Check for updates

Tracking and promoting the usage of a COVID-19 contact tracing app

Simon Munzert [1]✉, Peter Selb [2], Anita Gohdes [1], Lukas F. Stoetzer[3] and Will Lowe [1]

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 online 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 a 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 modelling 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 (17 December 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 suggested that a 60% adoption rate would be sufficient[3], although the simulations excluded 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 behaviour[8,9]. For instance, app users might practise more social distancing and thus have fewer social contacts than non-users[10], which would link app usage unfavourably to exposure and transmission risk.

In this paper, we provide evidence on: (1) the selective uptake and usage of Germany's official Corona-Warn-App (henceforth, CWA, or simply the app); and (2) 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 behaviour can generate reporting biases[15,16]. Our study design evades these issues by linking temporally fine-grained behavioural data on app usage with a survey capturing individual measures of a rich set of subject characteristics.

Results

Our study was designed to both track and stimulate usage of the official German DCT app over a period of ~100 d (see Supplementary Discussion for background information). Figure 1 illustrates the study setup (see Methods 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 behaviour of a volunteer panel of $n = 1,132$ individuals, 649 of whom participated in the survey. Our study population therefore consisted of three subgroups: a tracking-only group for which we had sparse but high-quality behavioural data ($n = 482$) and which we used as a baseline control for the analysis of tracked app usage; a survey-only group for which we had rich self-reports but no behavioural data ($n = 1,395$); and a survey-tracking group for which we had both self-reports and behavioural data ($n = 649$).

After completing an initial survey (30 July to 11 August) about social characteristics, and COVID-19-related attitudes and behaviours, survey respondents were randomly assigned to one of two message interventions to stimulate app uptake, or a control group. Attitudinal and self-reported behavioural outcomes were surveyed again in a second panel wave (14–24 August) 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 (meaning they were offered a monetary incentive of €5, €2 or €1 for agreeing to install the app) or to a control group (no incentivization). A final panel wave with $n = 1,569$ respondents (77%) was fielded on 10–22 September to measure additional outcomes. Mobile tracking data that were used in this study were collected from the date of the CWA's rollout (16 June) to 22 September and were unaffected by the dropout of respondents in later waves.

This study design offers gains in measurement accuracy and granularity at the expense of generalizability to our target population.

[1]Hertie School, Berlin, Germany. [2]Department of Politics and Public Administration, University of Konstanz, Konstanz, Germany. [3]SCRIPTS, Humboldt University of Berlin, Berlin, Germany. ✉e-mail: munzert@hertie-school.org
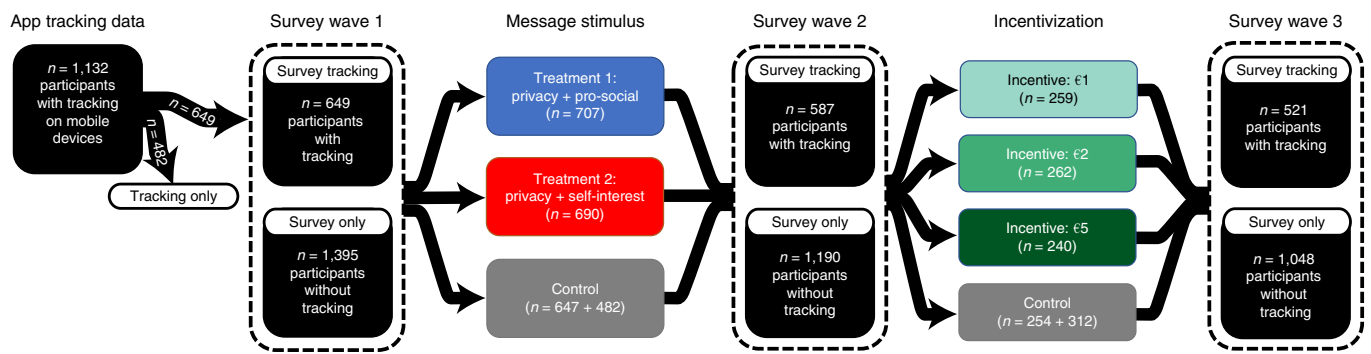
Fig. 1 | Overview of the study design. To collect tracking data (app usage, time stamps, duration and device information) from the app, members of the survey provider's passive tracking panel were incentivized to provide mobile app usage histories via passive metering software (Wakoopa). This was done from 15 June 2020 (1 d before the official app launch) until 21 September 2020 and included only panellists with Android devices. During survey wave 1, participants completed a 20-min survey about sociodemographic, attitudinal and behavioural characteristics. They were then assigned to one of two message interventions (message stimulus) or a control group. For analyses of tracked app usage, the 482 participants with mobile tracking only were used as additional controls. During survey wave 2, an average of 12 d after the initial survey, the participants were surveyed again to reassess their attitudes and behaviours. As part of the follow-up survey, self-reported non-users of the app were randomly assigned to one of three incentivization conditions or to a control group. When analysing tracked app usage, the 312 participants in the tracking-only group who did not have the app installed at the time of the follow-up survey were included as additional controls. Finally, during survey wave 3, an average of 28 d later, the survey wave 2 participants were re-invited to another follow-up survey during which their attitudes and behaviours were re-assessed.

However, compositional information about this population (that is, smartphone users in Germany whose devices satisfy technical minimum standards (at least Android 6 and iOS 13.5)) was not available. Survey-only respondents were potentially more representative of the target population. The Supplementary Information offers more detailed comparisons. It is important to note that our primary interest is in app usage conditional on covariates, not 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 we did model selection into the tracking sample and panel attrition to ensure that these were independent of treatment-related variables (see Supplementary Tables 2–4).

Differential uptake. Figure 2 displays app uptake, defined as opening the CWA at least once, across groups of survey-tracking respondents. In marked contrast with earlier survey work[14,18,19], uptake was more prevalent among older (50+ years) than younger (18–49 years) 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 COVID-19—usually older respondents—were 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, 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 there was no statistically significant evidence for a correlation with gender ($t = 1.38$; d.f. = 620.23; $P = 0.169$; 95% CI = −0.02, 0.13) or parental status ($t = 0.18$; d.f. = 607.97; $P = 0.859$; 95% CI = −0.07, 0.08).

Respondents who reported following the guidelines to wear a mask, wash their hands frequently and practise 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 unfavourably linked to compliance behaviour with other non-pharmaceutical interventions (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 ($t = 1.87$; d.f. = 158.03; $P = 0.063$; 95% CI = 0, 0.20), those who live in areas with registered COVID-19 outbreaks ($t = 1.09$;

d.f. = 44.18; $P = 0.280$; 95% CI = −0.08, 0.26) and those who 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, 0.15) or visiting restaurants and bars ($t = 0.87$; d.f. = 118.30; $P = 0.386$; 95% CI = −0.06, 0.16) once or less in the week before the survey were also 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 healthcare 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) compared with 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 6 weeks of deployment, adoption levels plateaued. We investigated whether and how app usage can be promoted by embedding two randomized experiments in the panel survey, using treatments designed to increase information (wave 1) and incentives (wave 2).

