Num. obs.	1875	1122

citizens not to install the app might include concern about privacy issues and data literacy, this makes self-selection into the tracking sample on the basis of these traits an important feature of the data. To measure these traits, we use a battery of questions that was included in wave 1 which applied to both panels. The items read:

- (A) It is important for me to control who is able to store and access information collected about me.
- (B) The government protects my privacy.
- (C) I am concerned about how many companies can access my personal without my knowledge.

The left column in Supplementary Table 2 reports estimates of a linear model of tracking panel membership. While there are no meaningful differences between the first two items, members of the tracking panel are significantly less concerned about companies accessing their data than members of the survey-only panel (controlling for a set of sociodemographic variables). Moreover, they exhibit higher levels of self-reported digital literacy, which we measured using another item battery following (11).

The right column in Supplementary Table 2 reports estimates of a linear model of survey-taking among tracking panelists. We find the higher age tiers to be over-represented in the survey sample compared to the tracking-only sample, which is likely a consequence of the target quotas in combination with the fact that the tracking panel has a much younger age structure. Older respondents in that panel were therefore much more likely to enter the survey in order to approximate the target quotas.

2.1.2 Panel attrition

We further consider panel attrition. Supplementary Table 3 provides aggregate attrition figures across all three waves. Attrition remained relatively low, likely due to the brief gaps between waves, with 13% of respondents dropping out after wave 1 and another 11% after wave 2. Respondents on the survey-plus-tracking panel remained somewhat more devoted to the panel.

Supplementary Table 4 explores potential predictors of attrition across waves using indicators of app usage and attitudes, treatment status in the experiments, and the set of sociodemographic factors. We find that age is relevant – older respondents were less likely to drop out than younger respondents. App and treatment status-related factors do not have a statistically significant effect on dropout.

Supplementary Table 3: Panel attrition across samples

Sample		wave 1	wave 2	wave 3
Pooled	N	2,044	1,777	1,569
	%	100	87	77
Survey-only	N	1,395	1,190	1,048
	%	100	85	75
Survey-tracking	N	649	587	521
	%	100	90	80

2.1.3 Selection bias

The analysis presented in Fig. 2 suggests that App uptake is more prevalent in older than in younger age groups. This result contradicts findings from earlier survey studies which report a *negative* relationship between age and intentions to use DCT apps (12–14). Differences between intentions and actual behavior aside, we need to carefully consider how sample selection might possibly affect central findings. This is all the more important as our respondents were recruited from an online access panel which is generally known for undercoverage of older cohorts (15). To illustrate, consider the graphical model of the sample selection process in Fig- 4.

Supplementary Table 4: Probit models of panel attrition. Outcome is coded 1 if respondent drops out before respective wave. The model reported in the last column is restricted to participants who report of not having installed the app in Wave 2 and therefore qualify for one of the treatment groups in the incentivization experiment.

	W2 dropout	W3 dropout	W3 dropout, app non-users
Intercept	0.16 (0.23)	−0.22 (0.25)	−0.12 (0.33)
Tracking sample	−0.26 (0.09)	−0.09 (0.09)	−0.16 (0.13)
Attitudes towards app	−0.01 (0.05)	−0.01 (0.06)	−0.09 (0.08)
App knowledge	−0.09 (0.04)	−0.03 (0.04)	0.00 (0.05)
App installed	0.08 (0.10)	0.09 (0.11)	
Message treatment W1	0.11 (0.09)	0.04 (0.10)	0.08 (0.13)
Incentivization treatment W2			0.11 (0.14)
Incentivization agreement W2			−0.20 (0.14)
Age/10	−0.24 (0.03)	−0.18 (0.03)	−0.22 (0.05)
Gender: Female	0.03 (0.08)	0.08 (0.09)	−0.13 (0.12)
Education: Medium	0.14 (0.11)	0.03 (0.12)	0.16 (0.15)
Education: High	0.08 (0.11)	0.01 (0.12)	0.01 (0.16)
HH income: Medium	−0.09 (0.10)	−0.11 (0.11)	−0.26 (0.14)
HH income: High	−0.16 (0.11)	0.12 (0.11)	0.04 (0.15)
Children	0.09 (0.09)	0.07 (0.09)	0.05 (0.12)
Trust in government	−0.09 (0.05)	−0.04 (0.05)	0.05 (0.07)
Trust in scientists	−0.01 (0.05)	0.04 (0.06)	0.04 (0.07)
Trust in HC system	−0.01 (0.05)	−0.07 (0.06)	−0.12 (0.07)
COVID threat: Self	0.07 (0.07)	−0.06 (0.07)	−0.07 (0.09)
COVID threat: Family/Friends	0.01 (0.07)	0.03 (0.07)	0.09 (0.09)
Social responsibility	−0.00 (0.04)	0.06 (0.05)	0.07 (0.06)
Self interest	−0.04 (0.04)	−0.03 (0.05)	0.00 (0.06)
Data control	−0.12 (0.05)	−0.01 (0.05)	−0.00 (0.07)
Digital literacy	0.05 (0.04)	−0.02 (0.04)	−0.08 (0.06)
AIC	1342.96	1169.09	704.91
BIC	1464.61	1287.80	815.97
Log Likelihood	−649.48	−562.55	−329.46
Deviance	1298.96	1125.09	658.91
Num. obs.	1862	1629	924

Here an individual's Age (presumably negatively) affects his digital literacy (DL), which in turn positively affects his propensity to be selected into the access panel, and, ultimately, into our survey (S). App usage affects (or is, through topic interest or salience, confounded with) selection into the survey. If we condition on S (depicted by the box in the intermediate panel), that is, look only into the survey sample, this may distort the population association between Age and App (depicted by the dashed arc) – a phenomenon known as collider bias in graph theory (16, 17). In this simplistic model, the problem can be solved by also conditioning on DL (right panel), and looking at the in-sample associations between Age and App usage by level of digital literacy. This is done in Supplementary Table 5. As one can see, the positive association between Age and App usage remains even after conditioning on Digital Literacy. Therefore, we are more confident that the sample association properly reflects the population association.

The App adoption rates over time provided in Fig. 4 suggest an increasing discrepancy between the survey-tracking and the tracking-only group. Parts of these differences were apparent before the initial survey wave. At the end of the study period, roughly 50% of the survey-tracking subjects but only 37% of the tracking-only subjects had adopted the App. This seems to have at least three drivers: the first is disproportionate self-selection from the tracking panel into the first wave of the panel survey starting on July 30, where we observe 40% versus 35% of App users in the survey-tracking and the tracking-only group. Topic interest and salience are well-known drivers of survey participation (18, 19), and may account for the increased participation of App users among tracking panelists in our survey. Respondents without tracking in our sample are potentially more comparable to the target population. Only a fraction of our survey respondents opt for membership in the mobile tracking panel (see Fig. 1. Additional analyses (see Section 2.1) show that mobile tracking panelists among survey respondents are, on average, more educated and digitally literate, but less concerned about data privacy. Both of these features are positively linked with App use. The second component is the treatment effect of the incentivization experiment conducted in wave 2 among self-reported non-users of the App (blue curves). The traced usage of the App increases from 8% to 20% among self-reported non-users who received incentives to install the app, which results in a total increase for both groups of survey-tracking users. A third and minor component may have been panel conditioning or placebo effects, that is, positive effects on App uptake of mere study participation, as indicated by the modest rise of App use from 9 to 11% among control group members during wave 2 (dashed blue curve). Note that the control group of the message experiment in wave 1 (yellow dashed curve) includes subjects which receive the incentivization treatment in wave 2. Therefore, the observed increase in App use during wave 2 is at least partly a treatment, not a conditioning or placebo effect.

2.1.4 Tracked device type

Due to a technical mistake by the provider, tracking data on CWA use were only available for users of mobile devices with Android, but not Apple, operating systems. Consequently, we excluded respondents from the survey-tracking and tracking-only panel that had an Apple

Supplementary Figure 4: Graphical causal model of a selection process that biases the population relationship between Age and App usage. DL is data literacy, S is sample selection.

Supplementary Table 5: Linear model of tracked app usage as a function of age and digital literacy.

	Tracked uptake
Intercept	0.31 (0.08)
Age: 30-39yrs	−0.08 (0.11)
Age: 40-49yrs	0.02 (0.11)
Age: 50-59yrs	0.12 (0.10)
Age: 60+yrs	0.17 (0.11)
Digital literacy: High	0.13 (0.11)
Age: 30-39yrs X Digital literacy: High	0.08 (0.14)
Age: 40-49yrs X Digital literacy: High	−0.09 (0.14)
Age: 50-59yrs X Digital literacy: High	−0.01 (0.13)
Age: 60+yrs X Digital literacy: High	−0.06 (0.14)
R ²	0.04
Adj. R ²	0.02
Num. obs.	586

device registered ($n = 129$ in total; already subtracted from the figures reported in the main text).

As soon as we had noticed the technical issue, we informed the data provider, who was then able to quickly set up tracking of CWA use on Apple devices. Unfortunately, this happened just after the tracking period when we were running our main analyses. However, we were granted access to about one additional month of tracking data. This section reports results from additional analyses using these and our existing data.

First, we consider pre-subject uptake and average differences on various variables by device type used by the panelists. This is done for tracking-survey panelists only, as those are the

subjects for which device type information is available. Reported pre-study uptake of the CWA share is 49% among Android device users and 68% among Apple device users. Next, Supplementary Table 6 provides mean differences on various variables by device type. In comparison to Android users in our sample, Apple users are significantly younger, have higher levels of education, lower prevalence of preconditions, and higher digital literacy. Interestingly, the care for privacy is not among the key differences. Apart from age, these factors are positively related to app uptake.

Supplementary Table 6: Covariate balance: Apple vs. Android device (means reported)

Covariate	Apple	Android	Diff. (Apple-Android)	p value
Age/10	3.87	4.26	-0.39	0.036
Gender: Female	0.51	0.46	0.05	0.485
Education: Medium	0.25	0.40	-0.14	0.022
Education: High	0.67	0.44	0.23	0.001
Household income: Medium	0.42	0.38	0.04	0.563
Household income: High	0.44	0.33	0.10	0.143
Children	0.44	0.49	-0.06	0.400
Public transport usage	0.16	0.21	-0.05	0.336
Restaurant/pub visit	0.25	0.14	0.11	0.066
Friend/family visit	0.60	0.47	0.13	0.069
NPI compliance	0.69	0.76	-0.07	0.311
Precondition	0.23	0.43	-0.20	0.001
COVID case in network	0.30	0.18	0.12	0.068
Lives in high COVID incidence region	0.09	0.06	0.03	0.505
Lives in urban region	0.51	0.43	0.08	0.282
Working	0.78	0.74	0.04	0.487
Trust in government	2.44	2.04	0.39	0.016
Trust in scientists	2.76	2.57	0.20	0.137
Trust in HC system	2.85	2.53	0.32	0.008
COVID threat perception: Self	1.33	1.45	-0.12	0.279
COVID threat perception: Family and Friends	1.75	1.75	-0.01	0.955
Social responsibility	-0.00	-0.00	0.00	0.996
Self interest	0.28	0.07	0.21	0.135
Data control	-0.10	-0.05	-0.05	0.712
Digital literacy	0.64	0.19	0.45	<0.001

Second, we used the additional tracking data in October to construct a new outcome of interest, tracked app usage in October, which could be used to compare Apple and Android users. Supplementary Figure 5 provides app adoption rates by device and tracking sample type for October. The final tracked average usage rates across sample tracking types on October 31st are 28% for Android users and 43% for Apple users (we use only the October data to generate the binary indicator ‘app used’ (yes/no) to make the figures comparable, which is why they are lower than the final uptake rates as reported for the end of the original study period).

Third, we estimated the ITT of the pooled incentivization treatments in W2 on October uptake separately for Android and Apple users. The results are reported in Supplementary

Table 21. The reported standardized effects translate into, and found a 19 percentage point effect for Apple users and 8 percentage point effect for Android users. Note that the effect estimated for the subset of Apple users comes with considerable uncertainty due to the small number of observations in the sample.

Overall, these additional analyses indicate the following: (1) Apple device usage and pre-study uptake of the CWA share common predictors, including age, education, lower prevalence of preconditions, and higher digital literacy. (2) Apple users have higher baseline adoption rates of the app. This is consistent with the published aggregate download figures by the RKI, according to which the app had been downloaded 12.6m times from the Google Play Store and 10.6m times from the Apple App Store. for comparison, the market share of Apple devices in Q2 2020 was 23% (see <https://www.counterpointresearch.com/germany-smartphone-sales-plunge-q2-2020/>). (3) Apple users seem to have been more receptive towards the incentivization treatments. This indicates that, had App usage of Apple users been properly tracked over the entire study period, the observed effects would have been somewhat more pronounced.

Supplementary Figure 5: App adoption rates by device and tracking sample type in October 2020. The figure shows adoption rates for iPhone-using survey tracking (red solid line) and tracking only respondents (yellow dotted line) and for Android phone users in the survey tracking (blue solid line) and tracking only groups (blue dotted line), each with 83% confidence intervals. Non-overlap indicates a significant difference at $p < 0.05$.

Supplementary Table 7: ITT estimates of pooled incentivization treatments (EUR 1, 2, 5) in W2 on tracked October uptake in survey-tracking sample. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	MDE
Uptake (tracked) (Pooled vs. control)	0.210 (0.092) [0.028, 0.392]	0.210 (0.092) [0.028, 0.392]	0.186 ($d=0.220$)
Uptake (tracked) (Pooled vs. control, Apple users)	0.454 (0.244) [-0.066, 0.974]	0.454 (0.244) [-0.059, 0.967]	0.807 ($d=0.909$)
Uptake (tracked) (Pooled vs. control, Android users)	0.196 (0.096) [0.006, 0.386]	0.196 (0.096) [0.006, 0.386]	0.192 ($d=0.227$)

2.2 Message experiment setup and implementation

2.2.1 Treatment delivery

This experiment was embedded at the end of the wave 1 questionnaire. Supplementary Table 8 provides the introductory text. Supplementary Table 9 provides the full transcription of the video message together with the distinct text treatment blocks at the end. Supplementary Figure 6 provides screenshots from the pro-social and self-interest content in the videos. The videos themselves can be watched at https://youtu.be/dyhDd_vrGEE (pro-social message) and at https://youtu.be/suOCvIW8_R0 (self-interest message). The first part of the message is based on the explainer video “Solltet Ihr die Corona-App installieren? Alle Infos zu Funktion & Datenschutz der Corona Warn-App” created by journalist Simon Lohmann and published on YouTube by *PC Welt* (the German edition of the *PC World*) on June 16, 2020 at <https://www.youtube.com/watch?v=I3C9BrC9I-8>. The original video was much longer than our version (8.5 minutes).

Supplementary Table 8: Message treatment: introduction

German (original)	English translation
Bitte schauen Sie sich das folgende Video (etwa 2 Minuten Länge) aufmerksam an. Bitte spulen Sie nicht vor oder überspringen Teile des Videos. Wir möchten Ihnen im Anschluss gerne einige kurze Fragen dazu stellen. [embedded video]	Please watch the following video (about 2 minutes length) carefully. Please do not fast forward or skip parts of the video. We would like to ask you some short questions about it afterwards. [embedded video]

2.2.2 Covariate balance

Tables 10 reports covariate balance between treatment and control groups. Each row reports group-specific variable means and associated two-sided p -values for treatment-control differences.

For both the pro-social and the self-interest message treatment groups vs. control, two differences are statistically significant at the 95% level: working status and digital literacy. Looking at the total number of comparisons across both treatments, this is within the range of what might be expected due to chance. We address potential confounding issues by including these variables for possible selection in all covariate-adjusted models.

2.2.3 Compliance

In this section, we provide evidence to which extent study participants complied with the different treatments. In particular, we address two questions: (1) what proportion of subjects

(a) Pro-social I

Bitte schauen Sie sich das folgende Video (etwa 2 Minuten Länge) aufmerksam an.
Bitte spulen Sie nicht vor oder überspringen Teile des Videos.
Wir möchten Ihnen im Anschluss gerne einige kurze Fragen dazu stellen.

(c) Self-interest I

Bitte schauen Sie sich das folgende Video (etwa 2 Minuten Länge) aufmerksam an.
Bitte spulen Sie nicht vor oder überspringen Teile des Videos.
Wir möchten Ihnen im Anschluss gerne einige kurze Fragen dazu stellen.

(b) Pro-social II

Bitte schauen Sie sich das folgende Video (etwa 2 Minuten Länge) aufmerksam an.
Bitte spulen Sie nicht vor oder überspringen Teile des Videos.
Wir möchten Ihnen im Anschluss gerne einige kurze Fragen dazu stellen.

