Early Indicators of COVID-19 Infection Prevention Behaviors: Machine

Learning Identifies Personal and Country-Level Factors

Caspar J. van Lissa[1]

Wolfgang Stroebe[2]

Michelle R. vanDellen[3]

N. Pontus Leander[2]

Maximilian Agostini[2]

Benjamin Gützkow[2]

Jannis Kreienkamp[2]

Edward P. Lemay, Jr.[4]

Georgios Abakoumkin[5]

Mioara Cristea[6]

Bertus F. Jeronimus[2]

Manuel Moyano[7]

Jocelyn J. Bélanger[8]

Tim A. Draws[9]

Andrii Grygoryshyn

Clara S. Vetter

On behalf of the PsyCorona Collaboration*

[1] Utrecht University
[2] University of Groningen
[3] University of Georgia
[4] University of Maryland College Park
[5] University of Thessaly
[6] Heriot Watt University

7 University of Cordoba
8 New York University Abu Dhabi
9 Delft University of Technology

Full coauthor information is supplied as supplement and should be added on publication: Jamilah H. B. Abdul Khaiyom, Vjollca Ahmedi, Handan Akkas, Carlos A. Almenara, Anton Kurapov, Mohsin Atta, Sabahat Cigdem Bagci, Sima Basel, Edona Berisha Kida, Anna Bertolini, Nicholas R. Buttrick, Phatthanakit Chobthamkit, Hoon-Seok Choi, Sára Csaba, Kaja Damnjanovic, Ivan Danyliuk, Arobindu Dash, Daniela Di Santo, Karen M. Douglas, , Violeta Enea, Daiane Gracieli Faller, Gavan Fitzsimons, Alexandra Gheorghiu, Ángel Gómez, Qing Han, Mai Helmy, Joevarian Hudijana, Ding-Yu Jiang, Veljko Jovanovic, Željka Kamenov, Anna Kende, Shian-Ling Keng, Tra Thi Thanh Kieu, Yasin Koc, Kamila Kovyazina, Inna Kozytska, Joshua Krause, Arie W. Kruglanski, Maja Kutlaca, Nóra Anna Lantos, Cokorda Bagus Jaya Lesmana, Winnifred R. Louis, Adrian Lueders, Najma Iqbal Malik, Anton Martinez, Kira McCabe, Mirra Noor Milla, Idris Mohammed, Erica Molinario, Hayat Muhammad, Silvana Mula, Hamdi Muluk, Solomiia Myroniuk, Reza Najafi, Claudia F. Nisa, Boglárka Nyúl, Paul A. O'Keefe, Jose Javier Olivas Osuna, Evgeny N. Osin, Joonha Park, Gennaro Pica, Antonio Pierro, Jonas Rees, Anne Margit Reitsema, Elena Resta, Marika Rullo, Michelle K. Ryan, Adil Samekin, Pekka Santtila, Edyta Sasin, Birga Mareen Schumpe, Heyla A Selim, Michael Vicente Stanton, Samiah Sultana, Robbie M. Sutton, Eleftheria Tseliou, Akira Utsugi, Jolien Anne van Breen, Kees Van Veen, Alexandra Vázquez, Robin Wollast, Victoria Wai-lan Yeung, Somayeh Zand, Iris Lav Žeželj, Bang Zheng, Andreas Zick, Claudia Zúñiga.

Abstract

The Coronavirus is highly infectious and potentially deadly. In the absence of a cure or a vaccine, the infection prevention behaviors recommended by the World Health Organization constitute the only measure that is presently available to combat the pandemic. The unprecedented impact of this pandemic calls for swift identification of factors most important for predicting infection prevention behavior. In this paper, we used a machine learning approach to assess the relative importance of potential indicators of personal infection prevention behavior in a global psychological survey we conducted between March-May 2020 ($N = 56,072$ across 28 countries). The survey data were enriched with society-level variables relevant to the pandemic. Results indicated that the two most important indicators of self-reported infection prevention behavior were individual-level injunctive norms—beliefs that people in the community *should* engage in social distancing and self-isolation, followed by endorsement of restrictive containment measures (e.g., mandatory vaccination). Society-level factors (e.g., national healthcare infrastructure, confirmed infections) also emerged as important indicators. Social attitudes and norms were more important than personal factors considered most important by theories of health behavior. The model accounted for 52% of the variance in infection prevention behavior in a separate test sample—above the performance of psychological models of health behavior. These results suggest that individuals are intuitively aware that this pandemic constitutes a social dilemma situation, where their own infection risk is partly dependent on the behaviors of others. If everybody engaged in infection prevention behavior, the virus could be defeated even without a vaccine.

Keywords: Machine learning; COVID-19; Health Behaviors; Social Norms; Social Dilemma

Significance statement

In the absence of a vaccine or cure, virus containment presently depends on individual-level compliance with behaviors recommended by the World Health Organization. However, it is unclear what psychological and societal factors most strongly correlate with such compliance. We used machine learning and database integration to isolate key indicators of compliance from a large international psychological survey. The most important indicators of compliance were not the personal threat of virus infection *per se*, but rather social attitudes—namely, that one's community should engage in such behavior and that society should take extraordinary virus containment measures. People appear to realize that their infection risk is partly dependent on behavior of others and that general compliance is necessary to defeat the pandemic.

Early Indicators of COVID-19 Infection Prevention Behaviors: Machine

Learning Identifies Psychological and Country-Level Factors

In the absence of a vaccine, behavioral measures are the only means by which individuals can lower the risk of getting infected with the COVID-19 virus. In the early phase of the COVID-19 pandemic between March and May 2020, the time when this research was conducted, three infection prevention behaviors were recommended by most governments: frequent hand washing, social distancing, and self-quarantining (MedicalXPress, April 24th, 2020). These behaviors were thought to provide protection because the virus spreads mainly through respiratory droplets that an infected person produces when coughing, sneezing or talking. The droplets can land on surfaces or be inhaled, so hand washing and social distancing were seen as effective behavioral means to reduce infection risk (TechnologyNetworks,com, 2020). People would need to maintain the behaviors even in private.

Behavioral science could help to identify factors that predict individuals' adherence to prevention behaviors. This is important, because – due to the way the virus is transmitted - infection prevention behavior involves both personal health behavior (hand washing) and social behavior (social distancing). Hence, in the midst of a communal effort to flatten the pandemic curve, individual compliance may partly depend on, for instance, social attitudes and norms, as well as broader societal conditions. Using a large international dataset, the present paper therefore set out to identify social indicators of individual compliance with infection prevention behaviors using machine learning. Machine learning can complement theory-driven approaches to identify important determinants or correlates of health behavior, because it maximizes predictive performance, is free from restrictive assumptions of classical linear models, and can identify blind spots in existing knowledge.

Our approach is based on two key assumptions: First, that individual-level psychological factors are likely to indicate compliance with infection prevention behaviors, even when individuals feel unobserved (e.g., Deutsch & Gerard, 1955; Hagger et al., 2014; Oosterhoff et al. 2020). A second key assumption is that infection prevention is, to some extent, a social dilemma. Without a cure or vaccine available at the pandemic onset, COVID-19 has some characteristics of the social dilemma known as a *public bad*, like air pollution: A situation that can only be controlled if most members of a society contribute to the effort (e.g., Sutter, 2002). A public bad is characterized by an incentive for people to profit from the efforts of others without making their own contribution (Olson, 2009). Research on social dilemmas suggests that social attitudes and perceived norms help motivate individual compliance (Biel & Thøgersen, 2007; Liebrand & van Run, 1985). However, this pandemic also diverges from other public bad-dilemmas due to the added personal health risks of the virus. Hence, engaging in infection prevention behavior both reduces the spread of the infection in society (i.e., reduces the public bad), as well as prevents one's own infection risk. Thus, we might expect a range of other individual differences, such as personal risk and health-related vulnerability, to predict compliance.