Wave 1. Wave 1 participants were assigned to one of two information treatment conditions or a control condition. Treated respondents were shown a video of ~2 min on the app that emphasized: (1) app functionality; (2) data privacy; and (3) the benefits of app usage for either vulnerable populations (pro-social message condition) or the respondent themself (self-interest message condition).

Our selection of functionality and data privacy messages in wave 1 was designed to mimic existing government campaigns, address
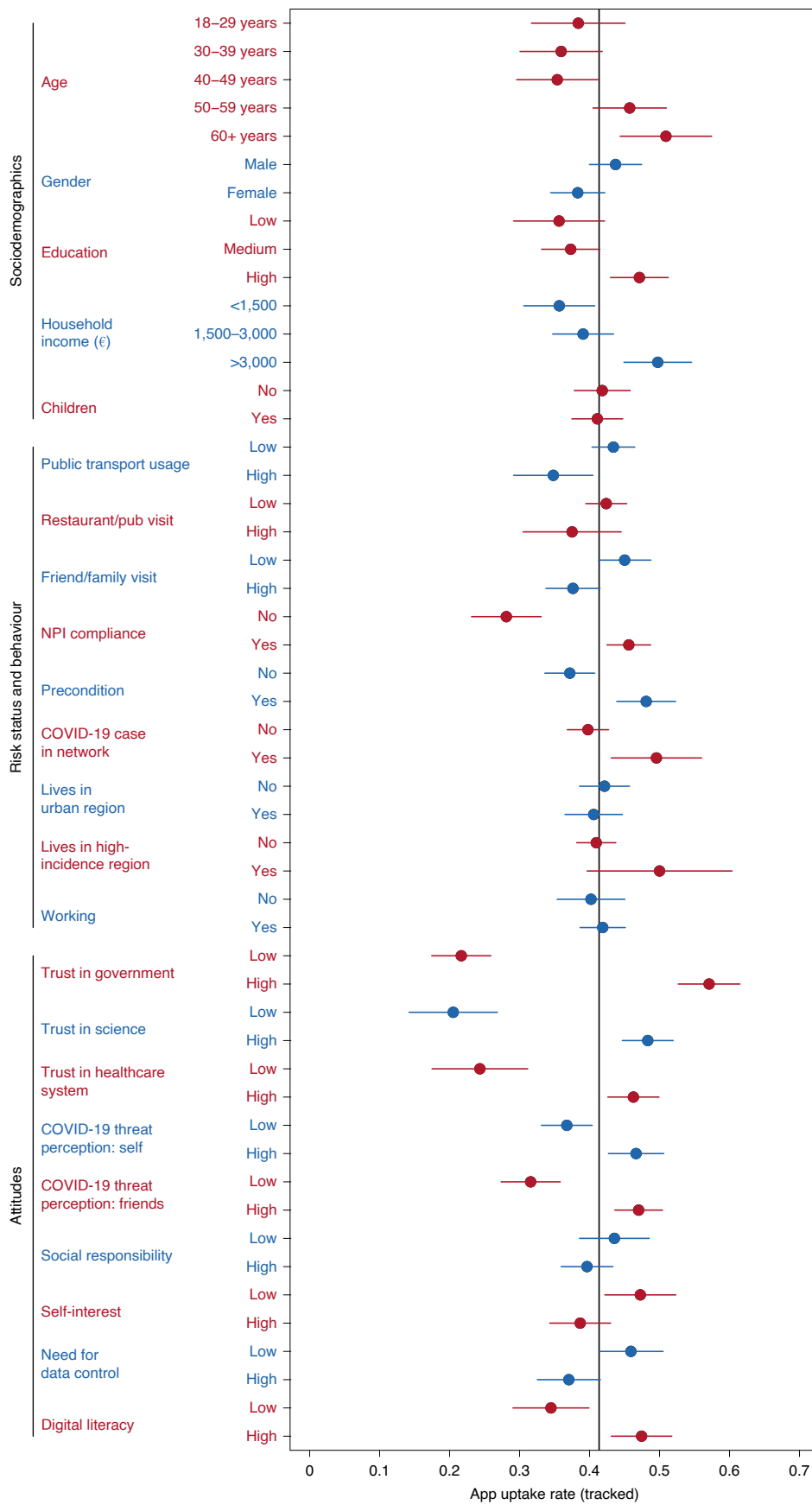
Fig. 2 | Pre-survey tracked app uptake by subgroups. Uptake rates, as observed among survey-tracking panel members ($n = 649$) on 28 July 2020. The uptake rate is reported for each sociodemographic stratum, risk status/behaviour and attitude type. Group means show 83% CIs so non-overlapping intervals indicate a significant difference at $P < 0.05$. The overall sample uptake rate (0.41) is indicated by a black line.

well-known public concerns and build on established social science literature on motivations for individuals to comply with public health and safety initiatives. Previous evidence from public opinion surveys indicated that individuals who do not use the app worry about the app's effectiveness and its data privacy[20]—a finding we replicated (see Supplementary Fig. 19). To address these concerns, the videos shown to participants followed the government's strategy of highlighting favourable arguments (see, for example, https://bit.ly/33MXnsX) and addressed some of the most frequent public concerns; namely, privacy, effectiveness and general knowledge about the app's functionality.

For the treatments, we built on existing research on pro-social[21] or self-interested[22] motivations as core factors underlying individual behaviour. 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 suggested that both pro-social and self-focused messaging frames can stimulate individuals to seek additional information about COVID-19 (ref. [27]). The pro-social and self-interest appeals were added as scrolling text on a filmed tablet screen at the end of the video (see https://youtu.be/dyhDd_vrGEE for the pro-social message and https://youtu.be/suOCvlW8_R0 for the self-interest message). For more details, see the Methods and Supplementary Information.

Wave 2. 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 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 and 9).

We build on a large body of research in behavioural 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 relied on reported and hybrid measures of uptake (see Supplementary Fig. 13). We further tested 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 tested whether the treatments sparked further interest and motivated 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 behaviour relevant for DCT to work. We asked whether they kept 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 they tested positive.

Given the sample size, pre-treatment covariates and repeated measures, the minimum detectable effect for most outcomes on the basis of a test with 80% power for our pre-registered level of significance of 0.05 was 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 estimated average treatment effects using difference-in-means and saturated regressions[31], to identify the intent-to-treat (ITT) effects (Methods and Supplementary Information; see Supplementary Results for estimates of the complier average causal effect).

Information treatments. Both video treatments successfully disseminated factual knowledge about the app. In a pre-registered manipulation check in the 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 $B_{\text{message, pooled}}^{\text{ITT}} = 0.25$ s.d.; 95% CI = 0.16, 0.35; $P < 0.001$; Fig. 3 and Extended Data Fig. 2) and also induced positive attitudes towards it ($B_{\text{message, pooled}}^{\text{ITT}} = 0.11$ s.d.; 95% CI = 0.05, 0.17; $P = 0.001$).