(d) Self-interest II

Bitte schauen Sie sich das folgende Video (etwa 2 Minuten Länge) aufmerksam an.
Bitte spulen Sie nicht vor oder überspringen Teile des Videos.
Wir möchten Ihnen im Anschluss gerne einige kurze Fragen dazu stellen.

Supplementary Figure 6: Screenshots of pro-social and self-interest messages in video treatment

Supplementary Table 9: Message treatment: transcription

German (original)	English translation
<p>[Wie funktioniert die App?] Also wenn ich mir die App heruntergeladen habe und einer Person begegne, die die App auch auf ihrem Handy hat, dann tauschen die Smartphones einen Zufallscode aus. Dabei handelt es sich um eine geheime Zahlenkombination, und das passiert jedes Mal, wenn ich anderen Nutzern der App begegne. Ja, und nur dieser Zufallscode wird auf dem Smartphone gespeichert. Und da wären wir schon bei einem wichtigen Punkt: Die Daten werden also nicht in eine Cloud oder einen Server irgendwo hochgeladen, wo ich gar nicht weiß, wo die Daten jetzt überhaupt landen, und die dann auch gehackt werden könnten, sondern die Daten werden dezentral auf Eurem Smartphone gespeichert. Was wird überhaupt gespeichert? Nun, im Prinzip eigentlich nur drei Dinge: Nämlich erstens, dass Ihr einer anderen Person begegnet seid, unabhängig jetzt davon ob die infiziert oder gesund war. Dann zweitens, welchen Abstand Ihr zu dieser Person hattet. Und drittens, wie lange die Begegnung gedauert hat. Die App weiß jetzt also nicht, dass ich jetzt beispielsweise Dennis, Takis, Michi oder Basti über den Weg gelaufen bin und vor allem auch nicht wo das dann war.</p> <p>[Was ist mit dem Datenschutz?] Ich habe es im Video ja schon mehrfach angesprochen: Datenschutz wird bei der App wirklich großgeschrieben. Alle Daten der Nutzer werden dezentral und anonymisiert gesammelt. Es werden auch keinerlei Standorte des Nutzers erfasst und gespeichert. Im Vorfeld gab es zwar Kritik an der Sicherheit der App unter anderem vom TÜV IT, aber diese Schwachstellen wurden geschlossen. Mittlerweile wird die App sogar von Datenschützern gelobt. Und wer wirklich Angst um seine Daten hat, der sollte vielleicht einmal nachschauen, welche Apps er denn sonst noch so auf seinem Smartphone hat.</p>	<p>[How does the app work?] So if I have downloaded the app and I meet a person who also has the app on their phone, the smartphones exchange a random code. This is a secret combination of numbers, and it occurs every time I meet other users of the app. Yes, and only this random code is stored on the smartphone. And that brings us to an important point: The data is not uploaded to a cloud or a server somewhere where I don't even know where the data is going to end up and could be hacked, but the data is stored locally on your smartphone. What is stored then? Well, in principle only three things: Firstly, that you met another person, regardless of whether they were infected or healthy. Then second, the distance you had to this person. And third, how long the encounter lasted. So the app doesn't know that I ran into Dennis, Takis, Michi or Basti, for example, and above all, it doesn't know where they were.</p>
<p>[Bringt die App jetzt noch was?] [Pro-social treatment (text only)] Gerade jetzt, wo es immer mehr Lockerungen gibt, ist die App so wichtig, weil der Kontakt mit anderen Menschen wieder deutlich zunimmt. Mit Nutzung der App tragen Sie aktiv zur Eindämmung der Pandemie bei. Dadurch können Sie Ihrer Familie, Freunden, und letztlich auch den schwächsten Mitmenschen helfen. Zum Beispiel sind Ältere und Menschen mit Vorerkrankungen besonders gefährdet und können so gezielter geschützt werden. Das ist ein kleiner Schritt für Sie, aber ein großer Beitrag für Ihre Mitmenschen. Vielen Dank!</p> <p>[Self-interest treatment (text only)] Gerade jetzt, wo es immer mehr Lockerungen gibt, ist die App so wichtig, weil der Kontakt mit anderen Menschen wieder deutlich zunimmt. Mit Nutzung der App tragen Sie aktiv zur Aufhebung der Einschränkungen bei. Dadurch könnten Sie schon bald wieder Ihrem fast normalen Alltag nachgehen. Zum Beispiel wieder uneingeschränkt verreisen, sich mit Freunden treffen, Sportevents, Kneipen oder Feste besuchen. Das ist ein kleiner Schritt mit potentiell großem Nutzen für Sie selbst. Vielen Dank!</p>	<p>[What about data privacy?] I have already mentioned it several times in the video: Data protection is really important to the app. All user data is collected locally and anonymously. No user locations are recorded and stored. In the run-up to the project, TÜV IT, among others, criticized the app's security, but these weaknesses have been removed. In the meantime, the app is even praised by data protectionists. And if you're really worried about your data, you might want to take a look at what other apps you have on your smartphone.</p> <p>[Is the app still of any use?] [Pro-social treatment (text only)] Especially now the restrictions are increasingly being relaxed, the app is very important because contact with other people is increasing again. By using it, you are actively helping contain the pandemic. In this way, you can help your family, friends, and ultimately also the weakest members of society. For example, the elderly and people with pre-existing conditions are particularly at risk and can therefore be protected in a more targeted manner. This is a small step for you, but a great contribution for your fellow human beings. Many thanks!</p> <p>[Self-interest treatment (text only)] Especially now the restrictions are increasingly being relaxed, the app is very important because contact with other people is increasing again. By using it you are actively contributing to the removal of restrictions. As a result, you could soon resume your almost normal everyday life. For example, travelling again without restriction, meeting friends, attending sporting events, pubs or parties. This is a small step with potentially great benefits for you. Thanks a lot!</p>

complied with the respective treatment? and (2) which subject characteristics predict compliance?

First, it is important to note that, as the control group is restricted access to the treatment, non-compliance is only possible in the treatment group and reduces to never-taking behavior.

Supplementary Table 10: Covariate balance: Message treatments vs. control (means and p values of treatment-control differences reported).

Covariate	Control	Pro-Social (p val.)	Self-Interest (p.val)
Age/10	4.73	4.74 (0.926)	4.74 (0.919)
Gender: Female	0.50	0.52 (0.649)	0.50 (0.972)
Education: Medium	0.36	0.39 (0.172)	0.38 (0.490)
Education: High	0.40	0.39 (0.746)	0.38 (0.407)
Household income: Medium	0.35	0.36 (0.507)	0.38 (0.167)
Household income: High	0.32	0.34 (0.354)	0.30 (0.431)
Children	0.56	0.56 (0.805)	0.56 (0.922)
Public transport usage	0.20	0.18 (0.482)	0.18 (0.497)
Restaurant/pub visit	0.15	0.16 (0.515)	0.15 (0.940)
Friend/family visit	0.48	0.49 (0.689)	0.46 (0.407)
NPI compliance	0.77	0.75 (0.277)	0.78 (0.677)
Precondition	0.42	0.42 (0.994)	0.42 (0.907)
COVID case in network	0.21	0.20 (0.509)	0.19 (0.459)
Lives in high COVID incidence region	0.07	0.07 (0.924)	0.07 (0.933)
Lives in urban region	0.39	0.39 (0.939)	0.40 (0.593)
Working	0.60	0.63 (0.241)	0.63 (0.298)
Trust in government	1.97	1.97 (0.928)	2.02 (0.405)
Trust in scientists	2.46	2.49 (0.655)	2.54 (0.153)
Trust in HC system	2.39	2.42 (0.582)	2.51 (0.028)
COVID threat perception: Self	1.47	1.45 (0.684)	1.44 (0.454)
COVID threat perception: Family and Friends	1.76	1.73 (0.509)	1.78 (0.676)
Social responsibility	-0.03	0.02 (0.356)	0.01 (0.409)
Self interest	-0.02	-0.01 (0.807)	0.03 (0.323)
Data control	-0.01	-0.02 (0.835)	0.03 (0.513)
Digital literacy	0.11	-0.01 (0.043)	-0.03 (0.021)

Our pre-registered definition of compliance takes as “treated” any respondent who spent at least the length of the treatment video in seconds (which equals 125 seconds for both videos) on the treatment exposure page. According to this definition, the compliance rate was 85% for the pro-social video message treatment and 86% for the self-interest video message treatment.

Next, we explored how individual-level characteristics relate to compliance with the treatment. Supplementary Figure 7 reports point estimates along with 83%/95% bootstrap confidence intervals on various covariates for the entire sample, compliers, and never-takers (note that always-takers are logically ruled out). We hardly find any differences between the two treatment conditions, which is plausible since they were similar except for the last segment in the video. Variables that predict never-taking, i.e. not watching the video in full despite being instructed to do so, include: age (being younger → less likely to comply), education (high levels of education → more likely to comply), having no precondition, trusting scientists less, being less digitally literate, being less concerned about COVID-19 for themselves, family, or friends, and having less positive pre-treatment attitudes about the app.

One implication of these findings is that information treatments of this type might be par-

ticularly difficult to deliver to those who could potentially benefit more.

2.2.4 Manipulation check

Following our pre-analysis plan, we use a battery of knowledge items asked directly after the video message in wave 1 as the basis for our manipulation check (see Section 2.5 for the full question wording of the item “Knowledge about app”). The item order was randomized. Note that this battery was replicated in wave 2 (providing the basis for the knowledge outcome in the main analysis) and wave 3. Because the items were implemented post-treatment already in wave 1, they are not used as pre-treatment variables in the main analyses.

Here, the battery serves as manipulation check because it reflects some of the key messages of the video treatment. The results are reported in Extended Data in the paper. In the following, we discuss the items in more detail:

- Two items (1) + (3) capture knowledge on privacy features, which respondents were briefed about in the treatment condition. The proportion of correct answers is significantly higher in both treatment groups vs. control group, while neither treatment group is significantly different from the other.
- One item (2) reflects the idea of better protection of vulnerable people, which was brought forward in the pro-social treatment. Both treatment groups exhibited a significantly higher share of correct answers than the control group, but the pro-social treatment group did not score significantly higher than the self-interest group.
- One item (4) reflecting the idea that returning to normal almost without restrictions is possible when many people use the app, was highlighted in the self-interest treatment. This somewhat bold (but not necessarily false) statement is supported by far the most often by members of the self-interest treatment group.
- The final item (5) captures general knowledge about app installment, something that respondents did not learn about in any of the treatments. While there is a statistically significant difference between both treatment groups, the substantive size of all differences is small compared to their differences on other items.

Overall, the findings demonstrate that the dissemination of information was done successfully with both treatments, albeit somewhat more so in the self-interest condition. The pro-social condition is not unique in leading treated respondents to support the idea that a higher number of app users helps the vulnerable.

Mean and 83%/95% confidence intervals

● Pro-social ● Self-interest

Supplementary Figure 7: Mean estimates and 83%/95% bootstrap confidence intervals on various covariates for the principal strata (message treatments), following the profiling method of (20).

2.3 Incentivization experiment setup and implementation

2.3.1 Treatment delivery

This experiment was embedded at the end of the wave 2 questionnaire. Respondents were then requested to state their agreement or non-agreement (see Supplementary Figure 8). When response (1) was selected, the respondent automatically qualified for the monetary payoff, which was dealt with by the survey provider. The payment was not conditional on further compliance measures, such as their reported app usage in the follow-up wave or the tracked usage. The provider was informed about respondent agreement at regular intervals while the survey was running.

Once a respondent agreed to install the app, they were shown a page with links to the Apple App Store and the Google Play store, along with further information about the app and a statement clarifying that the research was not affiliated with the government or other authorities (see Supplementary Figure 9). To facilitate installation for those respondents who did not fill out the survey on their mobile phones, they could request to see QR codes, which were then displayed on a follow-up page, and use them together with their phone to access the installation page.

Supplementary Figure 10 provides click-to-install rates by incentivization status, that is the proportion of those who clicked on one of the app store links provided to those who agreed to install the app after being incentivized, or who requested a QR link to the app stores. The click-to-install rates range between 45% and 55% and there are no significant differences between the different incentivization tiers.

Furthermore, while there are positive relationships between clicking behavior in wave 2 on the one hand and reported as well as tracked app uptake in wave 3 on the other, clicking on the app store links is not particularly predictive of future observed or reported behavior: 56% of those who clicked on any of the links in W2 **reported** to have installed the app in W3 (vs. 40% who did not click); 45% of those who clicked on any of the links in W2 **were observed** to have installed the app in W3) (vs. 27% who did not click) (see Tables 11 and 12).

Supplementary Table 11: Reported app adoption at wave 3 vs. click-to-install behavior observed in wave 2. Figures are based on incentivized subjects who agreed to install the app in wave 2.

		Reported app adoption, W3		
		No	Yes	Total
Click-to-install, W2	No	85 (60%)	57 (40%)	142 (50%)
	Yes	63 (44%)	80 (56%)	143 (50%)
	Total	148 (52%)	137 (48%)	285 (100%)

Sie haben angegeben, dass Sie die Corona-Warn-App derzeit nicht installiert haben.

Wir würden Ihnen 500 Mingle-Punkte gutschreiben, wenn Sie die Corona-Warn-App auf Ihrem eigenen Smartphone installieren.

Die Punkte würde Ihnen dann nach Ihrer Einwilligung in den nächsten Tagen gutgeschrieben werden.

Sie gehen dafür keinerlei Verpflichtung ein. Wir würden Sie lediglich bitten, die App zu installieren.

Sind Sie bereit, die App in den nächsten Tagen auf Ihrem Smartphone zu installieren?

Ja, ich bin bereit und werde die Corona-Warn-App in den nächsten Tagen auf meinem Smartphone installieren.

Nein, ich bin nicht bereit, die Corona-Warn-App zu installieren.

Ich habe die Corona-Warn-App bereits installiert.

Supplementary Figure 8: Screenshot of tracing app installation incentivization.

English translation: You have indicated that you currently do not have the Corona-Warn-App installed. We would credit you [100, 200, 500] Mingle Points if you install the Corona-Warn-App on your own smartphone. The points would then be credited to you within the next few days after your consent. There's no obligation. We only ask you to install the app. Are you ready to install the app on your smartphone within the next few days? (A) Yes, I am ready and will install the Corona-Warn-App on my smartphone within the next days (B) No, I am not ready to install the Corona-Warn-App. (C) I have already installed the Corona-Warn-App.

2.3.2 Covariate balance

Tables 13 reports covariate balance between treatment and control groups. Each row reports group-specific variable means and associated two-sided p -values for treatment-control differences.

For all treatment groups vs. control, employment status differences are significant at the 95% level. In addition, living in an urban region indicator is significant in the EUR 5 treatment group. Looking at the total number of comparisons across both treatments, this is within the range of what might be expected by chance. We address potential confounders by including

Supplementary Figure 9: Screenshot of messages following respondents' agreement with incentivization (Original Apple App Store / Google Play Store download badge removed for copyright reasons).

English translation, left page: Thank you very much for your agreement! Here you can download the Corona-Warn-App directly to your smartphone for free: [App Store/Play Store links] The app is an offer from the German government. Download and use of the app is completely voluntary. The research team behind this survey has no connection whatsoever with the federal government or subordinate authorities. If you do not complete the survey with your smartphone, but have it to hand, you can select "Show QR codes for app installation" below. You can also choose to install the app in the next few days. The points will be credited to your account within the next few days. (A) Show QR codes for app installation (B) Directly continue with survey

English translation, right page (shown when "Show QR codes for app installation" is selected): Here you can access the Google Play Store or Apple App Store via the QR Code scanner of your smartphone. To do so, open your camera application and hold the camera over the respective QR Code.

these variables for possible selection in all covariate-adjusted models.

2.3.3 Compliance

In this section, we provide evidence to which extent study participants complied with the different incentivization treatments. Our pre-registered definition of compliance takes as "treated" any respondent who installed the app after agreeing to do so. Supplementary Figure 11 reports shares of reported and tracked app uptake (based on the hybrid measure; see Section 2.4). Installation rates in the incentivization groups vary between 0.16 (EUR 2 tier) and 0.22 (EUR 5 tier). This is much higher than the installation rate in the control group (0.03), but also substantially lower than the initial agreement rates in W2.

Supplementary Figure 10: Click-to-install rates by incentivization status, restricted to those who agreed to install the app ($N_{\text{EUR } 1} = 259$, $N_{\text{EUR } 2} = 262$, $N_{\text{EUR } 5} = 240$). Two-sample t-test statistics: $T_{\text{EUR } 1 \text{ vs. EUR } 2} = 1.11$, d.f. = 207.45, $P = 0.266$, 95% CI $-0.06, 0.21$, $T_{\text{EUR } 1 \text{ vs. EUR } 5} = -0.28$, d.f. = 219.95, $P = 0.782$, 95% CI $-0.15, 0.11$, $T_{\text{EUR } 2 \text{ vs. EUR } 5} = -1.38$, d.f. = 206.12, $P = 0.169$, 95% CI $-0.23, 0.04$.