To identify key indicators of infection prevention behaviors, we launched a large-scale, rapid psychological survey in 28+ countries in the weeks after the World Health Organization (WHO) declared COVID-19 a pandemic. The survey included over one hundred individual-level variables, including basic demographic characteristics (e.g., gender, age, education, religion), brief self-report measures of various psychological factors (e.g., subjective states and well-being, work and financial concerns, societal attitudes, COVID-relevant attitudes and beliefs), and individual infection prevention behaviors (e.g., hand washing, avoiding crowds). The

a priori objective was to recruit tens of thousands of survey responses globally to

assess personal as well as social attitudes and beliefs towards the virus and society's

prescriptions, and examine how these factors relate to individual infection prevention

behaviors. The large global sample was designed to afford society-level database

integration, and employ machine learning analysis to isolate key indicators of

behavioral compliance with WHO advice.

Deductive and inductive approaches

The traditional scientific approach to identify predictors of infection

prevention behaviors is deductive: First, researchers identify relevant theories. Then, a

relatively narrow set of theoretically relevant variables are identified. Finally, using

parametric statistical models, the unique contributions of this set of variables is

examined. Such deductive research typically focuses on whether a theoretical model

can be validated empirically, with less emphasis on the model's predictive

performance. In this approach, a model's performance is typically evaluated by its

ability to adequately *describe the observed data*, defined as the percentage of variance

explained in the outcome variable (*R-square* or R^2). However, as the primary goal of

inductive research is typically to perform inference on the (significance of) theoretical

predictors, the R-square is often of secondary importance.

In recent years, inductive research has been gaining traction as a technique to

complement existing theories by identifying important omissions. In particular,

machine learning offers powerful new tools that can complement deductive models

by identifying potentially relevant predictors or complex relationships that have

eluded theoreticians (Brandmeier, Prindle, McArdle, & Lindenberger, 2016). Machine

learning uses flexible models to find reliable patterns and maximize predictive

performance in one subset of the data (the training sample), and then tests that model

in a separate subset (the test sample; e.g., Hastie, Tibsharani, & Friedman, 2009).

Instead of examining *descriptive performance*, machine learning examines *predictive performance* by computing the R^2 in the test sample, which tells us how well a model is likely to generalize. Machine learning models can distill a large set of candidate variables down to the ones that are most important (in terms of explained variance), and also indicate the direction and shape of the marginal association between those predictors and the outcome.

Relevant theory

Although we set out to use machine learning to inductively identify relevant predictors of infection prevention behavior, our work does not take place in a theoretical vacuum. The COVID-19 pandemic is a health emergency, but it has social, economic and societal ramifications that go beyond the health issues. Millions of people were expected to lose their jobs and experience economic hardship, and suffer psychological strains as result of the lockdowns or self-quarantining (Brooks et al., 2020). Accordingly, an international group of behavioral scientists proposed new research domains that may be relevant to numerous stages of the COVID-19 pandemic (Van Bavel et al., 2020), from individuals' internal states to their societal attitudes and beliefs. However, there was a lack of research to indicate which domains were most relevant in the early stage of the present pandemic.

In recognizing the broad social, economic, and personal ramifications of the pandemic, only a small part of our survey focused specifically on constructs relevant to the domain of health behavior, such as the Health Belief Model (Janz & Becker, 1984; Sheeran & Abraham, 2005), and the theory of planned behavior (TPB; Ajzen, 2005). According to the Health Belief Model, two conditions must be met to motivate people to engage in COVID-19 infection prevention behavior: They have to believe

that they are at risk of contracting the virus, and that engaging in the recommended virus protection behaviors would be effective in reducing that risk (Janz & Becker, 1984). A further assumption of the model is that perceived effectiveness of a health behavior will be moderated by the perceived costs of engaging in that behavior. If the behavior is too effortful, people might not adopt it, even if they think that doing so would be effective. According to the theory of planned behavior (TPB, Ajzen, 2005; Armitage & Conner, 2001; Robin, McEachan, Conner, Taylor et al., 2010), a more general psychological theory of behavior prediction, intentions to engage in a specific behavior would be predicted by three constructs: attitude towards the behavior (advantages and disadvantages), subjective norms (e.g., what do people who are important to me expect me to do), and perceived behavioral control (i.e., will I be able to do it).

Relevant theories such as these focus on a small number of theoretical constructs and are typically silent about other potentially important predictors not germane to the specific processes described. In line with this limited focus, models based on such theories typically explain limited variance in the outcome variable. For example, a meta-analysis based on 185 independent tests of the TPB found that attitudes, subjective norms and perceived control explain 39% of the variance in intention, with intention accounting for 22% of variance in behavior (Armitage & Conner, 2001). Although this performance is perceived as relatively strong in the field of social science, it leaves substantial room to consider potential predictors from other research domains, whose importance may idiosyncratically increase at various stages of a pandemic context. Thus, rather than focus on variables that exclusively target the health behavior, the present analysis sought to identify specific as well as general predictors unrelated to COVID-19.

The Present Study

We set out to distinguish important individual- and societal- level indicators of infection prevention behavior using machine learning. Our analysis is based on data from a large-scale psychological survey, enriched with publicly available country-level secondary data sources. We applied machine learning using the random forests algorithm. We used random forests because of its competitive performance, computational inexpensiveness, and ease of interpretation (Strobl, Malley, & Tutz, 2009). Our results could help focus policy on important indicators for which a causal mechanism is known or highly plausible, or spur additional research to substantiate causal mechanisms for other important predictors.

Method

PsyCorona survey

The survey was distributed online during the initial March-May virus wave. The survey was translated from English into 30 languages by bilingual members of the research team. Parallel sampling strategies were employed: convenience sampling, snowball sampling, and paid sampling. Given that age and gender were identified early as population vulnerability characteristics to the virus (Centers for Disease Control and Prevention, 2020; Wenham et al., 2020), the self-selected samples in 20 countries were supplemented with statistically representative (by age and gender) samples (n ~ 1000 per country).[1] Key demographic variables, such as age, gender, education level, and religiousness were included as predictors. Country of residence was included as a categorical predictor. The full survey is available in the

---

[1] The panel firms Qualtrics Panels and WJX achieved age-gender representative samples in 20 countries: Argentina, Australia, Brazil, Canada, China, France, Germany, Italy, Japan, Netherlands, Philippines, Romania, Russia, Serbia, South Africa, South Korea, Spain, Turkey, United Kingdom, and the United States. Four additional countries only achieved gender representativeness, due to insufficient access to the 55+ age group in Greece, Indonesia, Saudi Arabia, and Ukraine.

supplementary material, as well as codebooks and translation procedures for all

languages (Tables S1 & S2). All analysis code is available online

([https://github.com/cjvanlissa/COVID19_metadata)](https://github.com/cjvanlissa/COVID19_metadata), and has been publicly

documented since the start of the project.