We did not find significant effects on uptake, whether measured using tracking data ($B_{\text{message, pooled}}^{\text{ITT}} = 0.01$ s.d.; 95% CI = −0.02, 0.03; $P = 0.533$) or reported data ($B_{\text{message, pooled}}^{\text{ITT}} = 0.02$ s.d.; 95% CI = −0.02, 0.06; $P = 0.282$). Only when using the hybrid measure did we estimate a small significant positive effect ($B_{\text{message, pooled}}^{\text{ITT}} = 0.03$ s.d.; 95% CI = 0, 0.06; $P = 0.039$).

We found no evidence for a statistically significant effect of information treatment on sharing messages about the app ($B_{\text{message, pooled}}^{\text{ITT}} = -0.05$ s.d.; 95% CI = −0.14, 0.03; $P = 0.222$), looking up additional information about the app ($B_{\text{message, pooled}}^{\text{ITT}} = 0.07$ s.d.; 95% CI = −0.04, 0.18; $P = 0.208$) or activating Bluetooth ($B_{\text{message, pooled}}^{\text{ITT}} = 0.07$ s.d.; 95% CI = −0.03, 0.17; $P = 0.169$).

In addition, we found no clear evidence for statistically significant differences between pro-social and self-interest appeals (with the exception of positive attitudes ($B_{\text{pro-social versus self-interest}}^{\text{ITT}} = -0.07$ s.d.; 95% CI = −0.14, 0; $P = 0.045$) and activating Bluetooth ($B_{\text{pro-social versus self-interest}}^{\text{ITT}} = 0.12$ s.d.; 95% CI = 0.02, 0.22; $P = 0.022$)). Also, we found no evidence that participants who received the pro-social treatment were significantly more likely to install the app (tracked) compared with respondents who were assigned to the self-interest treatment ($B_{\text{pro-social versus self-interest}}^{\text{ITT}} = 0$ s.d.; 95% CI = −0.04, 0.05; $P = 0.848$).

Incentive treatments. Using the incentive experiment, we showed 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 ($B_{\text{incentive, pooled}}^{\text{ITT}} = 1.05$ s.d.; 95% CI = 0.71, 1.39; $P < 0.001$; Fig. 3 and Extended Data Fig. 3), hybrid ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.84$ s.d.; 95% CI = 0.66, 1.01; $P < 0.001$) and reported uptake ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.84$ s.d.; 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% participant agreement to install), there was still a 17-percentage-point increase in tracked app uptake across treatments ($B_{\text{incentive, pooled}}^{\text{ITT}} = 17$ percentage points; 95% CI = 12, 22; $P < 0.001$).

We did not find statistically significant evidence for higher uptake rates in the higher-incentive groups compared with the lower-incentive groups ($B_{\text{€5 versus €1}}^{\text{ITT}} = 0.54$ s.d.; 95% CI = −0.28, 1.37; $P = 0.197$; Extended Data Fig. 3). Note, however, that although the differences in effect sizes between the €1 and €5 incentive tiers were estimated to be large ($B_{\text{€1}}^{\text{ITT}} = 0.83$ s.d. (95% CI = 0.27, 1.39; $P < 0.001$; Fig. 3) versus $B_{\text{€5}}^{\text{ITT}} = 1.43$ s.d. (95% CI = 0.73, 2.13; $P < 0.001$)), we lack the power to distinguish these differences statistically. In contrast with our pre-registered hypothesis, we did not find statistically significant differences in reported willingness to install the app after incentivization between the different monetary incentive tiers (€1 versus €2: $t = 0.56$; d.f. = 518.83; $P = 0.579$; €1 versus €5: $t = -1.34$; d.f. = 493.62; $P = 0.180$; €2 versus €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 ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.16$ s.d.; 95% CI = 0.03, 0.28; $P = 0.014$) and made it slightly more likely that respondents intended to share a message about the app ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.07$ s.d.; 95% CI = 0.01, 0.12; $P = 0.019$), but the intervention did not significantly increase
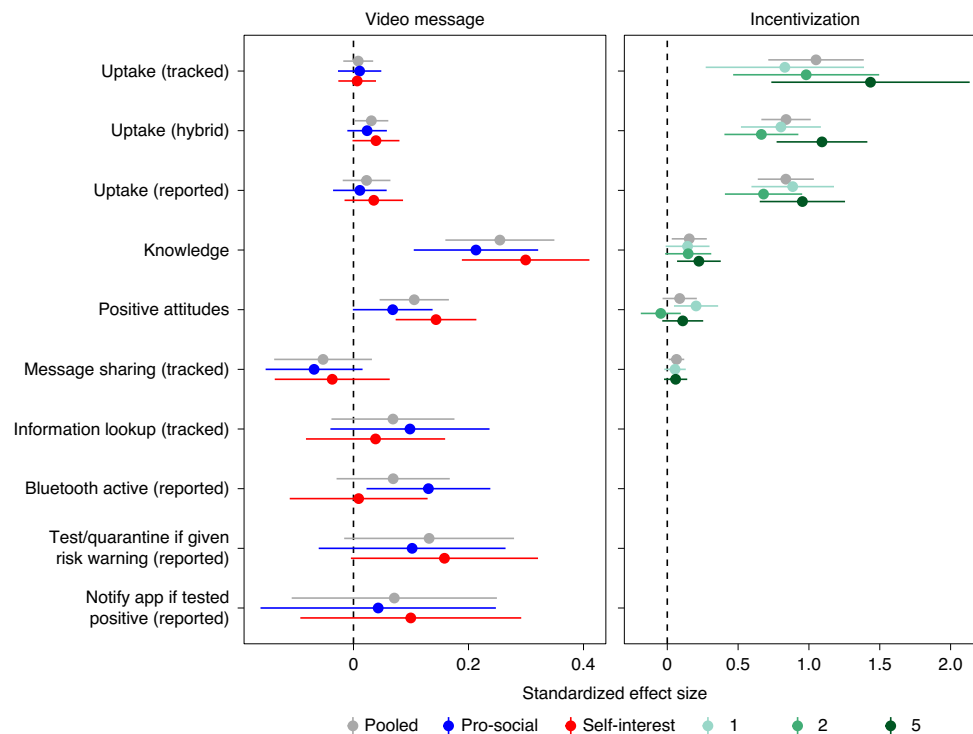
Fig. 3 | Effect of message and incentive treatments on uptake, knowledge, attitudes and behaviour. Each plot shows standardized ITT estimates with 95% CIs from fully saturated ordinary least squares regression models fit using the pre-registered LASSO covariate selection procedure. The video message sample comprises $n = 2,044$, 1,356 and 1,337 respondents for estimation of the pooled, pro-social and self-interest treatment effects, respectively. The incentive sample comprises $n = 1,015$, 513, 516 and 494 respondents for estimation of the pooled, €1, €2 and €5 treatment effects, respectively.

positive attitudes towards the app ($B_{\text{incentive, pooled}}^{\text{ITT}} = 0.09$ s.d.; 95% CI $= -0.03$, 0.21; $P = 0.156$).