Supplementary Table 12: Tracked app adoption at wave 3 vs. click-to-install behavior observed in wave 2. Figures are based on incentivized subjects who agreed to install the app in wave 2.

Tracked app adoption, W3				
		No	Yes	Total
Click-to-install, W2	No	47 (73%)	17 (27%)	64 (49%)
	Yes	36 (55%)	30 (45%)	66 (51%)
Total		83 (64%)	47 (36%)	130 (100%)

Next, we explored how individual-level characteristics may have affected compliance with the treatment. Supplementary Figure 12 reports point estimates along with 83%/95% bootstrap confidence intervals on various covariates for the entire sample, compliers, always-takers, and never-takers. Compared to the sample of reported app non-users that qualified for the incentivization treatment, compliers tend to be younger, more likely female, have more trust in the health care system, and be more concerned about COVID-19.

Supplementary Table 13: Covariate balance: Incentive treatments vs. control (means and p values of treatment-control differences reported).

Covariate	Control	EUR 1 (p val.)	EUR 2 (p.val)	EUR 5 (p.val)
Age/10	4.73	4.77 (0.770)	4.62 (0.413)	4.52 (0.134)
Gender: Female	0.55	0.53 (0.677)	0.49 (0.155)	0.54 (0.729)
Education: Medium	0.43	0.40 (0.587)	0.40 (0.515)	0.37 (0.232)
Education: High	0.30	0.34 (0.373)	0.38 (0.059)	0.38 (0.045)
Household income: Medium	0.39	0.41 (0.588)	0.37 (0.783)	0.44 (0.195)
Household income: High	0.27	0.27 (0.948)	0.32 (0.159)	0.25 (0.674)
Children	0.56	0.52 (0.344)	0.56 (0.965)	0.60 (0.428)
Public transport usage	0.19	0.15 (0.204)	0.20 (0.875)	0.23 (0.301)
Restaurant/pub visit	0.19	0.17 (0.639)	0.13 (0.099)	0.19 (0.981)
Friend/family visit	0.49	0.46 (0.576)	0.49 (0.997)	0.52 (0.496)
NPI compliance	0.71	0.78 (0.068)	0.70 (0.800)	0.72 (0.709)
Precondition	0.40	0.38 (0.596)	0.38 (0.511)	0.39 (0.747)
COVID case in network	0.18	0.14 (0.194)	0.15 (0.334)	0.19 (0.838)
Lives in high COVID incidence region	0.06	0.07 (0.618)	0.05 (0.949)	0.07 (0.600)
Lives in urban region	0.44	0.42 (0.645)	0.41 (0.423)	0.35 (0.035)
Working	0.65	0.61 (0.405)	0.60 (0.278)	0.63 (0.749)
Trust in government	1.68	1.75 (0.487)	1.64 (0.728)	1.69 (0.899)
Trust in scientists	2.26	2.36 (0.251)	2.31 (0.623)	2.35 (0.340)
Trust in HC system	2.34	2.32 (0.798)	2.21 (0.154)	2.26 (0.390)
COVID threat perception: Self	1.38	1.40 (0.811)	1.26 (0.123)	1.30 (0.357)
COVID threat perception: Family and Friends	1.65	1.71 (0.459)	1.56 (0.284)	1.68 (0.773)
Social responsibility	0.03	-0.03 (0.519)	-0.04 (0.449)	-0.03 (0.446)
Self interest	0.04	-0.15 (0.037)	0.16 (0.159)	0.04 (0.954)
Data control	0.08	-0.03 (0.185)	0.17 (0.376)	0.05 (0.682)
Digital literacy	-0.02	-0.03 (0.911)	-0.05 (0.752)	-0.04 (0.828)

Supplementary Figure 11: App installation rates in wave 3, by incentivization status ($N_{\text{Control}} = 533$, $N_{\text{EUR 1}} = 223$, $N_{\text{EUR 2}} = 228$, $N_{\text{EUR 5}} = 214$). Two-sample t-test statistics: $T_{\text{EUR 1 vs. Control}} = 6.25$, d.f. = 260.52, $P < 0.001$, 95% CI 0.12, 0.24, $T_{\text{EUR 2 vs. Control}} = 4.92$, d.f. = 277.25, $P < 0.001$, 95% CI 0.08, 0.18, $T_{\text{EUR 5 vs. Control}} = 6.70$, d.f. = 245.96, $P < 0.001$, 95% CI 0.14, 0.26, $T_{\text{EUR 1 vs. EUR 2}} = 1.44$, d.f. = 441.51, $P = 0.152$, 95% CI -0.02 , 0.13 , $T_{\text{EUR 1 vs. EUR 5}} = -0.57$, d.f. = 432.41, $P = 0.566$, 95% CI -0.10 , 0.06 , $T_{\text{EUR 2 vs. EUR 5}} = -2.00$, d.f. = 422, $P = 0.047$, 95% CI -0.15 , -0.00 .

Mean and 83%/95% confidence intervals

Supplementary Figure 12: Mean estimates and 83%/95% bootstrap confidence intervals on various covariates for the principal strata (incentivization tiers), following the profiling method of (20).

2.4 Variable specifications

In this section, we describe the specifications of the variables we used for our analysis. To see the full wording of the survey items, see Section 2.5.

2.4.1 App uptake

We use both a survey-based measure and the tracking data to measure app uptake. The survey-based measure is the basis for the outcome “reported app uptake”. The tracking-based measure is the basis for the outcome “tracked app uptake”. Furthermore, we construct a “hybrid app uptake” measure to allow us to pool the different samples to analyze effects on app uptake. Mostly in line with our pre-analysis plan, we use the scheme specified in Supplementary Figure 13 to code the hybrid variable (see Section 2.6 for a technical deviation from the pre-registered scheme):

Supplementary Figure 13: Coding scheme for hybrid app uptake indicator.

To arrive at the survey-based measure, two items are used. First, and in wave 1 only, respondents are asked whether they have access to a smartphone. If the answer is positive, they receive a follow-up question asking them whether they or someone else has installed or not installed the official Corona-Warn-App on their smartphone. The three possible answers are “App installed”, “App not installed”, and “App installed, but uninstalled since then”. In waves 2 and 3, the separate smartphone access question is replaced with an additional response option “I don’t have access to a smartphone”. The instrument was derived from an item of the Mannheim Corona Study (Questionnaire CW 14), which is based on the methodology and infrastructure of the German Internet Panel (GIP) (21).

2.4.2 Probability to install app

Those who report to not having the app installed despite having a smartphone were asked how likely is it that they would install the Corona-Warn-App on their smartphone in the future. The five answer categories “Very unlikely”, “Somewhat unlikely”, “Undecided”, “Somewhat likely”, and “Very likely” are recoded into numeric scores (Very unlikely = 0 to Very likely = 4). Due to a programming mistake in the wave 2 questionnaire, the item was only asked in wave 1 (pre-treatment) and wave 3. This inhibited us from testing the pre-registered hypothesis of message effects on the likelihood of app uptake with wave 2 data. However, we report longer-term effects using the wave 1 and wave 3 data in Section 3.

2.4.3 Bluetooth usage

Bluetooth usage could not be directly observed in the tracking data. Therefore, we use a survey item in all three waves, asking the respondents who report to have the app installed how often they use the Corona Warning App by activating Bluetooth communication on their smartphone. The five answer categories “Always”, “Mostly”, “Sometimes”, “Rarely”, and “Never” are recoded into numeric scores (Never = 0 to Always = 4). The instrument was derived from an item of the Mannheim Corona Study (Questionnaire CW 14), which is based on the methodology and infrastructure of the German Internet Panel (GIP) (21).

2.4.4 App knowledge

We used a battery of right-wrong-DK items to measure factual knowledge on the app. To avoid learning effects, these questions were not asked before treatment in wave 1. The outcome variable is calculated as the total number of correct answers divided by total number of items. In addition, the items “The more people who use the app, the better the protection of particularly vulnerable people” and “If a lot of people use the app, you could soon return to your daily life with almost no restrictions” are used as standalone alternative outcomes to check whether message-treatment condition-specific knowledge was picked up at different levels in the different treatment groups (see Section 2.2.4). The knowledge battery was repeated in waves 2 and 3. The data from wave 1 are exclusively used for the manipulation check, the data from wave 2 is used to estimate the effect of the message treatments on app knowledge reported in the main text, and wave 3 data is used to estimate the effect of the incentivization treatment on app knowledge as reported in the main text and as an alternative outcome for the message experiment (see Section 3).

2.4.5 App attitudes

We measure attitudes towards the app by conducting a principal components analysis (PCA) of their responses to a set of items. We run PCA in the set of responses in wave 1 to derive

an attitude index by taking the first principal component, and then predict the attitude index in waves 2 and 3 by using the same set of weights computed in wave 1.

The attitudinal items we consider are highly correlated, with a single principal component explaining 54% of the variance (see Supplementary Table 14. Higher values on the scale indicate more positive attitudes towards the app.

Supplementary Table 14: PCA loadings: App attitudes scale. Standardized loadings for first principal component based upon correlation matrix (with varimax rotation). The first principal component explains 54% of the variance.

Survey item	Loading
I don't think that the app is of any use in fighting the pandemic.	-0.80
I think it makes sense that the app is used as an instrument for tracking infections.	0.75
I do not feel well enough informed about the app.	-0.56
I am concerned about the privacy of the app.	-0.79

2.4.6 App message sharing behavior and app information lookup

At the end of the wave 1 and wave 2 questionnaires, respondents were shown an information screen that featured links to external pages with additional information about the app as well as links to the Corona Warning App in Apple's App Store and Google's Play Store (see Supplementary Figure 14). Furthermore, additional links offered ways to share information about the app via social media (Facebook, Twitter, Whatsapp) and email.

These links were internally tagged, which allowed us to track clicks on any of them. We used this to measure **sharing behavior**, which we coded 1 if a respondent clicked on any of the Facebook, Twitter, WhatsApp or Email share buttons on the information screen, and otherwise 0. Analogously, we created a measure of **information lookup**, which we coded 1 if a respondent clicked on any of the information or app store links, otherwise 0.

2.4.7 Hypothetical app usage

The tracking data did not allow use to observe internal app usage in detail. Obviously, for digital contact tracing to work, installing the app is only the first step. It furthermore requires compliant users that would be (a) responsive to warnings displayed by the app and (b) willing to report positive test results in the app. Therefore, we asked respondents about their hypothetical behavior in such scenarios. First, we asked how sure they were that they would have themselves tested and quarantined for COVID-19 when the app informed them about a recent risk encounter. The seven answer categories (1 = "Certainly not get tested/quarantined" to 7 = "Certainly get tested/quarantined") were recoded into numeric scores. Second, we asked how sure they were that they would report a positive test result in the app. The seven answer categories (1 = "Certainly not report" to 7 = "Certainly report") were recoded into numeric scores.

Zum Abschluss möchten wir Sie nochmals auf einige nützliche Informationen im Zusammenhang mit der Corona-Warn-App hinweisen.

- Fragen und Antworten zur Corona-Warn-App [bei der Verbraucherzentrale](#)
- Linus Neumann (Sprecher des Chaos Computer Clubs) über die Sicherheit der Corona-Warn-App [bei tagesschau.de](#)
- Bundeskanzlerin Angela Merkel zur Corona-Warn-App auf der [Seite der Bundesregierung](#)
- Kurzes Erklärvideo zur Funktionsweise der Corona-Warn-App auf der [Seite der Bundesregierung](#)

Hier können Sie die Corona-Warn-App kostenlos auf Ihr Smartphone herunterladen:

Im Apple App Store
downloaden

Im Google Play Store
downloaden

Die App ist ein Angebot der Bundesregierung. Download und Nutzung der App sind vollkommen freiwillig. Das Forscherteam hinter dieser Befragung steht in keinerlei Verbindung zur Bundesregierung oder nachgeordneten Behörden.

Hier können Sie Ihre Freunde auf die App hinweisen:

Facebook

Twitter

WhatsApp

E-Mail

Supplementary Figure 14: Screenshot of information page that respondents were shown post-treatment, that refers to additional information, links to the app stores, and offers ways to share information about the app via social media or email. (Original Apple App Store / Google Play Store download badge removed for copyright reasons).

2.4.8 Covariates

In the following, we briefly describe how further covariates that are used in the models as well as for descriptives are generated. We had not pre-registered these procedures.

Age is measured in five categories (18-29 years, 30-39 years, 40-49 years, 50-59 years, 60+ years). For some descriptives we use age as a continuous variable to facilitate interpretation.

Gender is measured using a dummy variable (male = 0, female = 1).

Education is measured in three categories (low = [did not finish school (yet), or finished school but holds no qualification to pursue education to satisfy university entrance requirements], intermediate = [finished school with qualification to pursue further education to satisfy

university entrance requirements], high = [finished school achieving university entrance requirements, and/or holds university degree and/or post-graduate degree]).

Number of children is measured in four categories (No children, 1 child, 2 children, 3 children, 4 or more children) and for some descriptives coded binary as children yes/no to facilitate interpretation).

Household income is measured in six categories (“Up to EUR 500”, “Between EUR 500 and 1,500”, “Between EUR 1,500 and 3,000”, “Between EUR 3,000 and 5,000”, “Between EUR 5,000 and 10,000”, “More than EUR 10,000”). In the main models, these are collapsed into three categories (“Up to EUR 1,499 EUR”, “EUR 1,500 to 2,999”, “EUR 3,000 and more”).

Living in region with high number of cases is measured by matching respondents’ zip code to district data published by the RKI as of August 25, 2020. Zip code areas with a mean number of cases larger than the 90% percentile of all zip code areas (here: more than 484 cases per 100k inhabitants) are coded as high-incidence regions. Also using zip code data, we classify the respondent as **living in an urban region** when the zip code belongs to a town or city.

Employment status: working is measured by classifying the respondent’s employment status as working if he/she reports to be working full time or part time, attending a school or university, or being in a trainee-ship or apprenticeship scheme.

Trust in government, scientists, and the healthcare system is measured using an item that asks respondents how much trust they have in the respective institutions. The five answer categories, ranging from “Not trust at all” to “Complete trust” are recoded into numeric scores (“Not trust at all” = 0 to “Complete trust” = 4).

Pre-existing health conditions are measured by asking respondents whether they “have any pre-existing conditions that increase the risk of a severe course of COVID-19 (e.g., high blood pressure, obesity, diabetes, COPD)” and encoding their answers to yes/no.

Incidences of COVID-19 infections in personal environment are measured by asking respondents whether they know of anyone (including yourself, family, friends, acquaintances, colleagues, or neighbors) who has been infected by COVID-19. Answers “Yes, 1-3 people” and “Yes, more than 3 people” are encoded as 1, “No” as 0.

Threat perception of COVID-19 for themselves and family/friends is measured by asking respondents how concerned they are about consequences of the coronavirus for their themselves and family and friends separately. The four-point scale (“Not at all concerned”, “Not too concerned”, “Somewhat concerned”, “Very concerned”) is recoded into numeric scores (“Not at all concerned” = 0 to “Very concerned” = 4).

Data privacy concerns are measured by conducting a principal components analysis (PCA) of responses to a set of items. We run PCA with two components on the set of responses in wave 1 to derive a data privacy concerns index by taking the first principal component. Two of the three items load highly on the first component. Two principal components together explain 78% of the variance (see Supplementary Table 15). Higher values on the scale indicate higher data privacy concerns.

Social responsibility is measured by conducting a principal components analysis (PCA) of responses to a set of items. We run PCA with one component on the set of responses in wave

Supplementary Table 15: PCA loadings: Data privacy concerns scale. Standardized loadings for first principal component based upon correlation matrix (with varimax rotation). First principal component explains 43% of the variance on all 3 items.

Survey item	Loading C1
It is important for me to control who is able to store and access information collected about me.	0.86
The government protects my privacy.	0.13
I am concerned about how many companies can access data about me without my knowledge.	0.74

1 to derive a social responsibility index by taking the first principal component. All three items load highly on the first component; the component explains 52% of the variance (see Table 16). Higher values on the scale indicate higher social responsibility. The items were taken from the Self-and Other-Interest Inventory (22), where they were among the high-loading items in a more extensive other-interest subscale.

Supplementary Table 16: PCA loadings: Social responsibility scale. Standardized loadings for first principal component based upon correlation matrix (with varimax rotation). First principal component explains 52% of the variance on all 3 items.

Survey item	Loading
I'm the kind of person that others can count on.	0.78
I generally carry out my tasks as best I can.	0.78
When a stranger needs my support, I try to give it to him.	0.59

Self-interest is measured by conducting a principal components analysis (PCA) of responses to a set of items. We run PCA with one component on the set of responses in wave 1 to derive a self-interest index by taking the first principal component. All three items load highly on the first component; the component explains 64% of the variance (see Supplementary Table 17). Higher values on the scale indicate higher self-interest. The items were taken from the Self-and Other-Interest Inventory (22), where they were among the high-loading items in a more extensive self-interest subscale.