A summary table of all variables entered as predictors is available in the

supplementary material (Table S3). Although the PsyCorona study was designed and

implemented prior to Van Bavel and colleagues's (2020) discussion of candidate

domains of inquiry for pandemic behavior, it touches on nearly all of these topics,

including navigating threats, stress and coping, science communication, moral

decision-making, and political leadership. To maximize topic breadth, the survey used

brief, face-valid indicators, with three overarching themes: First were personal factors

that could affect individuals' capacity to respond to the virus, such as psychological

coping and outlook, loneliness and deprivation, subjective emotional states, well-

being, employment, and financial (in)security. A second theme pertained to social

attitudes and norms, including general beliefs and attitudes about society, migrant

attitudes and prejudice, perceived and preferred social norms for infection prevention,

and endorsement of extraordinary virus containment and its economic rescue

measures. A third theme pertained to virus-relevant personal concerns, values, and

tendencies, including social contact and leaving the home, as well as the dependent

variable of interest: self-reported engagement in voluntary, infection prevention

behaviors recommended by the WHO. Established scales were used where possible;

ad hoc items were developed when necessary to account for social psychological

processes that may be specific to the circumstance; these were either adapted from other scales or informed by extant theoretical conceptualizations.[2]

Infection prevention behavior. At the time the survey was constructed, three infection prevention behaviors were advised across most countries and contexts: washing hands, avoiding crowds, and self-isolation/self-quarantine[3]. Participants were presented with a single screen that read, *"to minimize my chances of suffering from coronavirus, I..."* and indicated their agreement to *"1. ...wash my hands more often"*, *"2. ...avoid crowded spaces"*, and *"3. ...put myself in quarantine/self-isolate"*, each rated on a seven-point scale rated -3 (strongly disagree) to 3 (strongly agree). These three items were combined into a scale representing infection prevention behaviors ($M = 2.20$, $SD = 1.00$, $\alpha = .75$).[4] Note that the items were specifically framed to assess the behavioral intent to reduce the risk of infection, consistent with theories of health behavior that people engage in self-protective actions because they are perceived as instrumental for threat reduction (Stroebe, 2011).

---

[2] Personal factors adapted or informed by prior work included affective states (incl. valence and arousal, Russell, 1980); boredom (Fahlman et al., 2011); coping and avoidance (Carver et al, 1989; Sexton & Dugas, 2008); financial strain (Selenko & Batanic, 2011); loneliness (Huges et al., 2012); neuroticism (Hahn et al., 2012); happiness and well-being (Abdel-Kahlek, 2006; Hershfield et al., 2016; Seligman, 2011); time perception, management, and temporal focus (Fahlman et al., 2011; Macan, 1994; Shipp et al., 2009), working conditions & job insecurity (Konovsky & Cropanzano, 1991; Porath et al., 2012; Van der Elst, De Witte, & De Cuyper, 2014). The social attitudes and norms domain included generic conspiracy beliefs and paranoia (Bruder et al, 2013; Schlier et al., 2016); immigrant attitudes (American National Election Studies, 2017; European Social Survey, 2014; Zavala-Rojas, 2014); norm perceptions and preferences (adapted from Gelfand, 2019); societal discontent and disempowerment (adapted from Leander et al., 2019). Virus-relevant personal concerns included virus-related descriptive and injunctive norms (Leander et al., 2011); virus-relevant beliefs and perceived knowledge, virus exposure risk and economic risk, and severity of virus and economic consequences (informed by Stroebe, 2011; Stroebe et al., 2017); trust in government pandemic communication and response (informed by Gelfand, 2019; Stroebe et al., 2020; Van Zomeren et al., 2008), and attitudes towards prosocial responses and extraordinary societal responses (cf. Van Zomeren et al., 2008). This list is not exhaustive; see Table S3 for a full list and item details.

[3] Wearing a face covering was not universally recommended by the WHO until June 2020 (WHO, 2020); the data for the present study was collected between March – May 2020. Prior to June, countries differed in regards to national health advice on mask-wearing.

[4] The data were left-skewed, but random forest models are non-parametric models and thus robust to non-normality.

Data enrichment and data cleaning

Before applying the machine learning models, we enriched the individual-level PsyCorona data with publicly available country-level datasets. These datasets were selected due to their international relevance for affording, shaping, or guiding individual-level behavioral responses to the virus: First, pandemic severity, as indicated by the number of cases, deaths, and recovered patients. Second, pandemic-related policies including both preexisting policies and ongoing governmental response to the COVID-19 pandemic. Third, pandemic preparedness.[5] Table 1 presents an overview of the databases. Our data collection period extended from March to May, with most participants completing the survey in March and April (see supplementary Figure S1 for daily frequencies); this time range afforded variability in the degree to which people in a given country were seeing cases and/or engaging in different containment policies. When applicable, respondent's country-level data were matched to their date of participation (e.g., confirmed cases, lockdown severity). Altogether, there were 115 predictors (80 survey factors, 35 country-level factors).

Table 1: Summary of country-level databases

Database	Description
1. Johns Hopkins University COVID-19 Data Repository Center for Systems Science and Engineering (CSSE).	Number of confirmed COVID-19 infections, deaths, and recoveries by date per country.[6]
2. Global Health Security (GHS) Index	Country-level ratings of pandemic preparedness and general health security.[7]
3. World Health Organization (WHO) and Organisation for Economic Cooperation and Development (OECD)	Country-level health care resources and health infrastructure.[8]

---

[5] We intended a fourth theme, mobility, which played an early role in transmitting the virus (Kraemer, Yang, Guitierrez, Wu, Klein et al., 2020). However, we dropped this variable because publicly available Google mobility data included a disclaimer that the data should not be used for research.

[6] Available at https://github.com/CSSEGISandData/COVID-19; Dong et al., 2020.

[7] Available at https://www.ghsindex.org/.

[8] Available at https://apps.who.int/gho/data/node.main.HWF and https://stats.oecd.org/index.aspx?queryid=30183.

4. World Bank: Worldwide Governance Indicators (WGI)	Per-country data on aggregate ratings of: Voice and accountability, regulatory quality, political stability and absence of violence, rule of law, government effectiveness, and control of corruption.[9]
5. Oxford COVID-19 Government Response Tracker (OxCGRT)	Governmental responses and policies with respect to COVID-19 by date per country.[10]

We subsequently cleaned the data in several steps. First, to ensure that there was enough data on country-level, we excluded observations from countries that accounted for less than 1% of total observations. The final sample included $N = 56,072$ respondents across 28 countries (see supplementary Table S4 for samples that remained in the data). Second, we excluded any columns and rows from the data that had a proportion of missing values of more than 20%. Third, we computed mean scores for multi-item scales using the `tidySEM` R-package (Van Lissa, 2020). For instance, responses to all 4 items on job insecurity (Van der Elst, De Witte, & De Cuyper, 2014) were summarized by creating a single composite score for job insecurity. Scales with low reliability were excluded (*Cronbach's alpha* $\geq$ .65; see supplementary Table S5).

Data analytic plan

Prior to analysis, we split our data by randomly assigning 70% of observations to a training set and 30% of observations to a test set (Hastie et al., 2009). The test set was reserved for unbiased evaluation of the model's predictive performance, and not used or examined during model building to prevent cross-contamination. Thus, all models were trained using the training set and evaluated using the test set. We applied a random forest model using the `ranger` R-package (Wright & Ziegler, 2015). Random forests offer competitive predictive performance at a low computational cost,

[9] Available at http://info.worldbank.org/governance/wgi/

[10] Available at https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.

intrinsically capture non-linear effects and higher-order interactions, offer a single variable importance metric for multi-level categorical variables (such as country), and afford relatively straightforward interpretation of variable importance and marginal effects of the predictors (Breiman, 2001). With regard to the multilevel structure of the data, random forests inherently accommodate data nested within country, including cross-level interactions where a given predictor has a different effect in different countries. As random forests allow for different associations between variables across country, it partially accounts for potential measurement invariance. Random forests are also robust to multicollinearity, though variable importance tends to be shared among multicollinear variables.

The forest included a total of 1000 trees with two tuning parameters: the number of candidate variables to consider at each split of each tree in the forest, and the minimum node size. The optimal tuning parameters were selected by minimizing the out-of-bag *mean squared error* (MSE) using model-based optimization with the R-package `tuneRanger` (Probst, Wright, & Bloulesteix, 2019). The best model considered 31 candidate variables at each split, and a minimum of six cases per terminal node.