In contrast with our pre-registered expectations, we found 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 found some evidence for a negative effect of age on hybrid ($B_{\text{incentive, pooled}}^{\text{ITT}} = -0.15$ s.d.; 95% CI $= -0.27$, -0.03; $P = 0.016$) and reported uptake ($B_{\text{incentive, pooled}}^{\text{ITT}} = -0.16$ s.d.; 95% CI $= -0.28$, -0.04; $P = 0.008$), indicating that young respondents are more responsive to incentives (see Supplementary Tables 42, 43, 51 and 52).

App usage over time. Figure 4 displays app uptake rates in different treatment groups over the study period. The trajectories of observed app adoption in the tracking-only group (black curve) as well as the survey-tracking groups (red and orange curves) follow the official population download figures (see Supplementary Discussion), although they level out at a higher level than for the target population. At the beginning of wave 1, 35% (95% CI $= 30$, 39) of subjects in the tracking-only group and 41% (95% CI $= 37$, 45) of subjects in the survey-tracking group had installed the app, representing a significant difference ($t = 2.18$; d.f. $= 1,053$; $P = 0.030$; 95% CI $= 0.01$, 0.12). The top panel of Fig. 3 shows no evidence for the information treatments having a statistically significant effect on uptake. 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 the 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; bottom

panel of Fig. 3) and 25% of self-reported non-users were offered no incentive (dashed blue curve). Financial incentives increased uptake from 8% (95% CI $= 4$, 12) before 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 (versus 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 the uptake and usage of one of the most popular COVID-19 contact tracing apps, Germany's CWA, and explore experimentally how to make these apps work more effectively by increasing their usage in the population.

This study links survey information on sociodemographic, behavioural 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 over existing approaches, which probably suffer from reporting bias (see Supplementary Table 19 for illustration with our sample) and attrition.

Our results suggest scepticism about the current effectiveness of contact tracing apps and optimism about the potential to increase it by expanding the user base through incentives. The observed uptake patterns indicate a suboptimal distribution across population strata. High adoption rates among those with a high risk of virus exposure and transmission are crucial to maximize the effectiveness of contact tracing. We found significantly lower uptake rates among
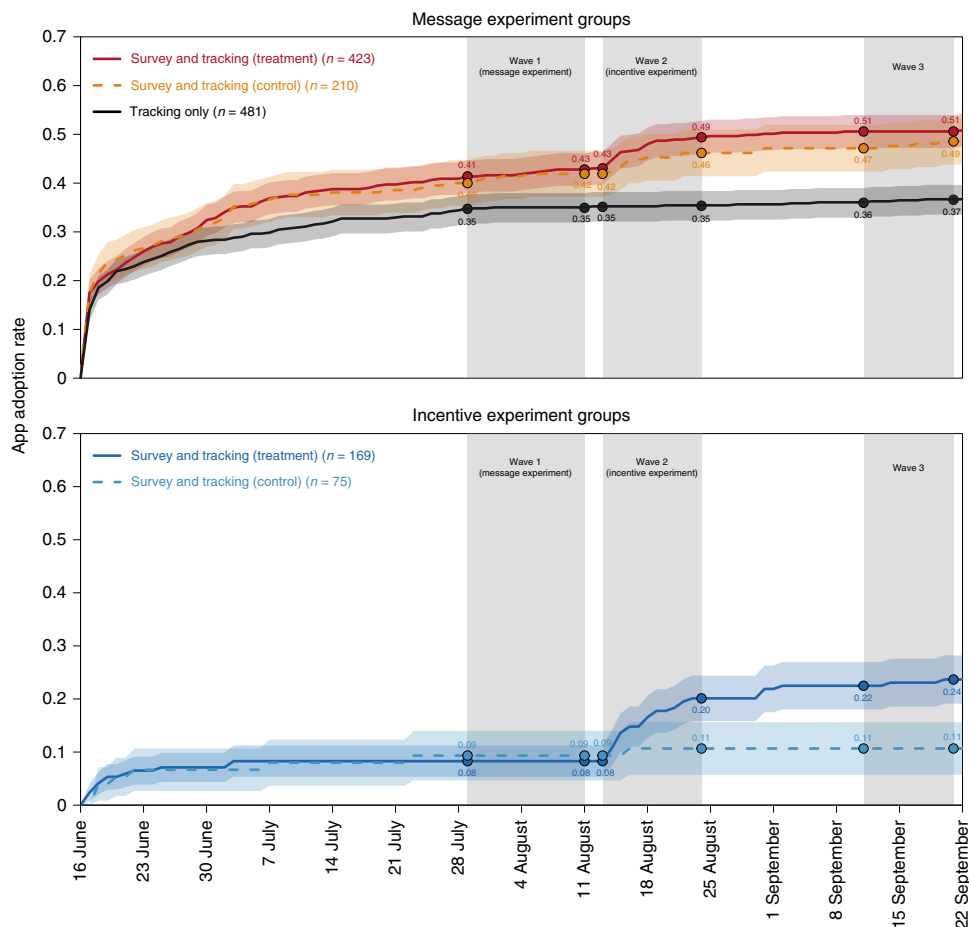
Fig. 4 | App adoption rates over time for message and incentive groups. The top plot shows the estimated rates and 83% CIs (shaded areas) for treated, control and baseline tracking-only respondents. The bottom plot shows the estimated rates and 83% CIs for treated and control respondents in the incentive group. Note that this group was limited to those who never adopted the app or uninstalled it by wave 2. Non-overlapping intervals indicate a significant difference at $P < 0.05$.

those who exhibited lower levels of compliance with other NPI measures and, somewhat surprisingly, among younger age groups. In the Supplementary Information, we provide additional evidence that the positive relationship between age and app use is probably 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 (14 million downloads at the beginning of the study period and 18.4 million 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 favourable 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 a 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 have not addressed 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 offer guidance for expanding app usage in an efficient and effective manner[43]. After substantial investments in advertising and 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; for example, by offering free credit on mobile app stores. An important limitation of the current DCT 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 labour 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 before data collection on 29 July 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 General Data Protection Regulation requirements and all relevant ethical regulations, as documented in the German Research Foundation 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 from the provider and with informed consent from the participants.