Digital literacy is measured by conducting a principal components analysis (PCA) of responses to a set of items. We run PCA with one component on the set of responses in wave 1 to derive a digital literacy index by taking the first principal component. All four items load reasonably highly on the first component; the component explains 43% of the variance (see Supplementary Table 18). Higher values on the scale indicate higher digital literacy. These items were taken from a validated digital literacy scale (11).

Compliance with anti-COVID rules is measured by asking respondents to what extent they comply with behavioral recommendations to contain the pandemic. The four-point scale

Supplementary Table 17: PCA loadings: Self-interest scale. Standardized loadings for first principal component based upon correlation matrix (with varimax rotation). The first principal component explains 64% of the variance.

Survey item	Loading
I am constantly looking for ways to get ahead.	0.85
Success is important to me.	0.78
I am constantly looking out for what will make me happy.	0.76

Supplementary Table 18: PCA loadings: Digital literacy scale. Standardized loadings for first principal component based upon correlation matrix (with varimax rotation). The first principal component explains 43% of the variance.

Survey item	Loading
I prefer to ask friends how to use any new technological gadget instead of trying to figure it out myself.	-0.47
I feel like information technology is a part of my daily life.	0.80
Using information technology makes it easier to do my work.	0.79
I often have trouble finding things that I've saved on my computer.	-0.49

(how concerned they are about consequences of the coronavirus for themselves and their family and friends separately. The four-point scale (“I almost always follow them”, “I try to adhere to them, but often I do not succeed (e.g., for professional reasons)”, “I barely adhere to them”, “I am not aware of these recommendations”) is recoded into a binary indicator of compliance (“I almost always follow them” = 1, otherwise = 0).

Risk behavior is measured using a battery of items, asking respondents how often in the last seven days they have (a) used public transport, (b) visited a restaurant, café or bar, and (c) met with friends, relatives or acquaintances in person. For each item, the four-point scale (“Never”, “Once this week”, “Several times this week”, “Every day”) was broken down into “Low” (Never, Once this week) and “High” (Several times this week, Every day) risk behavior.

2.5 Survey Questionnaire

This section shows the specific wordings of all the survey items we used as part of this study. The wave in which the questions were asked is indicated in square brackets. The original (German) Qualtrics implementation of the online questionnaire is available from the following URLs:

- wave 1: https://scripts.fra1.qualtrics.com/jfe/form/SV_cTsWKpNFdx2pS8B?Q_Language=DE
- wave 2: https://scripts.fra1.qualtrics.com/jfe/form/SV_0ULzi6WbDsvZMih?Q_Language=DE
- wave 3: https://scripts.fra1.qualtrics.com/jfe/form/SV_brXYcMjXcZ5o9GR?Q_Language=DE

2.5.1 Sociodemographics

Gender [W1]	
German (original)	English translation
Bitte geben Sie Ihr Geschlecht an.	Please state your gender.
(A) Männlich	(A) Male
(B) Weiblich	(B) Female
(C) Divers	(C) Other

Year of birth [W1]	
German (original)	English translation
In welchem Jahr sind Sie geboren? Bitte auswählen... [1920–2005]	What is your year of birth? Please select... [1920–2005]

Education (school degree) [W1]	
German (original)	English translation
Welches ist der höchste Bildungsabschluss, den Sie erhalten haben?	What is the highest level of school you have completed or the highest degree you have received?
(A) Kein oder noch kein Schulabschluss	(A) No school-leaving certificate (yet)
(B) Hauptschulabschluss, Volksschulabschluss	(B) Secondary school leaving certificate ("general school"), elementary school leaving certificate
(C) Mittlere Reife, Realschulabschluss, Polytechnische Oberschule mit Abschluss 10. Klasse	(C) Secondary school leaving certificate ("middle school"), "Middle Maturity"
(D) Fachhochschulreife, Abitur, erweiterte Oberschule mit Abschluss 12. Klasse, Hochschulreife	(D) A-Levels, advanced technical certificate, higher education entrance qualification
(E) Fachhochschul- oder Hochschulabschluss, Universitätsabschluss (z.B. Bachelor, Master, Diplom, Staatsexamen)	(E) College or university degree (e.g., Bachelor's degree, Master's degree, university diploma)
(F) Promotion, Habilitation	(F) Postgraduate degree (Ph.D., habilitation)

Children [W1]	
German (original)	English translation
Haben Sie Kinder?	Do you have children?
(A) Nein	(A) No
(B) Ja, 1 Kind	(B) Yes, 1 child
(C) Ja, 2 Kinder	(C) Yes, 2 children
(D) Ja, 3 Kinder	(D) Yes, 3 children
(E) Ja, 4 oder mehr Kinder	(E) Yes, 4 or more children

Household income [W1]	
German (original)	English translation
Wie hoch war das durchschnittliche monatliche Nettoeinkommen im letzten Jahr (2019) in Ihrem Haushalt insgesamt nach Abzug der Steuern und Sozialversicherungen?	What was your average monthly household income after deducting taxes and social security costs?
(A) Bis 500 Euro	(A) Up to 500 Euro
(B) Über 500 bis 1.500 Euro	(B) Between 500 and 1,500 Euro
(C) Über 1.500 bis 3.000 Euro	(C) Between 1,500 and 3,000 Euro
(D) Über 3.000 bis 5.000 Euro	(D) Between 3,000 and 5,000 Euro
(E) Über 5.000 bis 10.000 Euro	(E) Between 5,000 and 10,000 Euro
(F) Über 10.000 Euro	(F) More than 10,000 Euro
(G) Weiß nicht	(G) Don't know

Zip code [W1]	
German (original)	English translation
Wie lautet die Postleitzahl Ihres Wohnorts? [Offenes Textfeld]	What is your zip code? [Open text field]

Job/employment status [W1]	
German (original)	English translation
Wie ist Ihr aktueller Beschäftigungsstatus?	What is your current employment status?
(A) Zurzeit arbeitslos	(A) Unemployed
(B) In Vollzeit beschäftigt (mehr als 30 Stunden/Woche)	(B) Working full time (more than 30h/week)
(C) In Teilzeit beschäftigt (bis zu 30 Stunden/Woche)	(C) Working part time (up to 30h/week)
(D) In Rente/pensioniert	(D) Retired
(E) In schulischer oder universitärer Ausbildung	(E) Attending school or university
(F) In Ausbildung als Lehrling/Azubi	(F) In a trainee-ship or apprenticeship scheme
(G) Nicht berufstätig (Hausfrau/Hausmann)	(G) Stay-at-home partner
(H) In Mutterschaft oder Elternzeit	(H) On parental leave
(I) Anders, und zwar: [offenes Textfeld]	(I) Other (Please specify): [open text field]

2.5.2 Attitudes and opinions

COVID-19 threat perception [W1, W2, W3]	
German (original)	English translation
Wir würden Ihnen nun gerne einige Fragen zur COVID-19-(Coronavirus)-Pandemie stellen. Wie besorgt sind Sie über die Folgen des Coronavirus...	We would now like to ask you some questions related to the COVID-19 (coronavirus) pandemic. How concerned are you about consequences of the coronavirus...
(A) für Sie selbst?	(A) for yourself?
(B) für Ihre Familie und Freunde?	(B) for your family and friends?
(C) für Deutschland?	(C) for Germany?
1. Überhaupt nicht besorgt	1. Not at all concerned
2. Eher nicht besorgt	2. Not too concerned
3. Eher besorgt	3. Somewhat concerned
4. Sehr besorgt	4. Very concerned

Digital literacy [W1]	
German (original)	English translation
Bitte geben Sie an, inwiefern die folgenden Aussagen zutreffen oder nicht zutreffen.	Please indicate the extent to which you agree with the following statements.
<p>(A) Ich ziehe es vor, Freunde zu fragen, wie man ein neues technisches Gerät verwendet, anstatt es selbst herauszufinden.</p> <p>(B) Informationstechnologie ist ein Teil meines täglichen Lebens.</p> <p>(C) Der Einsatz von Informationstechnologie erleichtert mir die Arbeit.</p> <p>(D) Ich habe oft Probleme, Dinge zu finden, die ich auf meinem Computer gespeichert habe.</p> <p>1. Trifft überhaupt nicht zu</p> <p>2. Trifft eher nicht zu</p> <p>3. Trifft eher zu</p> <p>4. Trifft voll und ganz zu</p> <p>5. Weiß nicht</p>	<p>(A) I prefer to ask friends how to use any new technological gadget instead of trying to figure it out myself.</p> <p>(B) I feel like information technology is a part of my daily life.</p> <p>(C) Using information technology makes it easier to do my work.</p> <p>(D) I often have trouble finding things that I've saved on my computer.</p> <p>1. Strongly disagree</p> <p>2. Somewhat disagree</p> <p>3. Somewhat agree</p> <p>4. Strongly agree</p> <p>5. Don't know</p>

Data privacy concerns [W1]	
German (original)	English translation
Und inwiefern stimmen Sie den folgenden Aussagen zu oder nicht zu?	And to what extent do you agree with the following statements?
<p>(A) Für mich ist es wichtig zu kontrollieren, wer in der Lage ist, über mich gesammelte Informationen zu speichern und darauf zuzugreifen.</p> <p>(B) Die Regierung schützt meine Privatsphäre.</p> <p>(C) Ich mache mir Sorgen darüber, wie viele Unternehmen ohne mein Wissen auf Daten über mich zugreifen können.</p> <p>1. Trifft überhaupt nicht zu</p> <p>2. Trifft eher nicht zu</p> <p>3. Trifft eher zu</p> <p>4. Trifft voll und ganz zu</p> <p>5. Weiß nicht</p>	<p>(A) It is important for me to control who is able to store and access information collected about me.</p> <p>(B) The government protects my privacy.</p> <p>(C) I am concerned about how many companies can access data about me without my knowledge.</p> <p>1. Strongly disagree</p> <p>2. Somewhat disagree</p> <p>3. Somewhat agree</p> <p>4. Strongly agree</p> <p>5. Don't know</p>

Institutional trust [W1]	
German (original)	English translation
Wie sehr vertrauen Sie den folgenden Gruppen und Institutionen?	How much trust do you have in the following groups and institutions?
(A) Bundesregierung	(A) Federal government
(B) Landesregierung	(B) State government
(C) Wissenschaftliche Experten	(C) Scientific experts
(D) Medien	(D) Media
(E) Gesundheitssystem	(E) The healthcare system
1. 1 = Vertraue überhaupt nicht	1. 1 = Not trust at all
2. 2	2. 2
3. 3	3. 3
4. 4	4. 4
5. 5 = Vertraue voll und ganz	5. 5 = Complete trust

Self-interest and pro-social attitudes [W1]	
German (original)	English translation
Inwiefern stimmen Sie folgenden Aussagen zu?	To what extent do you agree with the following statements?
(A) Ich gehöre zu der Art von Menschen, auf die andere sich verlassen können.	(A) I'm the kind of person that others can count on.
(B) Ich erledige meine Aufgaben im Allgemeinen so gut ich kann.	(B) I generally carry out my tasks as best I can.
(C) Wenn ein Fremder auf meine Unterstützung angewiesen ist, versuche ich, sie ihm zu geben.	(C) When a stranger needs my support, I try to give it to him.
(D) Ich bin ständig auf der Suche nach Wegen, um voranzukommen.	(D) I am constantly looking for ways to get ahead.
(E) Erfolg ist mir wichtig.	(E) Success is important to me.
(F) Ich bin ständig auf der Suche nach dem, was mich glücklich macht.	(F) I am constantly looking out for what will make me happy.
1. Stimme überhaupt nicht zu	1. Completely disagree
2. Stimme eher nicht zu	2. Somewhat disagree
3. Stimme eher zu	3. Somewhat agree
4. Stimme voll und ganz zu	4. Completely agree
5. Weiß nicht	5. Don't know

2.5.3 Risk status and behavior

COVID-19 cases in network [W1]	
German (original)	English translation
Kennen Sie jemanden in Ihrer persönlichen Umgebung, Sie selbst eingeschlossen, Ihre Familie, Freunde, Bekannte, Kollegen oder Nachbarn, die mit COVID-19 (Coronavirus) infiziert wurden?	Do you know of anyone (including yourself, family, friends, acquaintances, colleagues, or neighbors) who has been infected by COVID-19?
(A) Nein	(A) No
(B) Ja, 1-3 Personen	(B) Yes, 1-3 people
(C) Ja, mehr als 3 Personen	(C) Yes, more than 3 people

Existing pre-conditions [W1]	
German (original)	English translation
Haben Sie eine Vorerkrankung, die das Risiko eines schweren Verlaufs von COVID-19 erhöht (z.B. Bluthochdruck, Adipositas, Diabetes, Lungenerkrankung)?	Do you have any pre-existing conditions that increase the risk of severe COVID-19 symptoms (e.g., high blood pressure, obesity, diabetes, COPD)?
(A) Ja	(A) Yes
(B) Nein	(B) No
(C) Weiß nicht	(C) Don't know

Risk behavior [W1, W2, W3]	
German (original)	English translation
Wenn Sie an die letzten sieben Tage zurückdenken: Wie häufig haben Sie folgendes getan?	In the last seven days: How often have you done the following?
(A) Öffentliche Verkehrsmittel genutzt	(A) Used public transport
(B) Restaurant, Café oder Bar besucht	(B) Visited a restaurant, café or bar
(C) Persönlich mit Freunden, Verwandten oder Bekannten getroffen	(C) Met with friends, relatives or acquaintances in person
1. Gar nicht	1. Never
2. An einem Tag	2. Once this week
3. An mehreren Tagen	3. Several times this week
4. Nahezu täglich oder täglich	4. Every day
5. Weiß nicht	5. Don't know

COVID-19 precautions [W1, W2, W3]

German (original)	English translation
Die Pandemie ist noch nicht überwunden. Weiterhin gilt: Abstand halten; Hände häufig waschen; Maske tragen, wenn Abstand halten nicht möglich ist. Wie halten Sie es mit diesen Empfehlungen?	With the pandemic that has taken place over the past few months, the following recommendations have been issued: Keep your distance; wash your hands frequently; wear a mask if keeping your distance is not possible. To what extent do you comply with these recommendations?
(A) Ich halte mich fast immer daran	(A) I almost always follow them
(B) Ich versuche, mich daran zu halten, häufig gelingt das aber nicht (z. B. aus beruflichen Gründen)	(B) I try to adhere to them, but often I do not succeed (e.g., for professional reasons)
(C) Ich halte mich kaum daran	(C) I barely adhere to them
(D) Ich kenne diese Empfehlungen nicht	(D) I am not aware of these recommendations

2.5.4 Tracing app usage and attitudes

Smartphone [W1]	
German (original)	English translation
Haben Sie Zugang zu einem Smartphone?	Do you have access to a smartphone?
(A) Ich habe mein eigenes Smartphone	(A) I have my own smartphone
(B) Ich habe kein eigenes Smartphone, habe aber täglich Zugriff auf ein Smartphone	(B) I don't have my own smartphone, but have daily access to a smartphone
(C) Ich habe kein eigenes Smartphone, aber wöchentlichen Zugriff auf ein Smartphone	(C) I don't have my own smartphone, but have weekly access to a smartphone
(D) Ich habe keinen Zugang zu einem Smartphone	(D) I don't have access to a smartphone

Tracing app install [W1, W2, W3]	
German (original)	English translation
<p>Zuletzt haben wir noch ein paar Fragen an Sie zur Corona-Warn-App, die das Robert-Koch-Institut (RKI) seit kurzem zur Verfügung stellt. Manche Leute installieren die Corona-Warn-App, während andere Leute die App nicht installieren können oder möchten. Wie ist es bei Ihnen? Haben Sie oder hat jemand für Sie die offizielle Corona-Warn-App auf Ihrem Smartphone installiert oder nicht installiert?</p> <p>(A) App installiert</p> <p>(B) App nicht installiert</p> <p>(C) App installiert, aber seitdem wieder deinstalliert</p> <p>(D) [W2, W3] Ich besitze kein Smartphone</p>	<p>Finally, we have a few questions for you about the Corona-Warn-App, which the Robert Koch Institute (RKI) has recently made available. Some people install the Corona-Warn-App, while other people cannot or do not want to install the app. How do you feel about it? Have you or someone else installed the official Corona-Warn-App on your smartphone?</p> <p>(A) App installed</p> <p>(B) App not installed</p> <p>(C) App installed, but uninstalled since then</p> <p>(D) [W2, W3] I don't have a smartphone</p>

App Bluetooth usage [W1, W2, W3]

German (original)

English translation

Es könnte sein, dass Sie die App installiert haben, diese aber nicht regelmäßig nutzen, z.B. weil Sie die Bluetooth-Technologie nicht immer aktiviert haben möchten oder der Smartphone-Akku schnell leer läuft. Wenn Sie unterwegs sind bzw. unter Leute gehen, wie häufig nutzen Sie die Corona-Warn-App, indem Sie Bluetooth-Kommunikation auf Ihrem Smartphone aktiviert haben?