The key output of random forests consists of 1) predictive performance, which reflects the model's ability to accurately predict new data; 2) variable importance, which reflects each predictor's relative role in accurately predicting the outcome measure, and 3) partial dependence plots, which indicate the direction and (non)linearity of a specific marginal effect (Breiman, 2001). Predictive performance is, essentially, a measure of explained variance (R^2), except that in the machine learning context, predictive performance is evaluated on the test sample, which was not used to estimate the model. Estimates of R^2 on the training sample should be

interpreted as a measure of descriptive performance (i.e., how well the model describes the data at hand), and can be (severely) positively biased when used as an estimate of predictive performance in new data. Given that we also recruited an age-gender representative subsample, we additionally computed predictive performance for the representative subsample of the test sample to better examine the generalizability of our findings to the target population.

The relative importance of predictor variables is based on permutation importance: Each predictor variable is randomly shuffled in turn, thus losing any meaningful association with the outcome, and the mean decrease in the model's predictive performance <u>after</u> permutation, as compared to the unpermuted model, is taken to reflect the (inverse) importance of that variable (Breiman, 2001).

The partial dependence plots are generated using the `metaforest` R-package (Van Lissa, 2018). Partial dependence plots display the marginal (bivariate) association between each predictor and the outcome. They are derived by computing predictions of the dependent variable across a range of values for each individual predictor, while averaging across all other predictors using Monte Carlo integration.

Results

For a complete archive of all analysis code and results, including fit tables and figures, see https://github.com/cjvanlissa/COVID19_metadata.

Total Variance Explained

The random forest model predicted approximately 52% of the variance in self-reported infection prevention behaviors. This was true for the cross-validated model performance using the training sample (training $R^{\wedge}2 = .518$), as well as both the full test sample ($R^2_{test} = .523$) and the age-gender representative test sample ($R^2_{rep} = .586$). The model performed somewhat better on the age-gender representative samples than

the full global sample that also included the convenience samples, which suggests that the findings generalize well.

The most important predictors are illustrated in Figure 1, ranked by relative importance and with an indication of whether the effect is generally positive, negative, or other (e.g., curvilinear). Table 2 serves as the legend for the variables illustrated in Figure 1. Note that Supplementary Table S3, which lists the summary statistics for each variable included in the analysis, is also sorted by relative importance, providing additional information about rank of each predictor.

Consistent with expectations, the most important predictors of infection prevention behavior included a mix of individual-level (survey) variables and country-level (database) indices. The partial dependence plots depicting the bivariate association between each predictor and outcome (Figure 2) indicate the precise patterns of each effect.

Figure 1. Machine learning results for self-reported personal infection prevention

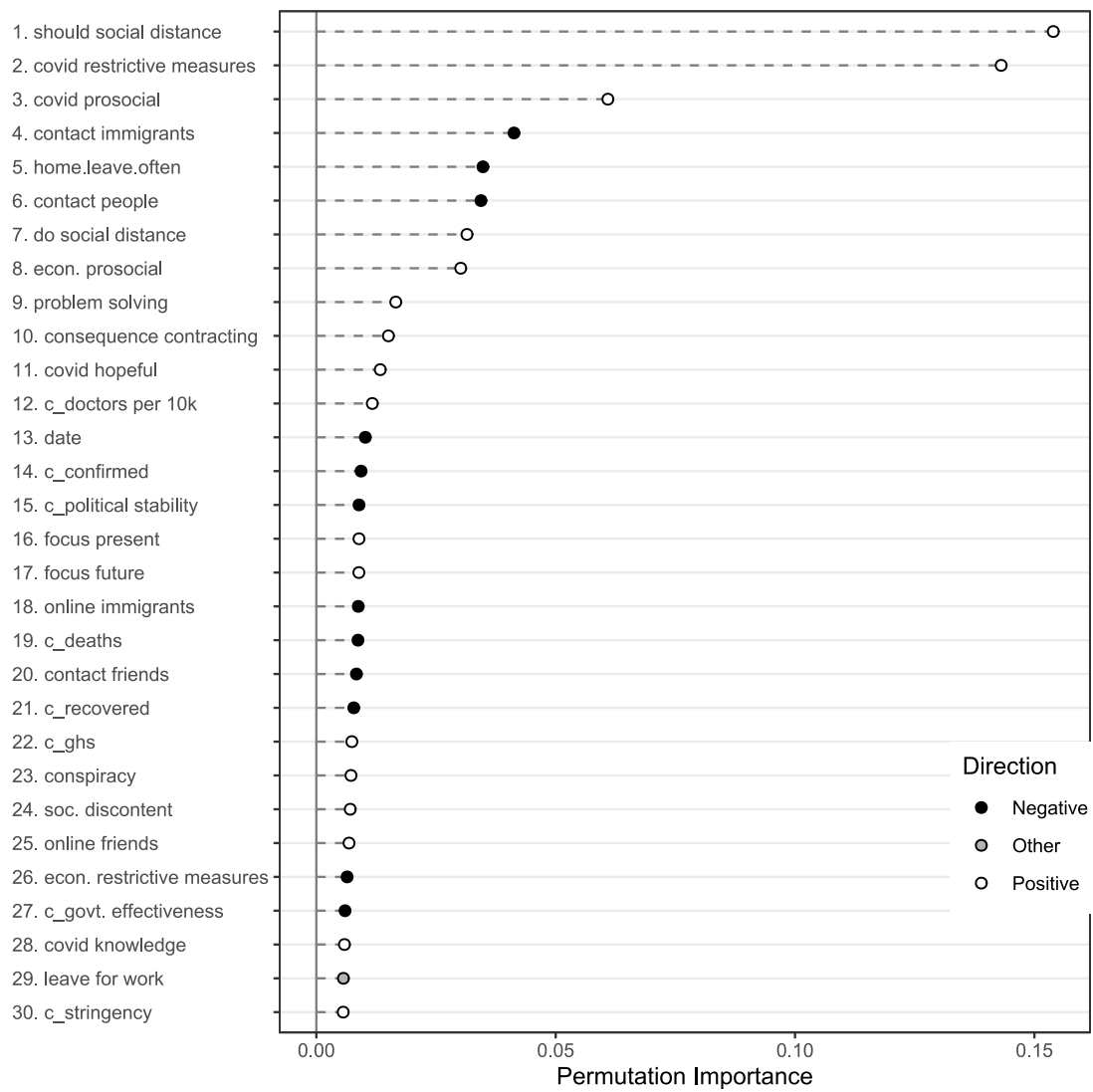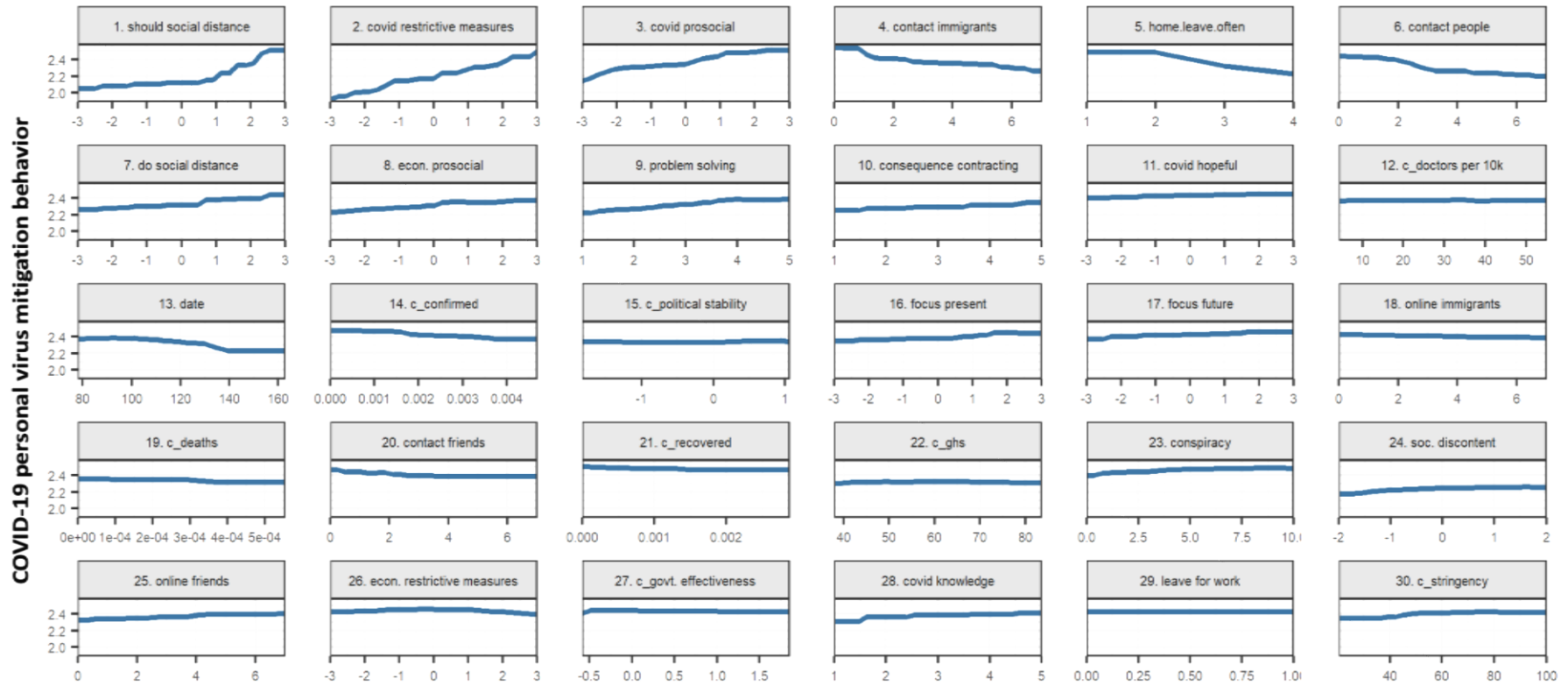behavior. Ranked in order of relative importance