Combining passive behavioural data and survey responses to investigate the usage behaviour of a contact tracing app calls for particular caution when that app has been specifically designed not to share with third parties any usage data about the user in conjunction with auxiliary data. First, following standard practice in survey research, respondents remained anonymous to the researchers. Information that is potentially compromising (such as a combination of zip code with sociodemographic data) has not been published as replication data; instead, we have only published 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 joined the passive tracking panel were informed about the type of data collected, and were kept anonymous. They were 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 Respondi, panellists 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 were implemented on our end, while Respondi rewarded respondents using their Mingle points system. The value of points awarded was nominal as a small compensation for the time taken to fill out the survey, and did not exert undue influence (participants of the regular (tracking) panel received the equivalent of €1 (€2) for the initial 20-min survey and €0.50 (€1) for each of the 10-min follow-up surveys. Participation was voluntary. Participants had to be at least 18 years old and had to reside in Germany. Panellists were selected according to their gender (two groups), age (five groups) and education (three 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 panellists 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 are sent via a secure connection to a cloud-based environment. Respondents provided informed consent and were 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 CWA (time stamp, connection and duration) and device metadata device type (operating system type, manufacturer and model). The tracking data we used covered the time period between 16 June (the launch date of the CWA) and 21 September. For the 1,132 users tracked in the survey-tracking or tracking-only samples, we identified 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 s per interaction with the app (that is, 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 registered ($n = 129$ in total; already subtracted from the numbers 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. Treatment group 1 (pro-social message) received an explainer video 2:05 min in length on the contact tracing app with an emphasis on: (1) privacy-related aspects; and (2) the benefits of app usage for vulnerable populations. Treatment group 2 (self-interest message) received an explainer video 2:05 min in length on the contact tracing app with an emphasis on: (1) privacy-related aspects (identical to treatment 1); and (2) the benefits of app usage for the respondent themself. The control group received no video. Due to a lack of time, we did not pilot the videos and only ran internal technical quality checks to adjust the order of the content, the design of the scrolling text and overall the 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 *PC World* on 16 June 2020 (see https://youtu.be/I3C9BrC9I-8).

Incentivization treatment. Following the pre-analysis plan, we implemented the following treatment conditions for those who reported not having installed the app. The control group received no incentivization and no encouragement to install the app. 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 provider, which is equivalent to €1, €2 and €5, or approximately US$1.30, US$2.60 and US$5.80, respectively). Participants 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, for use together with their phone to access the installation page.

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

Bluetooth usage. Bluetooth usage could not be observed directly in the tracking data. Therefore, we used a survey item in all three waves, asking the respondents who reported to have the app installed how often they used the CWA by activating Bluetooth communication on their smartphone. The five answer categories, 'Always', 'Mostly', 'Sometimes', 'Rarely' and 'Never' were recoded as 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 was calculated as the total number of correct answers divided by the total number of items. The knowledge battery was run in all waves. The data from wave 1 were exclusively used for the manipulation check and the data from wave 2 were used to estimate the effect of the message treatments on app knowledge reported in the main text.

App attitudes. We measured attitudes towards the app by conducting a principal component analysis (PCA) of 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 ran

PCA on the set of responses in wave 1 to derive an attitude index by taking the first principal component, and then predicted the attitude index in waves 2 and 3 by using the same set of weights computed in wave 1.

App message sharing behaviour 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 CWA in Apple's App Store and Google Play (see Supplementary Fig. 14). Furthermore, additional links offered ways to share information about the app via social media (Facebook, Twitter and WhatsApp) and email. These links were internally tagged, which allowed us to track clicks on any of them. We used this to measure sharing behaviour, 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 and 0 otherwise.

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 as numerical scores.

Covariates. Age was measured in five categories (18–29 years, 30–39 years, 40–49 years, 50–59 years and 60+ years). For some descriptives, we used age as a continuous variable to facilitate interpretation. Gender was measured using a dummy variable (male = 0; female = 1). Education was 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 was measured in five categories (no children, one child, two children, three children or four or more children) and, for some descriptives, coded binary as 'children yes/no' to facilitate interpretation. Household income was measured in six categories (up to €500, 'between €500 and 1,500', 'between €1,500 and 3,000', 'between €3,000 and 5,000', 'between €5,000 and 10,000' and 'more than €10,000). In the main models, these were collapsed into three categories (up to €1,499, €1,500 to 2,999 and €3,000 and more). Whether or not a participant lived in a region with a high number of cases was measured by matching respondents' zip code to the district data published by the Robert Koch Institute as of 25 August 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 100,000 inhabitants) were coded as high-incidence regions. Also using zip code data, we classified the respondent as living in an urban region when the zip code belonged to a town or city. A respondent's employment status was classed as 'working' if they reported that they were working full or part time, attending a school or university or in a traineeship or apprenticeship scheme. Trust in government, scientists and the healthcare system was measured using an item that asked respondents how much trust they had in the respective institutions. The five answer categories, ranging from 'Not trust at all' to 'Complete trust' were recoded as numerical scores ('Not trust at all' = 0 to 'Complete trust' = 4). Pre-existing health conditions were measured by asking respondents whether they had 'any pre-existing conditions that increase the risk of a severe course of COVID-19 (for example, high blood pressure, obesity, diabetes or COPD)' and encoding their answers as yes/no. Incidences of COVID-19 infections in the personal environment were measured by asking respondents whether they knew of anyone (including themself, family, friends, acquaintances, colleagues or neighbours) who had been infected by COVID-19. The answers 'Yes, 1–3 people' and 'Yes, more than 3 people' were encoded as 1, with 'No' encoded as 0. Threat perception of COVID-19 for themselves and family/friends was measured by asking respondents how concerned they were about the consequences of COVID-19 for their themselves and family and friends separately. The four-point scale ('Not at all concerned', 'Not too concerned', 'Somewhat concerned' and 'Very concerned') was recoded as numerical scores ('Not at all concerned' = 0 to 'Very concerned' = 4). Data privacy concerns were measured by conducting a PCA of responses to a set of three items. We ran 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 was measured by conducting a PCA of responses to a set of three items. We ran 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 was measured by conducting a PCA of responses to a set of three items. We ran 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 was measured by conducting a PCA of responses to a set of four items. We ran 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 was measured by asking respondents to what extent they complied with behavioural recommendations to contain the pandemic. The four-point scale ('I almost always follow them', 'I try to adhere to them, but often I do not succeed (for example, for professional reasons)', 'I barely adhere to them' and 'I am not aware of these recommendations') was recoded as a binary indicator of compliance ('I almost always follow them' = 1; otherwise = 0). Risk behaviour was measured using a battery of items, asking respondents how often in the past seven days they had: (1) used public transport; (2) visited a restaurant, café or bar; and (3) met with friends, relatives or acquaintances in person. For each item, the four-point scale ('Never', 'Once this week', 'Several times this week' and 'Every day') was broken down into low (never or once this week) and high risk behaviour (several times this week or every day).

Analysis. *Modelling approach.* To evaluate both the message and the incentivization experiment, we estimated average treatment effects (using difference-in-means and saturated regressions[31] to identify the ITT) and the complier average causal effect (using the instrumental-variables framework[51]). For all covariate-adjusted regressions, we used a LASSO procedure to select covariates. The estimates reported in the main text are the covariate-adjusted ITT estimates. All other estimates, along with the results from minimum detectable effect-size calculation, are reported in the Supplementary Results. Two-tailed statistical tests were used in all analyses and we did not correct the *P* values for multiple hypotheses. Due to the nature of our experimental manipulations, respondents were aware of which treatment they were receiving, but 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 pooled the survey-only and survey-plus-tracking samples. For app usage as the outcome, we ran three types of analyses. Under the first type, we restricted our analyses to the survey-plus-tracking sample and reported both unadjusted and covariate-adjusted estimates. Under the second type, we pooled the survey-plus-tracking sample and the tracking-only sample, which provided additional control units, and reported both unadjusted and covariate-adjusted estimates. In the latter case, the set of possible covariates was reduced to those that were also available for members of the tracking-only panel. Under the third type, we pooled all sub-samples and generated a hybrid outcome measure that used reported app usage in the survey-only sample and tracked app usage in the tracking sample, if available.