It could be that you have installed the app but do not use it regularly, e.g. because you do not always want to have Bluetooth technology activated or the smartphone's battery runs out too quickly. When you are travelling or socializing, how often do you use the Corona Warning App by activating Bluetooth communication on your smartphone?

(A) Immer

(A) Always

(B) Meistens

(B) Mostly

(C) Manchmal

(C) Sometimes

(D) Selten

(D) Rarely

(E) Nie

(E) Never

App positive test report [W1, W2, W3]	
German (original)	English translation
Nehmen wir an, Sie würden positiv auf COVID-19 getestet werden. Wie sicher sind Sie, dass Sie das Testergebnis in der App melden würden?	Let's assume you tested positive for COVID-19. How sure are you that you would report the test result in the app?
(A) 1 = Sicher nicht melden	(A) 1 = Certainly not report
(B) 2	(B) 2
(C) 3	(C) 3
(D) 4	(D) 4
(E) 5	(E) 5
(F) 6	(F) 6
(G) 7 = Sicher melden	(G) 7 = Certainly report
(H) Weiß nicht	(H) Don't know

App risk display [W1, W2, W3]	
German (original)	English translation
Hat Ihnen die App schon einmal ein erhöhtes Risiko bzw. eine Risiko-Begegnung angezeigt?	Has the app ever displayed an increased risk or risk encounter?
(A) Ja	(A) Yes
(B) Nein	(B) No
(C) Weiß nicht	(C) Don't know

App risk reaction [W1, W2, W3]	
German (original)	English translation
Haben Sie sich daraufhin testen lassen bzw. unter Quarantäne gestellt?	Were you tested or quarantined?
(A) Ja, ich habe mich testen lassen und unter Quarantäne gestellt.	(A) Yes, I had myself tested and quarantined.
(B) Ich habe mich testen lassen, mich aber nicht unter Quarantäne gestellt.	(B) I had myself tested but not quarantined.
(C) Ich habe mich unter Quarantäne gestellt, aber nicht testen lassen.	(C) I quarantined but was not tested.
(D) Nein, ich habe mich nicht testen lassen und auch nicht unter Quarantäne gestellt.	(D) No, I did not get tested and I have not quarantined.
(E) Möchte ich nicht sagen	(E) I do not want to say

App risk reaction, hypothetical [W1, W2, W3]	
German (original)	English translation
Nehmen wir an, die App informiert Sie, dass Sie kürzlich eine Risiko-Begegnung hatten. Wie sicher sind Sie, dass Sie sich selbst auf COVID-19 testen lassen und unter Quarantäne stellen würden?	Suppose the app informs you that you recently had a risky encounter. How sure are you that you would have yourself tested and quarantined?
(A) 1 = Sicher nicht testen / nicht unter Quarantäne stellen	(A) 1 = Certainly not get tested/quarantined
(B) 2	(B) 2
(C) 3	(C) 3
(D) 4	(D) 4
(E) 5	(E) 5
(F) 6	(F) 6
(G) 7 = Sicher testen / unter Quarantäne stellen	(G) 7 = Certainly get tested/quarantined
(H) Weiß nicht	(H) Don't know

Attitudes towards app [W1, W2, W3]	
German (original)	English translation
Inwiefern stimmen Sie den folgenden Aussagen zur Corona-Warn-App zu?	To what extent do you agree with the following statements about the Corona-Warn-App?
(A) Ich glaube nicht, dass die App für die Bekämpfung der Pandemie etwas nützt.	(A) I don't think that the app is any use in fighting the pandemic.
(B) Ich finde es sinnvoll, dass die App als Zusatzinstrument zur Nachverfolgung von Infektionen genutzt wird.	(B) I think it makes sense that the app is used as an instrument for tracking infections.
(C) Ich fühle mich nicht gut genug über die App informiert.	(C) I do not feel well enough informed about the app.
(D) Ich Sorge mich um den Datenschutz der App.	(D) I am concerned about the privacy of the app.
(E) Ich habe bereits viel Werbung für die App gesehen.	(E) I have already seen a lot of advertisements for the app.
1. Stimme überhaupt nicht zu	1. Completely disagree
2. Stimme eher nicht zu	2. Somewhat disagree
3. Stimme eher zu	3. Somewhat agree
4. Stimme voll und ganz zu	4. Completely agree
5. Weiß nicht	5. Don't know

Likelihood to install app [W2, W3]	
German (original)	English translation
Wie wahrscheinlich ist es, dass sie die Corona-Warn-App zukünftig auf Ihrem Smartphone installieren werden?	How likely is it that you will install the Corona-Warn-App on your smartphone in the future?
(A) Überhaupt nicht wahrscheinlich	(A) Very unlikely
(B) Eher nicht wahrscheinlich	(B) Somewhat unlikely
(C) Unentschieden	(C) Undecided
(D) Eher wahrscheinlich	(D) Somewhat likely
(E) Sehr wahrscheinlich	(E) Very likely

Knowledge about app [W1, W2, W3]	
German (original)	English translation
Welche dieser Aussagen über die Corona-Warn-App halten Sie für richtig und welche für falsch?	Which of these statements about the Corona Warning App do you consider correct and which ones are wrong?
<p>(A) Über die App gesammelte Daten werden zentral beim Robert-Koch-Institut gespeichert.</p> <p>(B) Je mehr Menschen die App nutzen, desto besser sind besonders gefährdete Menschen geschützt.</p> <p>(C) Trotz Anonymität kann die App mir helfen, schneller zu wissen, dass ich möglicherweise infiziert wurde.</p> <p>(D) Bis heute wurde die App etwa 10 Millionen mal installiert.</p> <p>(E) Wenn viele Menschen die App nutzen, könnte man bald wieder zu einem Alltag fast ohne Einschränkungen übergehen.</p> <p>1. Falsch</p> <p>2. Richtig</p> <p>3. Weiß nicht</p>	<p>(A) Data collected via the app is stored centrally at the Robert Koch Institute.</p> <p>(B) The more people who use the app, the better particularly vulnerable people are protected.</p> <p>(C) Despite anonymity, the app can help me to know more quickly that I may have been infected.</p> <p>(D) To date, the app has been installed about 10 million times.</p> <p>(E) If a lot of people use the app, you could soon return to your daily life with almost no restrictions.</p> <p>1. Wrong</p> <p>2. Right</p> <p>3. Don't know</p>

Knowledge of app icon [W1, W2, W3]	
German (original)	English translation
<p>Welches der folgenden Logos ist das der offiziellen Corona-Warn-App? Bitten sehen Sie die Antwort nicht nach, sondern wählen Sie einfach das Logo, das Ihnen richtig erscheint.</p> <p>[display of six icons, including the correct icon]</p>	<p>Which of the following logos is for the official Corona-Warning-App? Please do not look up the answer, just choose the logo that seems right to you.</p>

Knowledge of app screenshot [W3]	
German (original)	English translation
Und welcher der folgenden Screenshots stammt von der offiziellen Corona-Warn-App? Bitten sehen Sie die Antwort nicht nach, sondern wählen Sie einfach den Screenshot, der Ihnen richtig erscheint.	And which of the following screenshots comes from the official Corona-Warn-App? Please do not look up the answer, just choose the screenshot that seems right to you.
[display of six app screenshots, including the correct screenshot]	

Reasons not to install app [W3]

German (original)	English translation
Können Sie uns sagen, weshalb Sie die App nicht installiert haben oder wieder deinstalliert haben?	Can you tell us why you did not install the app or uninstalled it again?
(A) Ich glaube nicht, dass die App etwas nützt.	(A) I don't think that the app is of any use.
(B) Ich Sorge mich um den Datenschutz.	(B) I am concerned about privacy.
(C) Ich möchte nicht, dass mich der Staat überwacht.	(C) I do not want the state to monitor me.
(D) Das Bluetooth meines Smartphones müsste dauerhaft angeschaltet sein.	(D) The Bluetooth functionality on my smartphone would have to be turned on permanently.
(E) Ich will nicht, dass die App verfolgt, wen ich treffe.	(E) I do not want the app to track who I meet.
(F) Ich finde, das Thema Corona wird übertrieben.	(F) I think the Corona topic is exaggerated.
(G) Ich will erst die Erfahrungen von anderen abwarten.	(G) I want to wait for the experiences of others first.
(H) Mein Smartphone ist zu alt.	(H) My smartphone is too old.
(I) Die App würde zu viel Akku verbrauchen.	(I) The app would use too much battery.
(J) Die App produziert Fehlermeldungen.	(J) The app produces error messages.
(K) Ich fühle mich nicht gut genug informiert.	(K) I do not feel well enough informed.
(L) Ich habe die App noch nicht installiert, habe dies aber vor.	(L) I have not yet installed the app, but I intend to.
(M) Andere Gründe, und zwar: [offenes Textfeld]	(M) Other reasons, namely: [open text field]

Frequency of app usage [W3]	
German (original)	English translation
Wie häufig nutzen Sie die App aktiv, indem Sie sie öffnen?	How often do you actively use the app by opening it?
(A) Mehrmals täglich	(A) Several times a day
(B) Etwa einmal am Tag	(B) About once a day
(C) Mehrmals pro Woche	(C) Several times per week
(D) Etwa einmal pro Woche	(D) About once a week
(E) Seltener als einmal pro Woche	(E) Less than once a week
(F) Nie	(F) Never

2.6 Deviations with respect to pre-analysis plan

Although the results we present here are based on the pre-analysis plan we registered prior to obtaining the data (see details at <https://osf.io/6jstp>), there were a few areas in which we had to deviate from the plan. In most cases, these deviations were due to minor errors in questionnaire design or omissions in our pre-analysis plan. For the sake of transparency, we report each of these changes here:

- Due to a mistake in the programming of the questionnaire filters, the item to measure subjects' likelihood of installing the app item was missing in wave 2. Therefore, we could not test the pre-registered hypothesis on the effect of the information treatments on the reported likelihood of installing the app (among previous non-users). The item was included again in wave 3 and could then be used to test longer-term effects (see Fig. 24).
- The smartphone access variable had to be dropped from the list of predictors of the app uptake models because, as a necessary condition for installing the app, it is deterministically related to the outcome.
- Due to a technical mistake by the provider, tracking data on CWA use were only available for users of mobile devices with Android (but not Apple) operating systems. Consequently, we excluded respondents from the survey-plus-tracking and tracking-only panel that had an Apple device registered.
- In addition to the reported risk behavior covariates, we added a measure of compliance with NPIs as an additional covariate in the models and descriptives.
- While we did not pre-register any hypothesis or analysis on time of app adoption, we registered a way to measure it on the basis of the tracking data. We did not pursue this idea further in the analysis and have therefore not constructed such a variable.
- In the incentivization experiment, CACE is only computed for hypotheses on the downstream effects of app usage (H4, H5, and H6 in the pre-analysis plan) and not on the hypotheses on app uptake (H2a, H2b, H3a, H3b). This is a logical consequence of defining actual treatment (compliance of those in treatment assignment group) as having installed the app after agreeing to do so. In the pre-analysis plan, we committed to reporting CACE estimates for tests of H2a, H2b, H3a, H3b, too.
- We deviate slightly from the pre-registered coding scheme of the hybrid app usage variable. Originally, for the survey-plus-tracking sample and in the case of a mismatch between reported ("app installed") and tracked ("app not installed") data, we had planned to give the respondent the benefit of the doubt and code these as "app installed" if no smartphone tracking data was available but respondent had a smartphone. These cases did not exist because tracking panelists were sampled on the basis of having a smartphone registered for tracking. Therefore, we dropped that rule and exclusively relied on the tracking data.

- In the pre-analysis plan we had committed to analyze predictors of risk behavior with regards to COVID-19 exposure. We did not pursue this idea further. Instead, we integrated the analysis of risk behavior in our analyses of app uptake.

3 Supplementary Results

In this section, we report analyses of app usage as well as the results of testing each of our pre-registered hypotheses. If not stated otherwise, we report four quantities of interest:

1. The unadjusted difference in means (Intent-To-Treat estimate or ITT) in the outcome variable between treatment and control groups, as well as the standard error (in parentheses) and a 95% confidence interval (in square brackets).
2. The covariate-adjusted ITT estimate computed using the saturated regression approach (23), as well as the standard error (in parentheses) and a 95% confidence interval (in square brackets). These estimates are reported in the main text.
3. The Complier Average Causal Effect (CACE) estimates. We also report the standard error (in parentheses) and a 95% confidence interval (in square brackets). For all outcomes in the message experiment and for the downstream effects of app usage in the incentivization experiment, we use an instrumental-variables framework (24) to estimate the treatment effects among compliers, or the Complier Average Causal Effect (CACE). In the message experiment, we define as “treated” any respondent who spent at least the length of the treatment video in seconds on the treatment exposure page (125 seconds or longer). 85% of the respondents assigned to one of the treatment groups satisfied this condition. In the incentivization experiment, we define as “treated” any participant who is observed (via the tracking data) to use the app before wave 3 or, for the survey-only respondents, reports to have the app installed in wave 3.
4. The Minimum Detectable Effect (MDE) for the covariate-adjusted ITT estimate, which we compute based on the percentage of variance reduced by the covariate adjustment assuming a test with 80% power for a level of significance of 0.05. We also report in parentheses the equivalent Cohen’s d , computed by dividing the MDE by the standard deviation of the outcome variable. When interpreting this value, we follow the standard convention of assuming that $d \leq 0.20$ corresponds to a small effect.

We employ several different ways to filter our sample to test the hypotheses:

1. We pool both message treatments and compare them to the control condition.
2. We compare the pro-social treatment to the control condition.
3. We compare the self-interest treatment to the control condition.
4. We compare the pro-social treatment to the self-interest treatment.

As a robustness check, we analyze the effects among respondents in the treatment conditions who pass the manipulation check (selecting the correct answer on the item linked to the treatment condition in the knowledge question directly following the treatment; see Section 2.2.4): “Je mehr Menschen die App nutzen, desto besser sind besonders gefährdete Menschen geschützt.” in the pro-social behavior condition and “Wenn viele Menschen die App nutzen, könnte man bald wieder zu einem Alltag fast ohne Einschränkungen übergehen.” in the self-interest condition).

3.1 App usage

We report on tracked vs. reported App adoption among survey-tracking panelists in wave 1 in Supplementary Table 19. The results indicate one-sided misreporting: 18% reported app adopters could not be validated as such in the tracking data. On the other hand, only 3% percent of reported app non-adopters were found to have had used the app at least once on their registered device.

Supplementary Figure 15 provides app adoption rates (both tracked and reported) by sample, indicating a substantially higher adoption rate in the tracking than in the survey-only sample.

Supplementary Figures 16 to 18 provide information on actual or hypothetical behavior related to the App’s functionality. Supplementary Figure 16 indicates that a vast majority of app users have their smartphone’s Bluetooth communication activated always or mostly, which is necessary for the app to register contact with other users. Supplementary Figure 17 reports people’s certainty about getting tested and quarantined for COVID-19 after a risk notification by the app. About 61% indicate they would certainly do so. Supplementary Figure 18 provides people’s certainty about reporting a positive test result in the app. About 82% indicate they would certainly do so.

Supplementary Figure 19 lists reported reasons not to install the app by frequency. Concerns about the effectiveness of the app as well as privacy considerations rank first, followed by skepticism about the state monitoring them and the Bluetooth-on requirement.

Figures 20 and 21 replicate Figure 2 from the main text with additional outcomes and wave 1 vs. wave 3 comparisons.

Finally, Supplementary Figure 22 reports on estimates from a Probit model of tracked, reported, and hybrid measure of app uptake prior to wave 1.

Supplementary Table 19: App adoption prior to wave 1, report-based vs. tracking-based measure. Total counts together with row percentages. Figures based on data from the survey-plus-tracking panel.

		App adoption, tracked		
		No	Yes	Total
App adoption, reported	No	315 (97%)	9 (3%)	324 (51%)
	Yes	54 (18%)	252 (82%)	306 (49%)
Total		369 (59%)	261 (41%)	630 (100%)

Supplementary Figure 15: App adoption rates in wave 1, by sample and measure. Sample sizes: $N_{\text{Survey-only}} = 1,395$, $N_{\text{Survey-tracking}} = 649$, $N_{\text{Tracking-only}} = 482$.