Table 2. Brief descriptions of the top 30 predictors listed in Figure 1. Full variable

descriptions are in the supplementary material

	Variable	Brief description
1	Should social distance	Injunctive norm (Right now, people in my area..."-...should self-isolate and engage in social distancing.")
2	Covid restrictive measures	Support for severe collective virus containment measures (3 items: mandatory quarantines, mandatory vaccinations, report people suspected to be infected with COVID-19)
3	Covid prosocial	Pro-social willingness to protect vulnerable groups from the coronavirus (4 items)
4	Contact immigrants	Days of in-person (face-to-face) contact with immigrants
5	Home.leave.often	How many days in the last week did you leave your home?
6	Contact people	Days of in-person (face-to-face) contact with other people in general
7	Do social distance	Descriptive norm (Right now, people in my area..."-...do self-isolate and engage in social distancing.")
8	Econ prosocial	Pro-social willingness to protect vulnerable groups from economic consequences of the coronavirus (3 items)
9	Problem solving	Problem-focused coping style (3 items)
10	Consequence contracting	How personally disturbing would it be if… "You were infected with coronavirus"
11	Covid hopeful	"I have high hopes that the coronavirus situation will soon improve"
12	c_doctors_per10k	Number of doctors per 10,000 residents (Country-level; WHO)
13	Date	Date of survey participation (March 19-May 25).
14	c_confirmed	Number of confirmed coronavirus infections (Country-level; Johns Hopkins CSSE)
15	c_political stability	Political stability and absence of violence/terrorism (Country-level; WGI)
16	Focus_present	Temporal focus on the present moment
17	Focus_future	Temporal focus on the future
18	Online_immigrants	Days of online (virtual) contact with immigrants in the past week
19	c_deaths	Number of confirmed COVID-19 deaths (Country-level; Johns Hopkins CSSE)
20	Contact friends	Days of in-person (face-to-face) contact with friends & relatives in the past week
21	c_recovered	Number of confirmed COVID-19 recoveries (Country-level; Johns Hopkins CSSE)
22	c_ghs	Global health security index: pandemic preparedness and health security (Country-level). Source: Global Health Security Index
23	Conspiracy	Generic conspiracy beliefs (3 items)
24	Societal discontent	Concern about direction of society (3 items)
25	Online friends	Days of online (virtual) contact with friends & relatives in the past week
26	Econ. Restrictive measures	Support for extraordinary governmental intervention in economy (3 items)
27	c_govt. effectiveness	Government effectiveness (Country-level; WGI)
28	Covid knowledge	"How knowledgeable are you about the situation regarding the coronavirus?"
29	Leave for work	"In the past week, how often did you leave your house for work?"
30	c_stringency	Government COVID response tracker, measured across 17 policy indicators (Country-level): Source: OxCGRT

Notes: Full details of each measure are provided in Table S3 (https://osf.io/kxtjf/), as well as the survey codebook (https://osf.io/qhyue/?view_only=d60116c8090d4ec696bfaa9ea14b9432). Country-level variables are denoted with a c_ at the beginning of each variable name.

Figure 2. Partial dependence plots depicting bivariate associations between each variable and infection prevention behaviors.

Individual-level Predictors

Social Attitudes and Norms

By far, the most important indicators of infection prevention behaviors were individual-level beliefs and attitudes about how other people do and should behave, and how society should compel infection prevention behavior. The two strongest predictors were injunctive norms targeting infection prevention–namely, respondents' belief that people in the community *should* engage in social distancing and self-isolation (ranked 1st) , and their endorsement of extraordinary restrictive measures to contain the virus (mandatory quarantines and vaccination; reporting suspected infected individuals, ranked 2nd). The third strongest predictor was a pro-social willingness to protect vulnerable groups from the coronavirus (3rd). Respondents who complied with the norm to engage in infection prevention behaviors indicated that they wanted to do their bit to help other people to cope with the pandemic. Other, related indicators included the normative belief that people in one's community do self-isolate and engage in social distancing (ranked 7th), a pro-social willingness to limit the economic consequences of the coronavirus on others (8th), and support for extraordinary intervention in the economy (26th). Partial dependence plots indicate that the injunctive norm had a positive, approximately exponential, marginal relationship with the outcome measure, whereas the other indicators had positive approximately linear marginal relationships.

Social and Public Behavior

The next most important indicators were behavioral correlates of the dependent measure, namely, self-reported days in the last week that the individual engaged in social and public contact. Each of these behaviors had a negative linear relationship with the outcome variable. This includes the number of days that

respondents reported leaving home (5th), the number of days in the past week they had

in-person (face-to-face) contact with people who live outside their home, including

"…immigrants[11]" (4th), "…other people in general" (6th). and "…friends and

relatives" (20th). Higher in-person contact, which is not recommended under lock-

down rules, generally corresponded with less infection prevention behavior. In

contrast, online (virtual) contact, specifically with friends and relatives—which does

not violate social distancing rules—positively predicted infection prevention behavior

(ranked 25th).

Personal Psychological Factors

A third set of individual-level predictors thematically pertained to personal

and psychological resources and all had positive linear relationships with the outcome

variable: a problem-focused coping style (9th), high hopes that the coronavirus

situation would soon improve (11th), and a temporal focus on the present (16th) and/or

the future (17th). Consistent with theories of health behavior, the perceived personal

consequences of coronavirus infection ranked 10th (e.g., Protection Motivation

Theory; Rippetoe & Rogers, 1987; Rogers, 1983; Stroebe, 2011). Relatedly, self-

reported knowledge about COVID-19 (important for risk-assessment), ranked 28th.