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

Data availability

Anonymized versions of the datasets used in this study are shared on GitHub, with permission of the data provider and in compliance with General Data Protection Regulation requirements, at https://github.com/simonmunzert/covid19-app-experiment. Information that would make subjects directly identifiable has been 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 numerical COVID-19 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. Bai, Y. et al. Presumed asymptomatic carrier transmission of COVID-19. *J. Am. Med. Assoc.* **323**, 1406–1407 (2020).
2. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
3. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
4. Whitelaw, S., Mamas, M. A., Topol, E. & Topol, H. G. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit. Health* **2**, E435–E440 (2020).
5. O'Neill, P. H., Ryan-Mosley, T. & Johnson, B. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. *MIT Technology Review* (7 May 2020).
6. Hinch, R. et al. OpenABM-COVID19—an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.16.20195925 (2020).
7. Parker, M. J., Fraser, C., Abeler-Dörner, L. & Bonsall, D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *J. Med. Ethics* **46**, 427–431 (2020).

8. Canetti, R., Trachtenberg, A. & Varia, M. Anonymous collocation discovery: harnessing privacy to tame the coronavirus. Preprint at *arXiv* http://arxiv.org/abs/2003.13670 (2020).
9. Dorélien, A. M. et al. Minnesota social contacts and mixing patterns survey with implications for modelling of infectious disease transmission and control. *Survey Practice* https://doi.org/10.29115/SP-2020-0007 (2020).
10. Von Wyl, V. et al. A research agenda for digital proximity tracing apps. *Swiss Med. Wkly* **150**, w20324 (2020).
11. Ahmed, N. et al. A survey of COVID-19 contact tracing apps. *IEEE Access* **8**, 134577–134601 (2020).
12. Trang, S., Trenz, M., Weiger, W. H., Tarafdar, M. & Cheung, C. M. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. *Eur. J. Inf. Syst.* **29**, 415–428 (2020).
13. Reelfs, J. H., Hohlfeld, O. & Poese, I. 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. Altmann, S. et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. *JMIR mHealth uHealth* **8**, e19857 (2020).
15. Munzert, S. & Selb, P. Can we directly survey adherence to non-pharmaceutical interventions? Evidence from a list experiment conducted in Germany during the early corona pandemic. *Surv. Res. Methods* **14**, 205–209 (2020).
16. Tourangeau, R. & Yan, T. Sensitive questions in surveys. *Psychol. Bull.* **133**, 859–883 (2007).
17. Bareinboim, E. & Pearl, J. Causal inference and the data-fusion problem. *Proc. Natl Acad. Sci. USA* **113**, 7345–7352 (2016).
18. Guillon, M. & Kergall, P. 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. Thorneloe, R. et al. Scoping review of mobile phone app uptake and engagement to inform digital contact tracing tools for COVID-19. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/qe9b6 (2020).
20. Sonnenberg, A.-K. Neue Corona-Warn-App: Zweifel am Nutzen ist häufigster Grund für eine Nicht-Nutzung [article in German] *YouGov* https://yougov.de/news/2020/06/22/neue-corona-warn-app-zweifel-am-nutzen-ist-haufigs/ (2020).
21. Engel, C. Dictator games: a meta study. *Exp. Econ.* **14**, 583–610 (2011).
22. Persky, J. The ethology of *Homo economicus*. *J. Econ. Perspect.* **9**, 221–231 (1995).
23. Bellato, A. 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. *Soc. Sci. Humanit. Open* **2**, 100062 (2020).
24. Bonell, C. et al. 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 (2020).
25. Van der Linden, C. & Savoie, J. Does collective interest or self-interest motivate mask usage as a preventive measure against COVID-19? *Can. J. Polit. Sci.* **53**, 391–397 (2020).
26. Pfattheicher, S., Nockur, L., Böhm, R., Sassenrath, C. & Petersen, M. B. The emotional path to action: empathy promotes physical distancing and wearing of face masks during the COVID-19 pandemic. *Psychol. Sci.* **31**, 1363–1373 (2020).
27. Banker, S. & Park, J. Evaluating prosocial COVID-19 messaging frames: evidence from a field study on Facebook. *Judgm. Decis. Mak.* **15**, 1037–1043 (2020).
28. Kamenica, E. Behavioral economics and psychology of incentives. *Annu. Rev. Econ.* **4**, 427–452 (2012).
29. Giles, E. L., Robalino, S., McColl, E., Sniehotta, F. F. & Adams, J. The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. *PLoS ONE* **9**, e90347 (2014).
30. Sawilowsky, S. S. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* **8**, 597–599 (2009).
31. Lin, W. Agnostic notes on regression adjustments to experimental data: reexamining Freedman's critique. *Ann. Appl. Stat.* **7**, 295–318 (2013).
32. Linden, M. et al. 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. *Aktuelle Kennzahlen zur Corona-Warn-App, u.a. Downloadzahl* (Robert Koch Institut, 2020); https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Kennzahlen.pdf
34. French Bourgeois, L., Harell, A. & Stephenson, L. B. To follow or not to follow: social norms and civic duty during a pandemic. *Can. J. Polit. Sci.* **53**, 273–278 (2020).
35. Murphy, K., Williamson, H., Sargeant, E. & McCarthy, M. Why people comply with COVID-19 social distancing restrictions: self-interest or duty? *Aust. NZ J. Criminol.* **53**, 477–496 (2020).
36. IJzerman, H. et al. Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* **4**, 1092–1094 (2020).
37. Baker, R. et al. Summary report of the AAPOR task force on non-probability sampling. *J. Surv. Stat. Methodol.* **1**, 90–143 (2013).
38. Keiding, N. & Louis, T. A. Web-based enrollment and other types of self-selection in surveys and studies: consequences for generalizability. *Annu. Rev. Stat. Appl.* **5**, 25–47 (2018).
39. Hargittai, E. & Shaw, A. Comparing internet experiences and prosociality in Amazon Mechanical Turk and population-based survey samples. *Socius* https://doi.org/10.1177/2378023119889834 (2020).
40. Klingwort, J. & Schnell, R. Critical limitations of digital epidemiology. *Surv. Res. Methods* **14**, 95–101 (2020).
41. Maccari, L. & Cagno, V. Do we need a contact tracing app? *Comput. Commun.* **166**, 9–18 (2021).
42. Almagor, J. & Picascia, S. Exploring the effectiveness of a COVID-19 contact tracing app using an agent-based model. *Sci. Rep.* **10**, e22235 (2020).
43. Farronato, C. et al. How to get people to actually use contact-tracing apps. *Harvard Business Review* https://hbr.org/2020/07/how-to-get-people-to-actually-use-contact-tracing-apps (2020).
44. Bundesregierung verrät Kosten: So viel kostet den Staat die Corona-Warn-App [article in German] *Frankfurter Allgemeine Zeitung* (13 October 2020).
45. Loi, M. How to fairly incentivise digital contact tracing. *J. Med. Ethics* https://doi.org/10.1136/medethics-2020-106388 (2020).
46. Oliver, N. et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci. Adv.* **6**, eabc0764 (2020).
47. Buckee, C. O. et al. Aggregated mobility data could help fight COVID-19. *Science* **368**, 145–146 (2020).
48. Zhou, Y. et al. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. *Lancet Digit. Health* **2**, e417–e424 (2020).
49. Ienca, M. & Vayena, E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat. Med.* **26**, 463–464 (2020).
50. Best for Planning. *Berichtsband 2019* https://gik.media/produkt/b4p-berichtband/ (GIK Media, 2019).
51. Gerber, A. S. & Green, D. P. *Field Experiments: Design, Analysis, and Interpretation* (W. W. Norton, 2012).