Supplementary Figure 16: Bluetooth activated among tracing app users. Item text: “It could be that you have installed the app but do not use it regularly, e.g. because you do not always want to have Bluetooth technology activated or the smartphone battery runs out quickly. When you are on the road or socializing, how often do you use the Corona Warning App by activating Bluetooth communication on your smartphone?” Sample sizes (subset to those who report having the CWA installed): $N_{\text{Wave 1}} = 725$, $N_{\text{Wave 2}} = 666$, $N_{\text{Wave 3}} = 733$

Supplementary Figure 17: Testing and quarantining after risk encounter. Item text: “Suppose the app informs you that you recently had a risk encounter. How sure are you that you would have yourself tested and quarantined for COVID-19?” Sample sizes (subset to those who report having the CWA installed): $N_{\text{Wave 1}} = 673$, $N_{\text{Wave 2}} = 621$, $N_{\text{Wave 3}} = 657$

Supplementary Figure 18: Reporting positive test result. Item text: “Let’s assume you would test positive for COVID-19. How sure are you that you would report the test result in the app?” Sample sizes (subset to those who report having the CWA installed): $N_{\text{Wave 1}} = 702$, $N_{\text{Wave 2}} = 658$, $N_{\text{Wave 3}} = 720$

Supplementary Figure 19: Reasons not to use tracing app, by frequency (wave 3). Item text “Can you tell us why you did not install the app or uninstalled it again?” Sample size (subset to those who report not having the CWA installed): $N = 749$.

Supplementary Figure 20: Tracked and reported app uptake measured at waves 1 and 3 for survey-tracking panel members (group means with 83% confidence intervals; non-overlapping intervals indicate significant difference at $p < .05$) by socioeconomic strata, risk status and behavior, and attitudes. Overall sample means indicated by black lines (wave 1: solid, wave 3: dashed).

Supplementary Figure 21: Reported Bluetooth usage at waves 1 and 3 and tracked average daily app uses measured 7 days prior to waves 1 and 3 for survey-tracking panel members (group means with 83% confidence intervals; non-overlapping intervals indicate significant difference at $p < .05$) by socioeconomic strata, risk status and behavior, and attitudes. Overall sample means indicated by black lines (wave 1: solid, wave 3: dashed).

Supplementary Figure 22: Determinants of app uptake. Probit model of tracked ($N = 559$), reported ($N = 1,774$), and hybrid ($N = 1,774$) measure of app usage prior to wave 1. Unstandardized coefficients along with 95% confidence intervals reported.

3.2 Message experiment

3.2.1 Replicated main results with W3 outcomes & conditional on passing the manipulation checks

Supplementary Figure 23: Summary of effects (with 95% confidence intervals) of video treatments on knowledge, attitudes, and behaviors. Replication of results in main text, **restricted to those who passed the manipulation check**. The plot shows standardized ITT estimates with 95% confidence intervals from fully saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. The video message sample in wave 1 subset to those who passed the manipulation check comprises $n = 1,645/1,210/1,082$ respondents for the estimation of the pooled/pro-social/self-interest treatment effects respectively.

Supplementary Figure 24: Summary of effects (with 95% confidence intervals) of video treatments on knowledge, attitudes, and behaviors. Replication of results in main text **using wave 3 outcome measures**. The plot shows standardized ITT estimates with 95% confidence intervals from fully saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. The video message sample in wave 3 comprises $n = 1,581/1,058/1,025$ respondents for the estimation of the pooled/pro-social/self-interest treatment effects respectively.

Supplementary Figure 25: Summary of effects (with 95% confidence intervals) of video treatments on knowledge, attitudes, and behaviors. Replication of results in main text, **restricted to those who passed the manipulation check and using wave 3 outcome measures**. The plot shows standardized ITT estimates with 95% confidence intervals from fully saturated OLS regression models fit using the pre-registered LASSO covariate selection procedure. The video message sample in wave 3 subset to those who passed the manipulation check in wave 1 comprises $n = 1,286/951/837$ respondents for the estimation of the pooled/pro-social/self-interest treatment effects respectively.

3.2.2 Hypothesis tests (pre-registered)

Supplementary Table 20: Results: Hypothesis 1 – Subjects assigned to any of the treatments will be more likely to install the app than subjects assigned to the control group. Survey-plus-tracking sample only. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (tracked) (Pooled vs. control)	0.023 (0.091) [-0.157, 0.202]	-0.010 (0.021) [-0.052, 0.032]	-0.012 (0.024) [-0.060, 0.036]	0.038 ($d=0.037$)
Uptake (tracked) (Prosocial vs. control)	0.102 (0.107) [-0.107, 0.312]	-0.008 (0.024) [-0.056, 0.040]	-0.009 (0.027) [-0.063, 0.045]	0.049 ($d=0.047$)
Uptake (tracked) (Self-interest vs. control)	-0.052 (0.104) [-0.256, 0.152]	-0.013 (0.024) [-0.060, 0.035]	-0.014 (0.027) [-0.068, 0.039]	0.047 ($d=0.046$)
Uptake (tracked) (Prosocial vs. self-interest)	0.155 (0.105) [-0.052, 0.361]	0.004 (0.022) [-0.039, 0.047]	0.452 (2.914) [-5.277, 6.181]	0.043 ($d=0.042$)

Supplementary Table 21: Results: Hypothesis 1 – Subjects assigned to any of the treatments will be more likely to install the app than subjects assigned to the control group. Survey-plus-tracking sample only; restricted to those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (tracked) (Pooled vs. control)	0.106 (0.096) [-0.082, 0.294]	-0.009 (0.022) [-0.052, 0.034]	-0.010 (0.024) [-0.058, 0.038]	0.041 ($d=0.040$)
Uptake (tracked) (Prosocial vs. control)	0.205 (0.112) [-0.014, 0.425]	-0.015 (0.023) [-0.060, 0.030]	-0.016 (0.025) [-0.065, 0.032]	0.047 ($d=0.046$)
Uptake (tracked) (Self-interest vs. control)	-0.004 (0.114) [-0.228, 0.221]	-0.003 (0.028) [-0.058, 0.052]	-0.004 (0.030) [-0.063, 0.056]	0.055 ($d=0.054$)
Uptake (tracked) (Prosocial vs. self-interest)	0.209 (0.120) [-0.027, 0.445]	-0.012 (0.024) [-0.059, 0.035]	0.491 (1.165) [-1.801, 2.783]	0.048 ($d=0.046$)

Supplementary Table 22: Results: Hypothesis 1 – Subjects assigned to any of the treatments will be more likely to install the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; hybrid outcome measure. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (tracked) (Pooled vs. control)	0.125 (0.065) [-0.003, 0.253]	0.008 (0.013) [-0.018, 0.034]	0.009 (0.015) [-0.020, 0.038]	0.023 ($d=0.023$)
Uptake (tracked) (Prosocial vs. control)	0.205 (0.085) [0.037, 0.373]	0.011 (0.019) [-0.027, 0.048]	0.011 (0.020) [-0.028, 0.050]	0.025 ($d=0.025$)
Uptake (tracked) (Self-interest vs. control)	0.051 (0.082) [-0.111, 0.212]	0.006 (0.017) [-0.026, 0.039]	0.007 (0.019) [-0.029, 0.044]	0.025 ($d=0.025$)
Uptake (tracked) (Prosocial vs. self-interest)	0.155 (0.105) [-0.052, 0.361]	0.004 (0.022) [-0.039, 0.047]	0.452 (2.914) [-5.277, 6.181]	0.043 ($d=0.042$)

Supplementary Table 23: Results: Hypothesis 1 – Subjects assigned to any of the treatments will be more likely to install the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; hybrid outcome measure; restricted to those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (tracked) (Pooled vs. control)	0.106 (0.096) [-0.082, 0.294]	-0.009 (0.022) [-0.052, 0.034]	-0.010 (0.024) [-0.058, 0.038]	0.041 ($d=0.040$)
Uptake (tracked) (Prosocial vs. control)	0.205 (0.112) [-0.014, 0.425]	-0.015 (0.023) [-0.060, 0.030]	-0.016 (0.025) [-0.065, 0.032]	0.047 ($d=0.046$)
Uptake (tracked) (Self-interest vs. control)	-0.004 (0.114) [-0.228, 0.221]	-0.003 (0.028) [-0.058, 0.052]	-0.004 (0.030) [-0.063, 0.056]	0.055 ($d=0.054$)
Uptake (tracked) (Prosocial vs. self-interest)	0.209 (0.120) [-0.027, 0.445]	-0.012 (0.024) [-0.059, 0.035]	0.491 (1.165) [-1.801, 2.783]	0.048 ($d=0.046$)

Supplementary Table 24: Results: Hypothesis 1 – Subjects assigned to any of the treatments will be more likely to install the app than subjects assigned to the control group. Survey-plus-tracking and tracking-only samples combined; reduced set of controls. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (hybrid) (Pooled vs. control)	0.063 (0.042) [-0.020, 0.146]	0.031 (0.015) [0.002, 0.061]	0.036 (0.017) [0.002, 0.070]	0.030 ($d=0.030$)
Uptake (hybrid) (Prosocial vs. control)	0.059 (0.051) [-0.041, 0.159]	0.024 (0.018) [-0.011, 0.058]	0.028 (0.020) [-0.013, 0.068]	0.032 ($d=0.031$)
Uptake (hybrid) (Self-interest vs. control)	0.067 (0.052) [-0.035, 0.168]	0.039 (0.021) [-0.002, 0.080]	0.044 (0.024) [-0.002, 0.091]	0.036 ($d=0.035$)
Uptake (hybrid) (Prosocial vs. self-interest)	-0.008 (0.058) [-0.122, 0.107]	-0.015 (0.024) [-0.061, 0.032]	0.780 (1.489) [-2.141, 3.700]	0.047 ($d=0.046$)

Supplementary Table 25: Results: Hypothesis 1 – Subjects assigned to any of the treatments will be more likely to install the app than subjects assigned to the control group. Survey-plus-tracking and tracking-only samples combined; reduced set of controls; subsetting on those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (hybrid) (Pooled vs. control)	0.159 (0.055) [0.052, 0.267]	0.028 (0.021) [-0.013, 0.069]	0.031 (0.023) [-0.014, 0.076]	0.040 ($d=0.039$)
Uptake (hybrid) (Prosocial vs. control)	0.182 (0.063) [0.059, 0.305]	0.029 (0.023) [-0.016, 0.074]	0.033 (0.026) [-0.018, 0.084]	0.045 ($d=0.044$)
Uptake (hybrid) (Self-interest vs. control)	0.130 (0.068) [-0.003, 0.262]	0.027 (0.027) [-0.027, 0.080]	0.028 (0.029) [-0.029, 0.086]	0.051 ($d=0.050$)
Uptake (hybrid) (Prosocial vs. self-interest)	0.052 (0.070) [-0.086, 0.190]	0.005 (0.026) [-0.047, 0.056]	-0.101 (0.613) [-1.305, 1.102]	0.050 ($d=0.049$)

Supplementary Table 26: Results: Hypothesis 2 – Subjects assigned to any of the treatments will be more likely to report using the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (reported) (Pooled vs. control)	0.057 (0.051) [-0.043, 0.157]	0.023 (0.021) [-0.019, 0.064]	0.027 (0.024) [-0.021, 0.075]	0.041 ($d=0.041$)
Uptake (reported) (Prosocial vs. control)	0.036 (0.058) [-0.078, 0.150]	0.011 (0.024) [-0.036, 0.058]	0.013 (0.028) [-0.041, 0.067]	0.047 ($d=0.047$)
Uptake (reported) (Self-interest vs. control)	0.079 (0.059) [-0.037, 0.195]	0.035 (0.026) [-0.016, 0.086]	0.040 (0.030) [-0.018, 0.099]	0.051 ($d=0.051$)
Uptake (reported) (Prosocial vs. self-interest)	-0.043 (0.058) [-0.157, 0.071]	-0.024 (0.026) [-0.075, 0.027]	1.353 (2.057) [-2.682, 5.387]	0.051 ($d=0.051$)

Supplementary Table 27: Results: Hypothesis 2 – Subjects assigned to any of the treatments will be more likely to report using the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; restricted to those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Uptake (reported) (Pooled vs. control)	0.179 (0.054) [0.072, 0.286]	0.029 (0.022) [-0.014, 0.073]	0.033 (0.025) [-0.016, 0.081]	0.043 ($d=0.042$)
Uptake (reported) (Prosocial vs. control)	0.176 (0.062) [0.054, 0.299]	0.028 (0.025) [-0.022, 0.077]	0.031 (0.029) [-0.025, 0.088]	0.049 ($d=0.049$)
Uptake (reported) (Self-interest vs. control)	0.183 (0.067) [0.051, 0.314]	0.032 (0.028) [-0.024, 0.087]	0.034 (0.030) [-0.026, 0.094]	0.053 ($d=0.052$)
Uptake (reported) (Prosocial vs. self-interest)	-0.006 (0.070) [-0.143, 0.131]	-0.003 (0.029) [-0.060, 0.053]	0.081 (0.688) [-1.270, 1.432]	0.056 ($d=0.055$)

Supplementary Table 28: Results: Hypothesis 3 – Subjects assigned to any of the treatments will be more likely to report having their smartphone’s Bluetooth function activated than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Bluetooth active (reported) (Pooled vs. control)	0.084 (0.076) [-0.065, 0.233]	0.069 (0.050) [-0.029, 0.167]	0.080 (0.057) [-0.033, 0.192]	0.090 ($d=0.105$)
Bluetooth active (reported) (Prosocial vs. control)	0.161 (0.080) [0.004, 0.318]	0.130 (0.055) [0.023, 0.238]	0.144 (0.061) [0.024, 0.264]	0.107 ($d=0.132$)
Bluetooth active (reported) (Self-interest vs. control)	0.007 (0.089) [-0.169, 0.183]	0.009 (0.061) [-0.111, 0.129]	0.010 (0.067) [-0.122, 0.143]	0.120 ($d=0.130$)
Bluetooth active (reported) (Prosocial vs. self-interest)	0.154 (0.077) [0.003, 0.304]	0.121 (0.053) [0.017, 0.224]	-51.949 (613.310) [-1257.440, 1153.543]	0.104 ($d=0.126$)

Supplementary Table 29: Results: Hypothesis 3 – Subjects assigned to any of the treatments will be more likely to report having their smartphone’s Bluetooth function activated than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; restricted to those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Bluetooth active (reported) (Pooled vs. control)	0.166 (0.075) [0.020, 0.313]	0.095 (0.050) [-0.004, 0.193]	0.109 (0.057) [-0.003, 0.221]	0.094 ($d=0.121$)
Bluetooth active (reported) (Prosocial vs. control)	0.188 (0.080) [0.031, 0.345]	0.138 (0.055) [0.029, 0.246]	0.151 (0.061) [0.030, 0.271]	0.109 ($d=0.136$)
Bluetooth active (reported) (Self-interest vs. control)	0.138 (0.086) [-0.031, 0.307]	0.049 (0.066) [-0.081, 0.178]	0.046 (0.069) [-0.090, 0.181]	0.127 ($d=0.153$)
Bluetooth active (reported) (Prosocial vs. self-interest)	0.050 (0.071) [-0.090, 0.190]	0.086 (0.057) [-0.027, 0.199]	-2.594 (2.676) [-7.856, 2.669]	0.107 ($d=0.156$)

Supplementary Table 30: Results: Hypothesis 5 – Subjects assigned to any of the treatments will show higher levels of knowledge about the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Knowledge (Pooled vs. control)	0.277 (0.051) [0.176, 0.377]	0.255 (0.048) [0.160, 0.349]	0.294 (0.055) [0.185, 0.402]	0.090 ($d=0.088$)
Knowledge (Prosocial vs. control)	0.238 (0.059) [0.123, 0.352]	0.213 (0.055) [0.105, 0.321]	0.248 (0.064) [0.124, 0.373]	0.110 ($d=0.108$)
Knowledge (Self-interest vs. control)	0.317 (0.059) [0.201, 0.434]	0.299 (0.057) [0.189, 0.410]	0.342 (0.064) [0.217, 0.467]	0.112 ($d=0.110$)
Knowledge (Prosocial vs. self-interest)	-0.080 (0.058) [-0.194, 0.035]	-0.074 (0.055) [-0.182, 0.034]	6.534 (11.998) [-17.008, 30.075]	0.109 ($d=0.107$)

Supplementary Table 31: Results: Hypothesis 5 – Subjects assigned to any of the treatments will show higher levels of knowledge about the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; restricted to those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Knowledge (Pooled vs. control)	0.492 (0.052) [0.390, 0.595]	0.433 (0.049) [0.337, 0.530]	0.483 (0.055) [0.375, 0.590]	0.096 ($d=0.098$)
Knowledge (Prosocial vs. control)	0.447 (0.059) [0.331, 0.562]	0.397 (0.056) [0.287, 0.507]	0.446 (0.063) [0.323, 0.570]	0.113 ($d=0.115$)
Knowledge (Self-interest vs. control)	0.552 (0.064) [0.426, 0.677]	0.495 (0.061) [0.376, 0.614]	0.536 (0.065) [0.409, 0.664]	0.122 ($d=0.120$)
Knowledge (Prosocial vs. self-interest)	-0.105 (0.063) [-0.229, 0.019]	-0.097 (0.060) [-0.215, 0.022]	3.419 (3.064) [-2.595, 9.433]	0.118 ($d=0.128$)