Several individual-level variables rounded out the bottom of the list, which are

difficult to interpret. Having to leave one's house for work (ranked 29th) has a slight

negative association with infection prevention behavior, perhaps because having to

leave the house for extrinsic reasons hinders social distancing and self-isolation. The

positive association between conspiracy beliefs and infection prevention behavior

(ranked 23rd) might seem paradoxical, as one might expect a negative association if

we had measured whether respondents believed the virus was a hoax. However, this

---

[11] The questions about immigrants were included in the survey to assess specific research questions
about prejudice, unrelated to the present study.

variable instead assessed generic conspiracy beliefs (Bruder et al, 2013) – whether respondents believe that politicians do not always disclose the motives behind their decisions, that important things happen without public knowledge, and that government agencies closely monitor citizens. It makes sense that participants who endorse these beliefs tend to take infection prevention into their own hands.

Country-level Predictors

General Societal Conditions

Five (out of 9) general societal indices were ranked among the important indicators of infection prevention behaviors. The most important country-level predictor was a WHO/OECD indicator of national health care resources and infrastructure: the number of doctors per 10,000 inhabitants (ranked 12th). Other country-level predictors were the Global Health Security index (ranked 22nd), which pertains to pandemic preparedness and general health security, and two (out of six) World Governance Indicators: political stability (15th) and government effectiveness (27th). Country-level COVID-19 policy stringency (i.e., severity of lockdown conditions) ranked 30th, which potentially illustrates the limits of government lockdowns in compelling individual-level behavior, relative to other predictors.

COVID-19 Conditions

All three indicators of objective COVID-19 virus spread conditions were important predictors of personal infection prevention behavior–and in a negative direction: the cumulative number of confirmed COVID-19 cases (ranked 14th), deaths (19th) and recoveries (21st). All three patterns indicate that self-reported infection prevention behavior was lower among respondents who lived in countries with higher virus case counts, deaths, and recoveries on the day that they responded to the survey.

Calendar Date

Aside from the individual-level survey variables and country-level database indices, the calendar date of each survey response was a negative predictor (13th). Endorsement of infection prevention behavior generally decreased between March and May 2020.

Discussion

The machine learning analysis predicted over 50% of the variance in respondents' self-reported infection prevention behavior, far above the standards of social and health psychological theories. Moreover, whereas primary indicators in theoretical models are focused specifically on a target health behavior, this machine learning analysis additionally identified specific as well as general predictors unrelated to COVID-19 (e.g., temporal focus), thus offering complementary insights.

Who Complies with Infection Prevention Behavior?

A coherent picture emerged from our analysis of the type of person that showed early compliance virus infection prevention behavior recommendations. The underlying pattern of individual-level indicators points toward an intuitive understanding that infection prevention is a social dilemma and a conviction that the only way to fight the virus at present is widespread compliance with recommended behaviors. People appear to understand that factors such as personal risk (which was not indicated as highly important) is managed through expecting efforts from others. If everybody engaged in infection prevention behavior, the number of infected people in society would be reduced. Furthermore, if the people who did contract the virus maintained physical distancing, they would be less likely to infect others. This would explain why the strongest correlates of infection prevention behavior were beliefs that others in the community should engage in social distancing and self-isolation and that society should take restrictive measures to enforce that behavior, such as mandatory

quarantine, reporting people suspected to be infected, and (eventually) mandatory vaccination. Endorsement of these approaches to prevent infections seems to outweigh concerns that they infringe upon people's liberties and autonomy.

Given that being the only person complying with protective measures in an environment of non-compliers would undermine motivation to comply, it is reasonable that the descriptive normative belief that other people in the respondent's area were also compliant emerged as a strong predictor. Furthermore, according to their self-reports about their own behavior, compliant individuals did not engage in behavior that would be inconsistent with self-protections, such as leaving their homes or having personal contacts with other people. If they had contacts with their family and friends, it was not in face-to-face meetings, but online.

The findings also point to the idea that people who comply with recommended infection prevention behaviors are problem-solvers who have accepted that one way to combat the virus in the absence of a vaccine or cure is through these behaviors. That is, they tended to engage in a problem-focused coping style, focus on the present and the future (rather than dwell on the past), and maintained high hopes that the coronavirus situation would soon improve. This optimistic view is important because these individuals were likely aware of the costs of these infection prevention behaviors and perhaps needed psychological resources to alleviate these costs. In this vein, other strong predictors were a pro-social willingness to self-sacrifice to protect vulnerable groups from the virus, to limit the economic consequences of the coronavirus on such groups, and to support collective interventions in the economy such as tax increases. These patterns of data also explain the tension between members of society who do and do-not engage in updated recommendations. Because the largest predictor of infection prevention behaviors—at least those originally

recommended by the WHO—is the injunctive normative belief that one *should*

participate in the behaviors, people who do not engage in those behaviors are likely to

be seen as immoral, or at the very least norm-violators. Supportive of this, a large

British survey indicated in September 2020—three months after the WHO started to

universally recommended mask wearing—that 58% of the mask wearers in Britain

had severely negative attitudes towards those who do not wear masks and 68% of

Brits who complied with lockdown rules had strong negative views about lockdown

rule breakers. In fact, significant minorities who kept to the rules said that they

"hated" those who did not (Guardian, 2020).

The data also suggest that infection prevention behavior is a societal-level

challenge, in that individual-level compliance is more likely in a society that has the

political stability and health care infrastructure to take effective action to contain the

virus and treat people who have become infected. The findings regarding country-

level indicators are consistent with this analysis: government stability and

effectiveness, pandemic preparedness and health care resources (i.e., number of

doctors), pandemic preparedness and lockdown stringency, were all relatively

important indicators of infection prevention behavior.

Respondents in countries with higher confirmed COVID-19 infections, deaths,

and recoveries reported less adherence to infection prevention behavior

recommendations. Such findings might suggest reverse causality, as a country is

likely to experience increased pandemic severity if its citizens do not endorse

infection prevention behaviors Alternatively, it is possible that higher virus counts

demotivated infection prevention efforts—though, this assumes widespread

individual-level knowledge about virus rates. The fact that we observed self-reported

knowledge about COVID-19 to be an important positive indicator suggests that it is

more plausible that, in a society in which there is less compliance, there will be more infections, deaths, and recoveries.

Finally, one worrisome association is that respondents' date of participation emerged as an important predictor (13th) and had a negative linear relationship with personal health behavior. This suggests that between March and May, there was already a decrease in compliance with government advice. It could be that with time, self-isolation and social distancing became unbearable for many people. This is consistent with the popular notion of 'covid fatigue,' and highlights the need to study changing predictors as well as policies that might assist with long-term implementation of infection prevention behaviors.

Unexpected Absences from Top Indicators

After discussing the most important indicators of infection prevention behavior identified in our study, it is interesting to consider some of the other 85 variables that were not among the top indicators. Most surprising from a health psychological perspective is that the perceived personal likelihood of getting infected was not among the important predictors, though the perceived personal consequence of infection was ranked 10th. According to the Health Belief Model (Janz & Becker, 1984), perceived vulnerability and severity are both central to health threat appraisal. However, the absence of the perceived infection risk – but not of the perceived severity of getting infected - is consistent with the assumption that people intuitively realized that the pandemic constituted a social dilemma situation in which their personal risk was very much dependent on the behavior of others.

A number of other theoretically relevant variables absent from the most important predictors included loneliness and boredom, emotional and affective states experienced during the last week, subjective well-being, various forms of

psychological and financial strain, and job insecurity. It is important to note the present analysis does not rule out the importance of these personal factors for other outcomes, nor does it imply they do not predict infection prevention behaviors; rather the analysis did not identify them as being among the most important.