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.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-020-01044-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-01044-x.

Correspondence and requests for materials should be addressed to S.M.

Peer review information *Nature Human Behaviour* thanks Jo Hart, Baobao Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Marike Schiffer.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
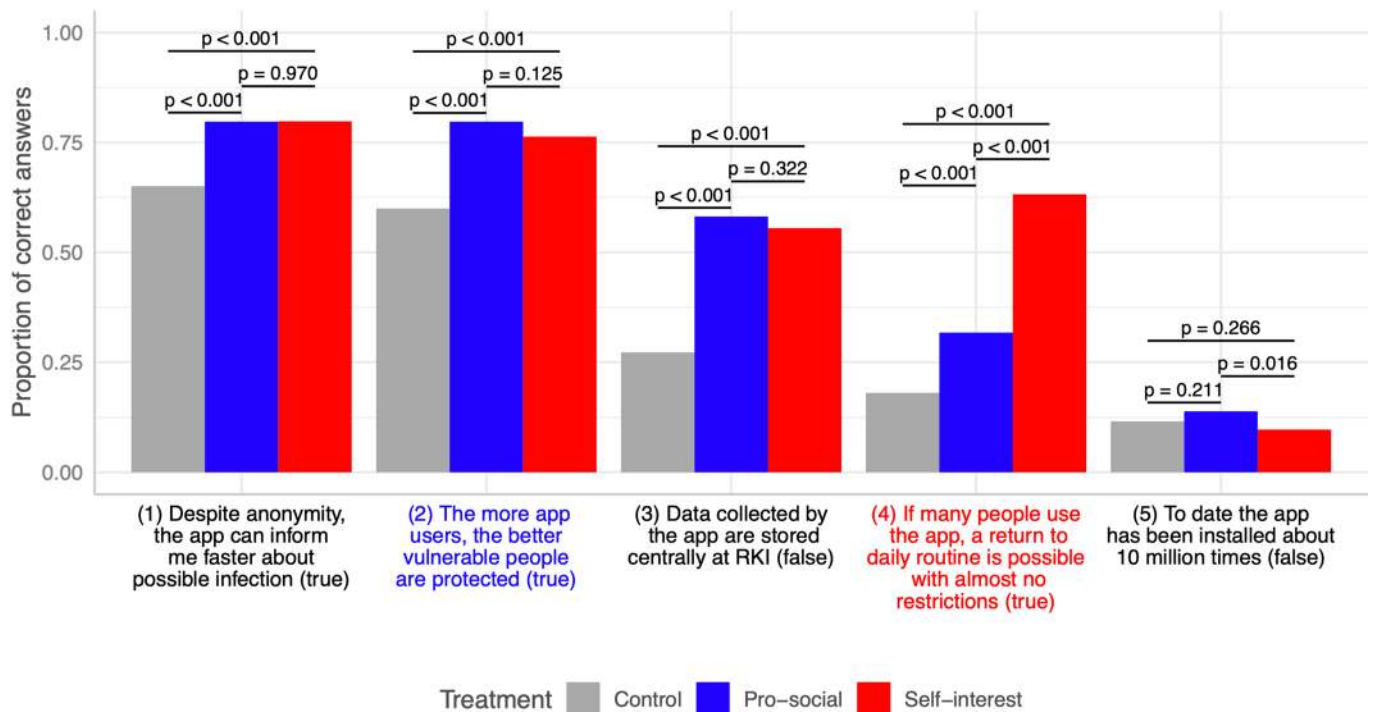
Extended Data Fig. 1 | Results of manipulation check of message experiment. Proportions of knowledge items correctly answered, by treatment group ($N_{\text{Pro-social}} = 707$, $N_{\text{Self-interest}} = 690$, $N_{\text{Control}} = 647$). Two-sample t-test statistics: $T^{Item(1)}_{\text{Pro-social vs. Control}} = 6.09$, d.f. $= 1268.4$, $P < 0.001$, 95% CI 0.10, 0.19, $T^{Item(2)}_{\text{Pro-social vs. Control}} = 8.07$, d.f. $= 1252$, $P < 0.001$, 95% CI 0.15, 0.25, $T^{Item(3)}_{\text{Pro-social vs. Control}} = 12.10$, d.f. $= 1348.70$, $P < 0.001$, 95% CI 0.26, 0.36, $T^{Item(4)}_{\text{Pro-social vs. Control}} = 5.89$, d.f. $= 1337.20$, $P < 0.001$, 95% CI 0.09, 0.18, $T^{Item(5)}_{\text{Pro-social vs. Control}} = 1.25$, d.f. $= 1349.8$, $P = 0.211$, 95% CI -0.01, 0.06, $T^{Item(1)}_{\text{Pro-social vs. Self-interest}} = -0.04$, d.f. $= 1392.3$, $P = 0.970$, 95% CI -0.04, 0.04, $T^{Item(2)}_{\text{Pro-social vs. Self-interest}} = 1.53$, d.f. $= 1384.1$, $P = 0.125$, 95% CI -0.01, 0.08, $T^{Item(3)}_{\text{Pro-social vs. Self-interest}} = 0.99$, d.f. $= 1389.6$, $P = 0.322$, 95% CI -0.03, 0.08, $T^{Item(4)}_{\text{Pro-social vs. Self-interest}} = -12.40$, d.f. $= 1386.8$, $P < 0.001$, 95% CI -0.36, -0.27, $T^{Item(5)}_{\text{Pro-social vs. Self-interest}} = 2.41$, d.f. $= 1386.8$, $P = 0.016$, 95% CI 0.01, 0.08, $T^{Item(1)}_{\text{Self-interest vs. Control}} = 6.10$, d.f. $= 1265.5$, $P < 0.001$, 95% CI 0.10, 0.20, $T^{Item(2)}_{\text{Self-interest vs. Control}} = 6.50$, d.f. $= 1280.9$, $P < 0.001$, 95% CI 0.11, 0.21, $T^{Item(3)}_{\text{Self-interest vs. Control}} = 10.95$, d.f. $= 1329.1$, $P < 0.001$, 95% CI 0.23, 0.33, $T^{Item(4)}_{\text{Self-interest vs. Control}} = 18.94$, d.f. $= 1299.1$, $P < 0.001$, 95% CI 0.40, 0.50, $T^{Item(5)}_{\text{Self-interest vs. Control}} = -1.11$, 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.
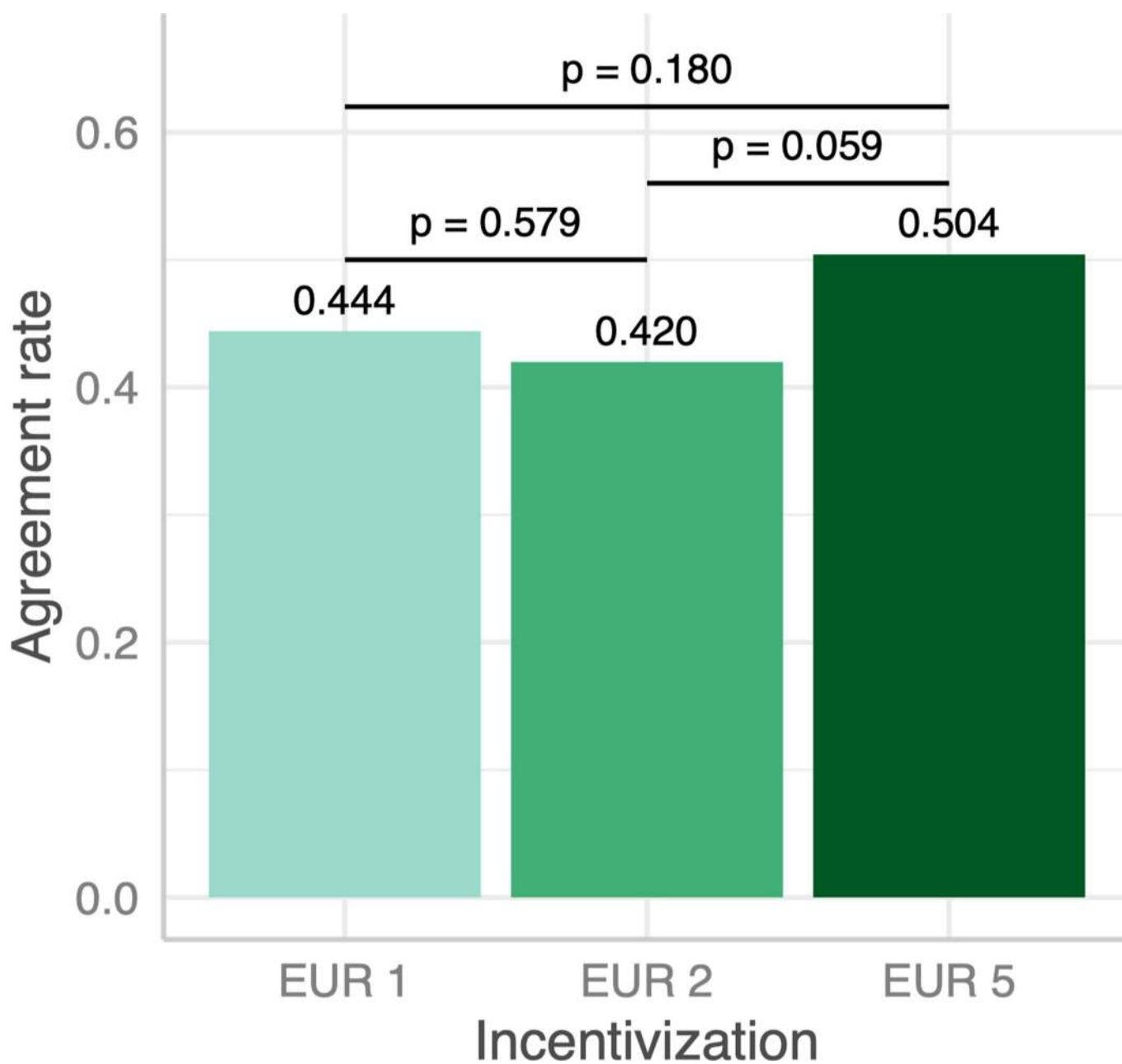
Extended Data Fig. 4 | Agreement rates of app installation incentivization by monetary incentive. ($N_{EUR\,1} = 259$, $N_{EUR\,2} = 262$, $N_{EUR\,5} = 240$). Two-sample t-test statistics: $T_{EUR\,1\,vs.\,EUR\,2} = 0.56$, d.f. $= 518.83$, $P = 0.579$, 95% CI -0.06, 0.11, $T_{EUR\,1\,vs.\,EUR\,5} = -1.34$, d.f. $= 493.62$, $P = 0.180$, 95% CI -0.15, 0.03, $T_{EUR\,2\,vs.\,EUR\,5} = -1.90$, 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. 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.