Supplementary Table 32: Results: Hypothesis 6 – Subjects assigned to any of the treatments will report more positive attitudes towards the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Positive attitudes (Pooled vs. control)	0.132 (0.052) [0.030, 0.235]	0.106 (0.031) [0.045, 0.166]	0.122 (0.036) [0.052, 0.192]	0.058 ($d=0.056$)
Positive attitudes (Prosocial vs. control)	0.074 (0.060) [-0.043, 0.192]	0.068 (0.035) [-0.001, 0.138]	0.080 (0.041) [-0.001, 0.160]	0.070 ($d=0.068$)
Positive attitudes (Self-interest vs. control)	0.193 (0.060) [0.075, 0.310]	0.144 (0.036) [0.073, 0.214]	0.164 (0.041) [0.084, 0.244]	0.072 ($d=0.070$)
Positive attitudes (Prosocial vs. self-interest)	-0.119 (0.059) [-0.234, -0.003]	-0.072 (0.036) [-0.142, -0.002]	6.941 (13.140) [-18.840, 32.721]	0.071 ($d=0.069$)

Supplementary Table 33: Results: Hypothesis 6 – Subjects assigned to any of the treatments will report more positive attitudes towards the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; subsetting on those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Positive attitudes (Pooled vs. control)	0.297 (0.054) [0.190, 0.403]	0.145 (0.032) [0.081, 0.208]	0.162 (0.036) [0.092, 0.232]	0.063 ($d=0.063$)
Positive attitudes (Prosocial vs. control)	0.240 (0.061) [0.120, 0.360]	0.108 (0.037) [0.035, 0.180]	0.120 (0.042) [0.038, 0.202]	0.074 ($d=0.073$)
Positive attitudes (Self-interest vs. control)	0.370 (0.067) [0.240, 0.501]	0.194 (0.040) [0.115, 0.274]	0.210 (0.043) [0.125, 0.295]	0.079 ($d=0.077$)
Positive attitudes (Prosocial vs. self-interest)	-0.130 (0.067) [-0.262, 0.001]	-0.093 (0.042) [-0.175, -0.012]	2.489 (1.662) [-0.774, 5.751]	0.082 ($d=0.083$)

Supplementary Table 34: Results: Hypothesis 7 – Subjects assigned to any of the treatments will be more likely to intend to share a message about the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Message sharing (Pooled vs. control)	-0.053 (0.043) [-0.138, 0.032]	-0.053 (0.043) [-0.138, 0.032]	-0.062 (0.051) [-0.162, 0.038]	0.070 ($d=0.088$)
Message sharing (Prosocial vs. control)	-0.068 (0.043) [-0.153, 0.016]	-0.068 (0.043) [-0.153, 0.016]	-0.081 (0.051) [-0.180, 0.019]	0.083 ($d=0.108$)
Message sharing (Self-interest vs. control)	-0.037 (0.051) [-0.137, 0.063]	-0.037 (0.051) [-0.137, 0.063]	-0.043 (0.059) [-0.159, 0.073]	0.100 ($d=0.108$)
Message sharing (Prosocial vs. self-interest)	-0.032 (0.037) [-0.104, 0.041]	-0.032 (0.037) [-0.104, 0.041]	2.590 (5.118) [-7.450, 12.629]	0.072 ($d=0.106$)

Supplementary Table 35: Results: Hypothesis 7 – Subjects assigned to any of the treatments will be more likely to intend to share a message about the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; subsetting on those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Message sharing (Pooled vs. control)	-0.063 (0.044) [-0.149, 0.022]	-0.063 (0.044) [-0.148, 0.022]	-0.070 (0.049) [-0.166, 0.025]	0.076 ($d=0.098$)
Message sharing (Prosocial vs. control)	-0.086 (0.041) [-0.167, -0.005]	-0.086 (0.041) [-0.167, -0.005]	-0.098 (0.047) [-0.190, -0.006]	0.087 ($d=0.114$)
Message sharing (Self-interest vs. control)	-0.033 (0.056) [-0.143, 0.076]	-0.033 (0.056) [-0.143, 0.076]	-0.037 (0.061) [-0.157, 0.084]	0.113 ($d=0.121$)
Message sharing (Prosocial vs. self-interest)	-0.053 (0.042) [-0.135, 0.029]	-0.053 (0.042) [-0.134, 0.029]	1.548 (1.494) [-1.384, 4.480]	0.074 ($d=0.125$)

Supplementary Table 36: Results: Hypothesis 8 – Subjects assigned to any of the treatments will be more likely to intend to look up additional information about the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Information lookup (Pooled vs. control)	0.069 (0.055) [-0.038, 0.176]	0.069 (0.055) [-0.038, 0.176]	0.080 (0.064) [-0.045, 0.206]	0.114 ($d=0.088$)
Information lookup (Prosocial vs. control)	0.098 (0.071) [-0.040, 0.237]	0.098 (0.071) [-0.040, 0.237]	0.116 (0.083) [-0.047, 0.279]	0.142 ($d=0.108$)
Information lookup (Self-interest vs. control)	0.038 (0.062) [-0.083, 0.159]	0.038 (0.062) [-0.083, 0.159]	0.045 (0.072) [-0.096, 0.186]	0.123 ($d=0.108$)
Information lookup (Prosocial vs. self-interest)	0.060 (0.075) [-0.088, 0.208]	0.060 (0.075) [-0.088, 0.208]	-4.914 (9.611) [-23.768, 13.940]	0.150 ($d=0.106$)

Supplementary Table 37: Results: Hypothesis 8 – Subjects assigned to any of the treatments will be more likely to intend to look up additional information about the app than subjects assigned to the control group. Survey-only and Survey-plus-tracking samples combined; subsetting on those who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Information lookup (Pooled vs. control)	0.050 (0.057) [-0.062, 0.162]	0.050 (0.057) [-0.062, 0.162]	0.055 (0.064) [-0.070, 0.181]	0.117 ($d=0.098$)
Information lookup (Prosocial vs. control)	0.115 (0.079) [-0.041, 0.270]	0.115 (0.079) [-0.041, 0.270]	0.130 (0.090) [-0.046, 0.306]	0.152 ($d=0.114$)
Information lookup (Self-interest vs. control)	-0.034 (0.052) [-0.136, 0.067]	-0.034 (0.052) [-0.136, 0.067]	-0.038 (0.057) [-0.149, 0.074]	0.108 ($d=0.121$)
Information lookup (Prosocial vs. self-interest)	0.149 (0.077) [-0.001, 0.299]	0.149 (0.077) [-0.001, 0.299]	-4.391 (3.312) [-10.889, 2.107]	0.165 ($d=0.125$)

3.2.3 Hypothesis tests (not pre-registered)

Supplementary Table 38: Results (not pre-registered) – Subjects assigned to any of the treatments will be more likely to report getting tested and quarantine when the app informs them about a recent risk encounter. Survey-only and Survey-plus-tracking samples combined, sub-setting on those who report having the app installed before treatment. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Tested after warning (Pooled vs. control)	0.086 (0.086) [-0.083, 0.255]	0.132 (0.075) [-0.016, 0.279]	0.138 (0.080) [-0.019, 0.295]	0.126 ($d=0.136$)
Tested after warning (Prosocial vs. control)	0.036 (0.097) [-0.154, 0.226]	0.102 (0.083) [-0.060, 0.264]	0.113 (0.092) [-0.068, 0.293]	0.159 ($d=0.166$)
Tested after warning (Self-interest vs. control)	0.137 (0.095) [-0.049, 0.324]	0.158 (0.083) [-0.004, 0.321]	0.167 (0.088) [-0.005, 0.340]	0.160 ($d=0.170$)
Tested after warning (Prosocial vs. self-interest)	-0.101 (0.085) [-0.269, 0.066]	-0.058 (0.074) [-0.202, 0.087]	1.335 (1.851) [-2.305, 4.974]	0.145 ($d=0.164$)

Supplementary Table 39: Results (not pre-registered) – Subjects assigned to any of the treatments will be more likely to report getting tested and quarantine when the app informs them about a recent risk encounter. Survey-only and Survey-plus-tracking samples combined, subsetting on those who report having the app installed before treatment and who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Tested after warning (Pooled vs. control)	0.100 (0.089) [-0.074, 0.275]	0.133 (0.076) [-0.016, 0.283]	0.138 (0.080) [-0.019, 0.296]	0.132 ($d=0.142$)
Tested after warning (Prosocial vs. control)	0.065 (0.098) [-0.128, 0.258]	0.106 (0.083) [-0.057, 0.269]	0.116 (0.091) [-0.063, 0.295]	0.160 ($d=0.167$)
Tested after warning (Self-interest vs. control)	0.146 (0.102) [-0.055, 0.347]	0.164 (0.086) [-0.006, 0.334]	0.171 (0.090) [-0.007, 0.348]	0.171 ($d=0.180$)
Tested after warning (Prosocial vs. self-interest)	-0.081 (0.094) [-0.266, 0.103]	-0.060 (0.079) [-0.214, 0.095]	1.103 (1.526) [-1.899, 4.105]	0.154 ($d=0.174$)

Supplementary Table 40: Results (not pre-registered) – Subjects assigned to any of the treatments will be more likely to report a positive test result in the app. Survey-only and Survey-plus-tracking samples combined, restricted to those who report having the app installed before treatment. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Reporting positive test (Pooled vs. control)	0.071 (0.091) [-0.108, 0.250]	0.071 (0.091) [-0.107, 0.249]	0.079 (0.101) [-0.119, 0.277]	0.155 ($d=0.155$)
Reporting positive test (Prosocial vs. control)	0.043 (0.104) [-0.162, 0.248]	0.043 (0.104) [-0.162, 0.248]	0.048 (0.117) [-0.181, 0.277]	0.204 ($d=0.192$)
Reporting positive test (Self-interest vs. control)	0.100 (0.098) [-0.092, 0.292]	0.100 (0.098) [-0.092, 0.292]	0.110 (0.108) [-0.102, 0.321]	0.190 ($d=0.193$)
Reporting positive test (Prosocial vs. self-interest)	-0.057 (0.088) [-0.230, 0.116]	-0.057 (0.088) [-0.230, 0.116]	3.702 (8.345) [-12.698, 20.102]	0.175 ($d=0.186$)

Supplementary Table 41: Results (not pre-registered) – Subjects assigned to any of the treatments will be more likely to report a positive test result in the app. Survey-only and Survey-plus-tracking samples combined, restricted to those who report having the app installed before treatment and who passed manipulation check. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Reporting positive test (Pooled vs. control)	0.086 (0.093) [-0.096, 0.269]	0.086 (0.093) [-0.096, 0.268]	0.093 (0.100) [-0.103, 0.290]	0.163 ($d=0.166$)
Reporting positive test (Prosocial vs. control)	0.069 (0.103) [-0.134, 0.273]	0.069 (0.103) [-0.134, 0.272]	0.076 (0.114) [-0.148, 0.300]	0.203 ($d=0.195$)
Reporting positive test (Self-interest vs. control)	0.109 (0.103) [-0.093, 0.312]	0.109 (0.103) [-0.093, 0.312]	0.115 (0.108) [-0.098, 0.328]	0.209 ($d=0.210$)
Reporting positive test (Prosocial vs. self-interest)	-0.040 (0.092) [-0.221, 0.140]	-0.040 (0.092) [-0.221, 0.140]	0.942 (2.189) [-3.362, 5.245]	0.184 ($d=0.204$)

3.2.4 Effect heterogeneity

Supplementary Table 42: Tests of effect heterogeneity by pre-treatment app use, age, and NPI compliance. Effects are computed in the covariate-adjusted ITT model.

Outcome	Any message treatment interacted with								
	App use			Age			NPI compliance		
	B	95% CI	P	B	95% CI	P			
Uptake (tracked)				0.01	[-0.00, 0.02]	0.184	-0.02	[-0.09, 0.05]	0.593
Uptake (hybrid)				0.00	[-0.02, 0.02]	0.909	-0.01	[-0.11, 0.09]	0.804
Uptake (reported)				-0.00	[-0.03, 0.02]	0.740	-0.02	[-0.13, 0.09]	0.742
Knowledge	-0.06	[-0.26, 0.13]	0.519	-0.08	[-0.14, -0.01]	0.018	0.05	[-0.19, 0.29]	0.681
Positive attitudes	-0.04	[-0.21, 0.12]	0.589	-0.03	[-0.07, 0.01]	0.139	0.08	[-0.08, 0.23]	0.312
Message sharing (tracked)	0.08	[-0.07, 0.23]	0.286	-0.03	[-0.09, 0.04]	0.427	-0.10	[-0.21, 0.02]	0.109
Information lookup (tracked)	-0.02	[-0.20, 0.16]	0.841	0.02	[-0.05, 0.09]	0.536	0.16	[-0.09, 0.41]	0.201
Bluetooth active (reported)	0.01	[-0.39, 0.41]	0.963	-0.02	[-0.09, 0.04]	0.537	0.12	[-0.14, 0.38]	0.376
Tested after warning	-0.06	[-0.60, 0.49]	0.840	-0.10	[-0.21, 0.00]	0.052	0.02	[-0.53, 0.57]	0.931
Reporting positive test	0.32	[-0.27, 0.90]	0.286	0.03	[-0.08, 0.14]	0.593	-0.62	[-1.33, 0.08]	0.082

Supplementary Table 43: Tests of effect heterogeneity by pre-treatment government trust, data privacy attitudes, and COVID threat perceptions for family and friends. Effects are computed in the covariate-adjusted ITT model.

Outcome	Any message treatment interacted with								
	Gov. trust			Data control			Threat perception		
	B	95% CI	P	B	95% CI	P			
Uptake (tracked)	-0.04	[-0.20, 0.13]	0.664	0.05	[-0.04, 0.14]	0.305	0.01	[-0.08, 0.10]	0.849
Uptake (hybrid)	-0.06	[-0.19, 0.06]	0.319	0.01	[-0.07, 0.09]	0.821	-0.07	[-0.15, 0.02]	0.114
Uptake (reported)	-0.05	[-0.18, 0.08]	0.432	-0.01	[-0.10, 0.07]	0.794	-0.03	[-0.12, 0.06]	0.463
Knowledge	-0.21	[-0.95, 0.54]	0.583	0.11	[-0.15, 0.37]	0.405	-0.31	[-0.69, 0.07]	0.108
Positive attitudes	-0.16	[-0.64, 0.31]	0.500	-0.03	[-0.15, 0.10]	0.656	-0.06	[-0.18, 0.07]	0.385
Message sharing (tracked)	0.11	[-0.10, 0.32]	0.289	-0.16	[-0.32, 0.01]	0.066	0.01	[-0.17, 0.20]	0.890
Information lookup (tracked)	-0.05	[-0.32, 0.22]	0.720	-0.22	[-0.43, -0.01]	0.042	-0.08	[-0.30, 0.15]	0.511
Bluetooth active (reported)	-0.21	[-0.58, 0.16]	0.258	0.17	[-0.04, 0.38]	0.106	0.10	[-0.13, 0.32]	0.391
Tested after warning	0.09	[-0.40, 0.58]	0.713	0.12	[-0.15, 0.39]	0.374	0.21	[-0.16, 0.58]	0.271
Reporting positive test	-0.05	[-0.79, 0.69]	0.895	0.04	[-0.32, 0.39]	0.837	0.12	[-0.32, 0.57]	0.590

3.3 Incentivization experiment

3.3.1 Hypothesis tests (pre-registered)

Supplementary Table 44: Results: Incentivization hypotheses 2a and 2b – Subjects in any of the incentivization groups will be more likely to install the app than subjects assigned to the control group. The differences between incentivized subjects and the control group will be larger for subjects offered higher monetary incentives. Tracked app usage measure. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	MDE
Uptake (tracked)	1.387 (0.192)	1.050 (0.172)	0.262
(Pooled vs. control)	[1.009, 1.766]	[0.713, 1.387]	($d=0.137$)
Uptake (tracked)	1.537 (0.339)	0.830 (0.284)	0.205
(EUR 1 vs. control)	[0.861, 2.212]	[0.271, 1.389]	($d=0.134$)
Uptake (tracked)	1.101 (0.280)	0.980 (0.262)	0.230
(EUR 2 vs. control)	[0.543, 1.658]	[0.465, 1.496]	($d=0.157$)
Uptake (tracked)	1.616 (0.371)	1.435 (0.356)	0.242
(EUR 5 vs. control)	[0.873, 2.358]	[0.734, 2.135]	($d=0.162$)
Uptake (tracked)	-0.436 (0.434)	0.127 (0.360)	0.713
(EUR 2 vs. EUR 1)	[-1.294, 0.421]	[-0.585, 0.839]	($d=0.270$)
Uptake (tracked)	0.079 (0.497)	0.541 (0.417)	0.831
(EUR 5 vs. EUR 1)	[-0.906, 1.063]	[-0.284, 1.367]	($d=0.298$)
Uptake (tracked)	0.515 (0.459)	0.515 (0.459)	0.883
(EUR 5 vs. EUR 2)	[-0.395, 1.425]	[-0.393, 1.423]	($d=0.334$)