No demographic variables emerged as most important, even though several are associated with increased risk of complications from COVID-19. For instance, elderly people are at higher risk die from a COVID-19 infection and are therefore strongly advised to take great care (e.g., Centers for Disease Control and Prevention, 2020). Furthermore, there is reason to assume that social distancing and self-isolation present more of a social dilemma to young rather than elderly people, especially those on a pension. For young people, the costs of social distancing and self-isolation are typically higher and–because they usually recover more easily from a COVID-19 infection–the rewards of those infection prevention behaviors are smaller. It is therefore surprising that age was not identified as impactful by our analysis. However, this could simply suggest that any potential effects of demographics (e.g., age, gender, education, country) are explained through other factors.

Limitations, Strengths, and Future Directions

There are limitations to the analysis to address in future research. The survey was a cross-sectional self-report study, which raises questions about causality and common method bias among survey items (but not the country-level variables). However, the use of self-report data is not inherently problematic for our analysis given that the key outcome is self-reported infection prevention behavior. Furthermore, the purpose of the present analysis was not to infer causality but to identify potential variables of interest from a larger variable set. Based on these

results, future research may be most productive and informative for policy if it focuses on elucidating potential causal effects of the most important indicators.

The questionnaire used in this study was designed by an interdisciplinary consortium of scientists from many different countries. Thus, it is broader than a questionnaire would be that was guided by a singular theoretical perspective. For this reason, it was ideally suited for a machine learning analysis. Given this origin, it is fascinating that a theoretically coherent picture emerged from this analysis. However, it remains just a small slice of potentially important variables, and thus similar studies are recommended to identify other important omissions.

The results from an inductive approach should also be interpreted differently than results from deductive approaches (Brandmaier et al., 2016). Deductive methods are used to fit a theoretical model to data and to test its predictive validity. Inductive methods are used to identify important predictors from a large number of candidates. Deductive methods have parameters that are easy to interpret (e.g., regression coefficients) – but the model often explains little variance and the true population effects are rarely linear, so interpretation might not be so sensible. Inductive methods explain more variance, and in the case of random forests, also capture any non-linear effects and interactions. However, the results are harder to interpret. Thus, machine learning may be useful for gaining early insights and to draw attention to specific variables or theoretical areas, but it is neither exhaustive nor conclusive. Inductive analysis can complement theories, but the output it is not a theory in itself that can be generalized. Nevertheless, for the purposes of the present research, the combination of a psychological survey, database integration, and machine learning suffices to answer our research question: when simultaneously testing individual-level factors and and societal-level factors, both offered important indicators of infection prevention

behavior. If the aim of social and behavioral sciences is to maximally predict consequential behaviors, this study offers a possible method to that end.

Given that enforcement of infection prevention behaviors is difficult (e.g., social distancing) or impossible (e.g., hand washing) and requires long-term behavioral maintenance, recommendations are most likely effective if they are internalized by individuals and supported by societal-level factors. The picture that emerges from this analysis is that early compliance with infection prevention behavior recommendations is partly psychological and partly contextual. Our findings suggest a strong emphasis on norms—both injunctive and descriptive[12]—and the societal conditions enabling these norms. We began with an assumption that the pandemic is analogous to a public bad, in that COVID-19 is a social challenge that, in the absence of a cure or vaccine, can only be addressed if enough individuals engage in infection prevention behavior. In accordance with this assumption, both social beliefs and society-wide factors, rather than exclusively personal psychological states, emerged as the main predictors in our analysis.

---

[12] Whereas descriptive norms are beliefs about how to behave that guide behavior, injunctive norms are beliefs about how other people should behave.

References

Abdel-Khalek, A. (2006). Measuring happiness with a single-item scale. *Social behavior and Personality*, *34*(2), 139-150. doi: 10.2224/sbp.2006.34.2.139

Abraham, C., & Sheeran, P. (2005). The Health Belief Model. *Predicting health behavior (pp.* 28-80).Maidenhead, U.K.: Open University Press.

American National Election Studies, User's Guide and Codebook for the ANES 2016 Time Series Study (2017). https://electionstudies.org/wp-content/uploads/2018/12/anes_timeseries_2016_userguidecodebook.pdf;

Ajzen, I (2005). *Attitudes, personality and behavior 2nd. Ed.*. Maidenhead: Open University Press.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, *40*, 471-499. doi:10.1348/014466601164939

Bedford, J., Enria, D., Giesecke, J., Heymann, D. L., Ihekweazu, C., Kobinger, G., Lane, H. C., Memish, Z., Oh, M., Sall, A. A., Schuchat, A., Ungchusak, K., Wieler, L. H., & WHO Strategic Tech Advisory Grp. (2020.). COVID-19: towards controlling of a pandemic. *Lancet*, *395*, 1015–1018. doi:10.1016/S0140-6736(20)30673-5

Biel, A., & Thøgersen, J. (2007). Activation of social norms in social dilemmas: A review of the evidence and reflections on the implications for environmental behaviour. *Journal of economic psychology*, *28*, 93-112. doi: 10.1016/j.joep.2006.03.003

Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods*, *21*, 566–582. doi:10.1037/met0000090

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

    doi:10.1023/A:1010933404324

Brooks, S. K., Webster, R.K., Smith, L. E., Woodland, L., Wessely, S., Greenberg,

    N. et al. (2020). The psychological impact of quarantine and how to reduce it:

    rapid review of the evidence. *Lancet*, 395, 912-920. doi: 10.1016/S0140-

    6736(20)30460-8

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring

    individual differences in generic beliefs in conspiracy theories across cultures:

    Conspiracy Mentality Questionnaire. *Frontiers in psychology*, *4*, 225. doi:

    10.3389/fpsyg.2013.00225

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies:

    a theoretically based approach. *Journal of personality and social psychology,*

    56(2), 267. doi: 10.1037/0022-3514.56.2.267

Centers for Disease Control and Prevention (2020). COVID-19 guidance for older

    adults. Retrieved from https://www.cdc.gov/aging/covid19-guidance.html

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track

    COVID-19 in real time. *Lancet, 20,* P533-545.

Deutsch, M., & Gerard, H.B. (1955). A study of normative and informational social

    influence upon individual judgments. *Journal of Abnormal and Social*

    Psychology, *51*, 629-636. doi:10.1037/h0046408

European Social Survey, (2014). ESS Round 7 Source Questionnaire. London: ESS

    ERIC Headquarters, Centre for Comparative Social Surveys, City University

    London.

https://www.europeansocialsurvey.org/docs/round7/fieldwork/source/ESS7_so

urce_main_questionnaire.pdf

Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., Eastwood, J. D. (2011).

Development and validation of the Multidimensional State Boredom Scale

(MSBS). Assessment. Advance online publication. *20*(1), 68-85.

doi:10.1177/107319111142130310.1177/1073191111421303

Fitzsimons, G. M., & Finkel, E. J. (2011). Outsourcing self-regulation. *Psychological

Science*, *22*, 369-375. doi:10.1177/095679761037955

Gelfand, M. (2019). *Rule makers, rule breakers: Tight and loose cultures and the

secret signals that direct our lives*. Scribner.

Guardian (2020). Covid rules more divisive than Brexit, survey finds. Retrieved

September 11, 2020, https://www.theguardian.com/world/2020/sep/11/covid-

lockdown-rules-more-divisive-than-brexit-survey-finds

Hagger, M., Hardcastle, S., Chater, A., Mallett, C., Pal, S., & Chatzisarantis, N.

(2014). Autonomous and controlled motivational regulations for multiple

health-related behaviors: Between- and within-participants analyses. *Health

Psychology and Behavioral Medicine, 2*, 565-601.

doi:10.1080/21642850.2014.912945

Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of

personality–Validity and reliability of the GSOEP Big Five Inventory (BFI-S).

Journal of Research in Personality, *46*(3), 355-359. doi:

10.1016/j.jrp.2012.03.008

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning:

Data mining, inference, and prediction* (Second). Springer.