nature research

Corresponding author(s):   Simon Munzert

Last updated by author(s):   Dec 14, 2020

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Data collection was carried out by external data provider.
Data analysis	The R software was used to analyze the data.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data and computer code that support the findings of this study is open source and available on GitHub at: https://github.com/simonmunzert/covid19-app-experiment.

Field-specific reporting

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

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

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

Behavioural & social sciences study design

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

Study description	Quantitative panel, experimental, and passive smartphone usage data
Research sample	Panelists of a commercial panel in Germany. 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 marginal distributions of the 2019 Best for Planning study. Participants from the tracking panel were recruited from the subset of panelists who had the tracking software installed on a mobile device.
Sampling strategy	In partnership with the online survey firm Respondi, panelists were invited to participate in a study on attitudes in the context of the COVID-19 pandemic with several survey waves. Participation was voluntary.
Data collection	The survey design and programming (including the randomized assignment to intervention groups) was implemented on our end, while Respondi rewarded respondents with their mingle points system.
Timing	June 16th to September 21st, 2020
Data exclusions	None
Non-participation	N/A
Randomization	Participants of the survey were randomly allocated to one of three conditions in the information intervention and, if they reported to have the contact tracing app not installed in Wave 2, randomly assignment to one of four conditions in the incentive intervention.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*

Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

| Ethics oversight | *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.* |

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

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	In partnership with the online survey firm Respondi, panelists were invited to participate in a study on attitudes in the context of the COVID-19 pandemic with several survey waves. Participation was voluntary. 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 anonymous and time-aggregated form with permission of the provider and with informed consent of the participants.
Ethics oversight	Ethics work at the Hertie School is run through the Dean's office. This process is in lieu of an American style committee. The study was approved by the Dean of Research and Faculty at the Hertie School.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*

| Outcomes | *Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.* |

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No | Yes
☐ | ☐ Public health
☐ | ☐ National security
☐ | ☐ Crops and/or livestock
☐ | ☐ Ecosystems
☐ | ☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No | Yes
☐ | ☐ Demonstrate how to render a vaccine ineffective
☐ | ☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☐ | ☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☐ | ☐ Increase transmissibility of a pathogen
☐ | ☐ Alter the host range of a pathogen
☐ | ☐ Enable evasion of diagnostic/detection modalities
☐ | ☐ Enable the weaponization of a biological agent or toxin
☐ | ☐ Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI   ☐ Used   ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a | Involved in the study
☐ | ☐ Functional and/or effective connectivity
☐ | ☐ Graph analysis
☐ | ☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*
Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*