Supplementary Table 45: Results: Incentivization hypotheses 2a and 2b – Subjects in any of the incentivization groups will be more likely to install the app than subjects assigned to the control group. The differences between incentivized subjects and the control group will be larger for subjects offered higher monetary incentives. Hybrid app usage measure. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	MDE
Uptake (hybrid)	0.910 (0.095)	0.838 (0.089)	0.194
(Pooled vs. control)	[0.724, 1.096]	[0.664, 1.013]	($d=0.107$)
Uptake (hybrid)	0.968 (0.155)	0.802 (0.144)	0.197
(EUR 1 vs. control)	[0.663, 1.273]	[0.520, 1.084]	($d=0.128$)
Uptake (hybrid)	0.682 (0.139)	0.664 (0.133)	0.188
(EUR 2 vs. control)	[0.409, 0.956]	[0.403, 0.925]	($d=0.134$)
Uptake (hybrid)	1.092 (0.163)	1.092 (0.163)	0.228
(EUR 5 vs. control)	[0.771, 1.413]	[0.772, 1.412]	($d=0.145$)
Uptake (hybrid)	-0.285 (0.199)	-0.143 (0.186)	0.368
(EUR 2 vs. EUR 1)	[-0.676, 0.105]	[-0.508, 0.223]	($d=0.175$)
Uptake (hybrid)	0.124 (0.216)	0.257 (0.204)	0.406
(EUR 5 vs. EUR 1)	[-0.301, 0.550]	[-0.145, 0.658]	($d=0.180$)
Uptake (hybrid)	0.410 (0.205)	0.410 (0.205)	0.406
(EUR 5 vs. EUR 2)	[0.006, 0.813]	[0.006, 0.813]	($d=0.189$)

Supplementary Table 46: Results: Incentivization hypotheses 3a and 3b – Subjects in any of the incentivization groups will be more likely to report using the app than subjects assigned to the control group. The differences between incentivized subjects and the control group will be larger for subjects offered higher monetary incentives. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	MDE
Uptake (reported)	0.837 (0.101)	0.837 (0.101)	0.238
(Pooled vs. control)	[0.639, 1.035]	[0.639, 1.035]	($d=0.133$)
Uptake (reported)	0.886 (0.148)	0.886 (0.148)	0.306
(EUR 1 vs. control)	[0.594, 1.177]	[0.594, 1.177]	($d=0.188$)
Uptake (reported)	0.680 (0.139)	0.680 (0.139)	0.284
(EUR 2 vs. control)	[0.407, 0.953]	[0.407, 0.952]	($d=0.187$)
Uptake (reported)	0.954 (0.153)	0.954 (0.153)	0.313
(EUR 5 vs. control)	[0.652, 1.255]	[0.653, 1.255]	($d=0.190$)
Uptake (reported)	-0.206 (0.180)	-0.206 (0.180)	0.356
(EUR 2 vs. EUR 1)	[-0.559, 0.147]	[-0.559, 0.147]	($d=0.187$)
Uptake (reported)	0.068 (0.191)	0.068 (0.191)	0.378
(EUR 5 vs. EUR 1)	[-0.307, 0.443]	[-0.307, 0.443]	($d=0.190$)
Uptake (reported)	0.274 (0.184)	0.274 (0.184)	0.363
(EUR 5 vs. EUR 2)	[-0.087, 0.635]	[-0.087, 0.635]	($d=0.189$)

Supplementary Table 47: Results: Incentivization hypothesis 4 – Subjects in any of the incentivization groups will show higher levels of knowledge about the app than subjects assigned to the control group. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Knowledge (Pooled vs. control)	0.145 (0.079) [-0.010, 0.299]	0.155 (0.063) [0.032, 0.279]	1.135 (0.465) [0.221, 2.048]	0.112 ($d=0.108$)
Knowledge (EUR 1 vs. control)	0.205 (0.097) [0.014, 0.395]	0.143 (0.079) [-0.012, 0.298]	0.996 (0.548) [-0.081, 2.073]	0.156 ($d=0.151$)
Knowledge (EUR 2 vs. control)	0.010 (0.098) [-0.182, 0.202]	0.148 (0.083) [-0.015, 0.310]	1.664 (0.985) [-0.272, 3.599]	0.165 ($d=0.160$)
Knowledge (EUR 5 vs. control)	0.225 (0.096) [0.037, 0.414]	0.223 (0.078) [0.069, 0.377]	1.293 (0.481) [0.347, 2.238]	0.157 ($d=0.156$)

Supplementary Table 48: Results: Incentivization hypothesis 4 (test with alternative knowledge measure - not preregistered) – Subjects in any of the incentivization groups will show higher levels of applied knowledge about the app than subjects assigned to the control group. Applied knowledge is measured with two items to test respondents' recognition of (a) the app icon among six alternative icons and (b) a screenshot of the app among six alternative screenshots. The outcome is constructed as the sum of correct answers (from 0 to 2). The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Knowledge (icon, screenshot) (Pooled vs. control)	0.235 (0.055) [0.126, 0.344]	0.195 (0.045) [0.106, 0.284]	1.427 (0.344) [0.753, 2.101]	0.084 ($d=0.114$)
Knowledge (icon, screenshot) (EUR 1 vs. control)	0.318 (0.070) [0.180, 0.456]	0.276 (0.059) [0.160, 0.391]	1.827 (0.454) [0.935, 2.719]	0.118 ($d=0.160$)
Knowledge (icon, screenshot) (EUR 2 vs. control)	0.129 (0.068) [-0.004, 0.263]	0.111 (0.055) [0.002, 0.219]	1.119 (0.564) [0.010, 2.228]	0.110 ($d=0.157$)
Knowledge (icon, screenshot) (EUR 5 vs. control)	0.258 (0.070) [0.121, 0.395]	0.198 (0.059) [0.082, 0.314]	1.221 (0.353) [0.527, 1.916]	0.117 ($d=0.163$)

Supplementary Table 49: Results: Incentivization hypothesis 5 – Subjects in any of the incentivization groups will report more positive attitudes towards the app than subjects assigned to the control group. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Positive attitudes (Pooled vs. control)	0.136 (0.086) [-0.033, 0.305]	0.088 (0.062) [-0.033, 0.209]	0.656 (0.452) [-0.231, 1.542]	0.105 ($d=0.093$)
Positive attitudes (EUR 1 vs. control)	0.249 (0.107) [0.039, 0.460]	0.203 (0.079) [0.047, 0.359]	1.403 (0.557) [0.308, 2.497]	0.158 ($d=0.140$)
Positive attitudes (EUR 2 vs. control)	-0.004 (0.104) [-0.208, 0.201]	-0.046 (0.072) [-0.187, 0.095]	-0.495 (0.820) [-2.107, 1.117]	0.143 ($d=0.130$)
Positive attitudes (EUR 5 vs. control)	0.168 (0.106) [-0.042, 0.377]	0.109 (0.073) [-0.035, 0.254]	0.654 (0.425) [-0.182, 1.490]	0.147 ($d=0.132$)

Supplementary Table 50: Results: Incentivization hypothesis 6 – Subjects in any of the incentivization groups will be more likely to intend to share a message about the app than subjects assigned to the control group. The columns report: (1) unadjusted difference-in-means estimate, (2) covariate-adjusted ITT estimate, (3) Complier Average Causal Effect estimate using an IV framework, (4) Minimum Detectable Effect assuming power=0.80 and $\alpha=0.05$ in the covariate-adjusted estimate. Standard errors are reported in parentheses and 95% CIs are reported in square brackets.

Outcome	Unadjusted ITT	Adjusted ITT	CACE	MDE
Message sharing (Pooled vs. control)	0.065 (0.028) [0.010, 0.120]	0.065 (0.028) [0.011, 0.120]	0.467 (0.235) [0.005, 0.928]	0.083 ($d=0.124$)
Message sharing (EUR 1 vs. control)	0.055 (0.039) [-0.021, 0.131]	0.055 (0.039) [-0.021, 0.131]	0.430 (0.322) [-0.202, 1.062]	0.077 ($d=0.175$)
Message sharing (EUR 2 vs. control)	0.081 (0.060) [-0.038, 0.200]	0.081 (0.060) [-0.037, 0.200]	0.984 (0.804) [-0.596, 2.564]	0.122 ($d=0.175$)
Message sharing (EUR 5 vs. control)	0.059 (0.042) [-0.023, 0.141]	0.059 (0.042) [-0.023, 0.141]	0.194 (0.199) [-0.197, 0.585]	0.080 ($d=0.178$)

3.3.2 Effect heterogeneity

Supplementary Table 51: Tests of effect heterogeneity by pre-treatment age, NPI compliance, and government trust. Effects are computed in the covariate-adjusted ITT model.

Outcome	Any incentive treatment interacted with								
	Age			NPI compliance			Gov. trust		
	B	95% CI	P	B	95% CI	P			
Uptake (tracked)	-0.01	[-0.25, 0.24]	0.964	0.55	[-0.12, 1.22]	0.107	-0.20	[-1.40, 1.00]	0.741
Uptake (hybrid)	-0.15	[-0.27, -0.03]	0.016	0.50	[0.03, 0.97]	0.036	0.48	[-0.12, 1.08]	0.117
Uptake (reported)	-0.16	[-0.28, -0.04]	0.008	0.32	[-0.11, 0.76]	0.148	0.45	[-0.09, 0.99]	0.102
Knowledge	-0.03	[-0.11, 0.05]	0.518	0.29	[0.02, 0.56]	0.038	0.25	[-0.08, 0.57]	0.139
Knowledge (icon, screenshot)	-0.04	[-0.10, 0.01]	0.118	-0.12	[-0.30, 0.06]	0.185	0.11	[-0.12, 0.33]	0.339
Positive attitudes	0.00	[-0.07, 0.07]	0.930	0.18	[-0.12, 0.47]	0.239	-0.20	[-0.53, 0.13]	0.234
Message sharing (tracked)	0.05	[0.00, 0.10]	0.031	0.04	[-0.06, 0.14]	0.411	0.04	[-0.14, 0.22]	0.657

Supplementary Table 52: Tests of effect heterogeneity by pre-treatment COVID threat perceptions for family and friends, data privacy attitudes, and message treatment status. Effects are computed in the covariate-adjusted ITT model.

Outcome	Any incentive treatment interacted with								
	Threat perception			Data control			Message treatment		
	B	95% CI	P	B	95% CI	P			
Uptake (tracked)	0.44	[-0.27, 1.15]	0.224	0.40	[-0.42, 1.23]	0.333	0.50	[-0.13, 1.14]	0.121
Uptake (hybrid)	0.41	[-0.02, 0.84]	0.063	0.03	[-0.43, 0.49]	0.894	0.38	[0.02, 0.74]	0.038
Uptake (reported)	0.35	[-0.05, 0.75]	0.089	-0.15	[-0.56, 0.27]	0.482	0.52	[0.08, 0.95]	0.020
Knowledge	-0.00	[-0.25, 0.25]	0.989	-0.00	[-0.26, 0.25]	0.993	0.08	[-0.18, 0.35]	0.535
Knowledge (icon, screenshot)	0.06	[-0.12, 0.24]	0.504	-0.10	[-0.28, 0.07]	0.252	0.20	[0.00, 0.40]	0.045
Positive attitudes	0.14	[-0.12, 0.39]	0.293	-0.01	[-0.25, 0.23]	0.924	0.26	[-0.01, 0.54]	0.060
Message sharing (tracked)	0.03	[-0.07, 0.14]	0.519	-0.04	[-0.15, 0.07]	0.487	0.01	[-0.09, 0.12]	0.789

References

1. ZEIT ONLINE, “Nur jedes zweite Labor kann Testergebnisse digital übermitteln”, <https://www.zeit.de/wissen/gesundheit/2020-08/corona-warn-app-testergebnisse-digital-labor-kritik-fdp>, 6 August 2020.
2. Robert-Koch-Institut, “Kennzahlen zur Corona-Warn-App”, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Kennzahlen.pdf, 22 September 2020.
3. Robert-Koch-Institut, “Coronavirus SARS-CoV-2 - Infektionsketten digital unterbrechen mit der Corona-Warn-App”, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html_09, 2020.
4. Times of India, “Aarogya app enters 100 million users club”, <https://timesofindia.indiatimes.com/gadgets-news/aarogya-setu-app-enters-100-million-users-club/articleshow/75709726.cms>, 13 May 2020.
5. Les Echos, “StopCovid : l’appli gouvernementale attend toujours son public”, <https://www.lesechos.fr/tech-medias/hightech/stopcovid-lappli-gouvernementale-attend-toujours-son-public-1237962>, 31 August 2020.
6. P. H. O’Neill, T. Ryan-Mosley, B. Johnson, “A flood of coronavirus apps are tracking us. Now it’s time to keep track of them.”, 2020, (<https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/>).
7. European Commission, *How tracing and warning Apps can help during the pandemic*, 2020, (https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/travel-during-coronavirus-pandemic/how-tracing-and-warning-apps-can-help-during-pandemic_en).
8. BR, “Europäische Lösung: Corona-Warn-App wird international”, <https://www.br.de/nachrichten/netzwelt/europaeische-loesung-corona-warn-app-wird-international>, SBz942j, 29 September 2020.
9. NDR, “Campingplatz erlaubt Zutritt nur mit Corona-App”, <https://www.ndr.de/nachrichten/niedersachsen/oldenburg-ostfriesland/Krummhoern-Campingplatz-macht-Corona-App-zur-Pflicht,coronaapp154.html>, 28 August 2020.
10. Best for Planning, “Berichtsband 2019”, GIK Media, 2019.
11. A. Guess, K. Munger, “Digital Literacy and Online Political Behavior”, 2020, (<https://doi.org/10.31219/osf.io/3ncmk>).
12. S. Altmann, L. Milsom, H. Zillessen, R. Blasone, F. Gerdon, R. Bach, F. Kreuter, D. Nosenzo, S. Toussaert, J. Abeler, Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. *JMIR mHealth and uHealth* **8**, e19857 (2020).
13. M. Guillon, P. Kergall, Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. *Public Health* **188**, 21–31 (2020).

14. R. Thorneloe, T. Epton, W. Fynn, M. Daly, N. Stanulewicz, A. Kassianos, G. W. Shorter, S.-J. Moll, M. Campbell, S. Sodergren, *et al.*, Scoping Review of Mobile Phone App Uptake and Engagement to Inform Digital Contact Tracing Tools for COVID-19. Preprint at *PsyArXiv* <https://psyarxiv.com/qe9b6/> (2020).
15. J. Bethlehem, Selection bias in web surveys. *International Statistical Review* **78**, 161–188 (2010).
16. F. Elwert, C. Winship, Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology* **40**, 31–53 (2014).
17. J. Schuessler, P. Selb, Graphical causal models for survey inference. Preprint at *SocArXiv* <https://osf.io/preprints/socarxiv/hbg3m/> (2019).
18. R. M. Groves, S. Presser, S. Dipko, The role of topic interest in survey participation decisions. *Public Opinion Quarterly* **68**, 2–31 (2004).
19. F. Keusch, The role of topic interest and topic salience in online panel web surveys. *International Journal of Market Research* **55** (2013).
20. M. Marbach, D. Hangartner, Profiling Compliers and Noncompliers for Instrumental-Variable Analysis. *Political Analysis*, 1–10 (2020).
21. A. Blom, U. Krieger, K. Möhring, E. Naumann, M. Reifenscheid, R. Lehrer, S. Juhl, F. Heinemann, H. P. Grüner, T. Rettig, A. Wenz, M. Fikel, C. Cornesse, S. Friedel, J. Axenfeld, “Mannheim Corona Study”, <https://www.uni-mannheim.de/en/gip/corona-study/>, 2020.
22. M. E. Gerbasi, D. A. Prentice, The Self-and Other-Interest Inventory. *Journal of Personality and Social Psychology* **105**, 495 (2013).
23. W. Lin, Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique. *Annals of Applied Statistics* **7**, 295–318 (2013).
24. A. S. Gerber, D. P. Green, *Field Experiments: Design, Analysis, and Interpretation* (W. W. Norton, 2012).