Hershfield, H. E., Mogilner, C., & Barnea, U. (2016). People who choose time over money are happier. *Social Psychological and Personality Science*, *7*(7), 697-706. https://doi-org/10.1177/1948550616649239

Hughes, M. E., Waite, L. J., Hawkley, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on aging*, *26*(6), 655-672. doi: 10.1177/0164027504268574

Janz, N. K., & Becker, M. H. (1984). The Health Belief Model: A decade later. *Health education quarterly*, *11*, 1-47. doi:10.1177/109019818401100101

Kahle, L. R., Beatty, S. E., & Homer, P. (1986). Alternative measurement approaches to consumer values: the list of values (LOV) and values and life style (VALS). *Journal of Consumer Research*, *13*(3), 405-409. doi: 10.1086/209079

Konovsky, M. A., & Cropanzano, R. (1991). Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. *Journal of applied psychology*, *76*(5), 698.

Stroebe, W., Kreienkamp, J., Leander, N. P., & Agostini, M. (2020, in press). Do Canadian and US American handgun owners differ? *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*. Advance online publication: https://psycnet.apa.org/record/2020-65584-001

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., ... & Brownstein, J. S. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*, 493-497. doi: 10.1126/science.abb4218

Leander, N. P., Chartrand, T. L., & Wood, W. (2011). Mind your mannerisms: Behavioral mimicry elicits stereotype conformity. *Journal of Experimental Social Psychology*, *47*(1), 195-201. doi: 10.1016/j.jesp.2010.09.002

Leander, N. P., Stroebe, W., Kreienkamp, J., Agostini, M., Gordijn, E. H., &
      Kruglanski, A. W. (2019). Mass shootings and the salience of guns as means
      of compensation for thwarted goals. *Journal of Personality and Social
      Psychology*, *116*(5), 704. doi: 10.1037/pspa0000150

Liebrand, W. B., & van Run, G. J. (1985). The effects of social motives on behavior
      in social dilemmas in two cultures. *Journal of Experimental Social
      Psychology, 21,* 86-102. doi:10.1016/0022-1031(85)90008-3

Macan, T. H. (1994). Time management: Test of a process model. *Journal of applied
      psychology*, 79(3), 381.

MedicalXpress (April 24, 2020). Coronavirus: Distancing and handwashing could
      lower flu rates too. Retrieved from: https://medicalxpress.com/news/2020-04-
      coronavirus-distancing-handwashing-flu.html

Olson, M. (2009). *The Logic of Collective Action: Public Goods and the Theory of
      Groups (2nd Printing):* Boston: Harvard University Press.

Oosterhoff, B., Palmer, C. A., Wilson, J., & Shook, N. (2020). Adolescents'
      motivations to engage in social distancing during the COVID-19 Pandemic:
      Associations with mental and social health. *Journal of Adolescent Health, 67,*
      179-185. doi:10.1016/j.jadohealth.2020.05.004

Porath, C., Spreitzer, G., Gibson, C., & Garnett, F. G. (2012). Thriving at work:
      Toward its measurement, construct validation, and theoretical refinement.
      Journal of Organizational Behavior, *33*(2), 250-275. doi: 10.1002/job.756

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and tuning
      strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining
      and Knowledge Discovery*, e1301. doi:10.1002/widm.1301

Robin, R., McEachan, C., Conner, M., Taylor, N. J. & Lawton. R. J. (2010).

  Prospective prediction of health-related behaviours with the Theory of Planned

  Behaviour: A meta-analysis. *Health Psychology Review, 5*, 97 - 144. doi:

  https://doi.org/10.1080/17437199.2010.521684

Rokeach, M. (1973). The nature of human values. Free press.; Rokeach, M. (1968).

  The Role of Values in Public Opinion Research. Public Opinion Quarterly,

  32(4), 547. doi:10.1086/267645;

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social*

  psychology, 39(6), 1161. doi: 10.1037/h0077714

Schlier, B., Moritz, S., & Lincoln, T. M. (2016). Measuring fluctuations in paranoia:

  Validity and psychometric properties of brief state versions of the Paranoia

  Checklist. Psychiatry Research, 241, 323-332.

Selenko, E. & Batinic B. (2011). Beyond debt, a moderator analysis of the

  relationship between perceived financial strain and mental health. *Social*

  Science and Medicine, *73*, 1725-1732. doi: 10.1016/j.socscimed.2011.09.022

Seligman, M. (2011). *Flourish: A new understanding of happiness, well-being, and*

  how to achieve them. Nicholas Brealey Pub.

Sexton, K. A., & Dugas, M. J. (2008). The cognitive avoidance questionnaire:

  validation of the English translation. *Journal of Anxiety Disorders*, *22*(3), 355-

  370. doi:10.1016/j.janxdis.2007.04.005

Shipp, A. J., Edwards, J. R., & Lambert, L. S. (2009). Conceptualization and

  measurement of temporal focus: The subjective experience of the past,

  present, and future. *Organizational Behavior and Human Decision Processes*,

  110(1), 1-22. doi: 10.1016/j.obhdp.2009.05.001

Strobl, C., Malley, K., & Tutz, G. (2009). An introduction to recursive partitioning:

      Rationale, application, and characteristics of classification and regression

      trees, bagging, and random forests. *Psychological Methods*, *14*, 323–48.

      doi.org/10.1037/a0016973.

Stroebe, W. (2011). Social psychology and health (3rd. ed.) Maidenhead: Open

      University Press.

Stroebe, W., Leander, N. P., & Kruglanski, A. W. (2017). Is it a dangerous world out

      there? The motivational bases of American gun ownership. *Personality and*

      Social Psychology Bulletin, *43*(8), 1071-1085. doi:

      10.1177/0146167217703952

Sutter, M. (2002). Public bad prevention by majority voting on redistribution–

      experimental evidence. *Group Decision and Negotiation*, *11*, 415-428.

      doi.org/10.1023/A:1020473620752

TechnologyNetwork.com (2020). The rationale behind social distancing and hand

      washing. Retrieved from:

      https://www.technologynetworks.com/immunology/articles/the-rationale-

      behind-social-distancing-and-hand-washing-332559

vanDellen, M. R., Shah, J. Y., Leander, N. P., Delose, J. E., & Bornstein, J. X. (2015).

      In good company: Managing interpersonal resources that support self-

      regulation. *Personality and Social Psychology Bulletin*, *41*, 869-882.

      doi:10.1177/0146167215580778

Van Bavel, J.J. et al. (2020). Using social and behavioural science to support COVID-

      19 pandemic response. Nature: Human Behaviour, 4, 460-471.

Van Lissa, C. J. (2018). *Metaforest: Exploring Heterogeneity in Meta-Analysis using Random Forests* (0.1.2) [R-package]. https://CRAN.R-project.org/package=metaforest

Van Lissa, C. J. (2020). *TidySEM: Generate tidy SEM-syntax* (0.1.0.5) [Computer software]. www.github.com/cjvanlissa/tidySEM

Van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. *Psychological bulletin*, *134*(4), 504. doi: 10.1037/0033-2909.134.4.504

Wenham, C., Smith, J., Morgan, R., et al (2020). COVID-19: The gendered impacts of the outbreak. *Lancet*, *395*, 846-848. doi:10.1016/S0140-6736(20)30526-2

World Health Organization (2020). Advice on the use of masks in the context of COVID-19: Interim guidance. https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the- community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak

Wright, M. N., & Ziegler, A. (2015). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *ArXiv:1508.04409 [Stat]*. http://arxiv.org/abs/1508.04409

Zavala-Rojas D. (2014). Thermometer Scale (Feeling Thermometer). In Michalos A.C. (eds), *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. doi: 10.1007/978-94-007-0753-5_1028