1 Title

2 COVID-Dynamic: A large-scale multifaceted longitudinal study of socioemotional and behavioral
3 change across the pandemic

4 Authors

5 Tessa Rusch^{*,a}, Yanting Han^{*,b}, Dehua Liang^{*,c}, Amber Hopkins^c, Caroline V. Lawrence^d, Uri
6 Maoz^{b,c}, Lynn K. Paul^{†,a}, and Damian A. Stanley^{†,+,a,e} and the COVID-Dynamic team[°]

7 * these authors contributed equally

8 † senior authors

9 + corresponding author: dstanley@adelphi.edu

10 ° see Figure 5 for a list of all members of the COVID-Dynamic team and their individual
11 contributions

12 ^a Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA,
13 USA

14 ^b Division of Biology and Biological Engineering, California Institute of Technology, Pasadena,
15 CA, USA

16 ^c Institute for Interdisciplinary Brain and Behavioral Sciences, Crean College of Health and
17 Behavioral Sciences, Schmid College of Science and Technology, Chapman University, Orange,
18 CA, USA

19 ^d Yale Law School, New Haven, CT, USA

20 ^e Derner School of Psychology, Adelphi University, 1 South Avenue, Garden City, NY 11530, USA

1 Abstract

2 The COVID-19 pandemic is a catastrophe that has caused societal upheaval globally. In the US,
3 it triggered an economic shock and laid bare societal inequities. However, it also afforded a unique
4 opportunity to elucidate the dynamics of psychosocial processes during extreme and volatile
5 conditions. From April 2020 through January 2021, we conducted a within-participant longitudinal
6 study—in 16 waves—to capture the COVID-19 experiences of >1000 US residents, together with
7 curated, external data to contextualize participants' responses. Each hour-long wave combined
8 standard psychological assessments with surveys of emotion, social/political/moral attitudes,
9 COVID-19-related behaviors, and tasks assessing implicit attitudes and social decision-making.
10 Here we characterize the sample, assess data quality, and showcase examples to demonstrate
11 the promise of this dataset, which is intended to be a resource for researchers interested in a
12 multitude of psychological questions, as well as clinicians and policy makers looking to mitigate
13 the psychological effects of future calamities.

14 Significance Statement

15 In the coming years, much will come to light about the impacts of the personal, social, economic,
16 and public-health challenges of 2020. Given that human memory is fallible, it is critical to record
17 the psychological and social experiences of individuals, in real-time, during periods of high
18 societal volatility. Here we introduce, describe, and assess the COVID-Dynamic dataset, a rich
19 collection of within-participant longitudinal data that captures the psychological states, personality
20 traits, attitudes, beliefs, and behaviors of more than 1000 US residents from April 2020 to January
21 2021. These data constitute a unique resource for scientists interested in both COVID-19-specific
22 research questions and basic psychological phenomena, as well as policymakers interested in
23 how we might do better in future disasters.

1 Introduction

The COVID-19 pandemic has been a devastating and ubiquitous global catastrophe. As of early 2021, (Figure 1) the pandemic had triggered tremendous societal upheaval in the United States with over 400,000 COVID-19-related deaths, an economic downturn not seen since the Great Depression^{1,2} and prolonged social-isolation, the detrimental effects of which will take years to understand. In mid-March of 2020, it was already clear that the societal effects of COVID-19 would be extreme, idiosyncratic, and highly dynamic, providing a unique opportunity to better understand the human psychological response to crisis. To capture this psychological moment in extensive detail, we developed the COVID-Dynamic longitudinal study (www.coviddynamic.caltech.edu): a multifaceted set of survey- and task- based measures (Table 1; see also <https://osf.io/sb6qx> for pre-registration), designed to characterize and quantify the dynamics of psychological, emotional, moral, attitudinal, and behavioral change over the COVID-19 pandemic. We were motivated by the following specific goals:

1) Characterizing within-participant psychological dynamics. To what extent do individual characteristics (e.g., states, traits, attitudes, etc.) change in the face of severe environmental stress? This question has clear clinical relevance and also addresses more basic questions about the stability of psychological characteristics.

2) Breadth of assessment. It was clear that, in addition to mental health, the pandemic would impact many other aspects of personal and social life in an unpredictable manner. Therefore, we were inclusive in our selection of measures. We assessed mental and physical health, behavior, personality, emotion, racial and political attitudes (implicit and explicit), social and moral decision-making, along with demographic detail, and more (see Table 1). To situate these data in context, we compiled a variety of state- and county- level measures (e.g., COVID-19- and restriction-related metrics, unemployment numbers, anti-racism-protest counts).

3) Methodological diversity. Recent work³ highlights the importance of methodological diversity for psychological studies. We therefore balanced standardized self-report psychological measures (clinical and basic), with novel targeted surveys, implicit measures, consequential social decision-making tasks, free response measures, and behavioral reports. As such, the dataset also provides a rare opportunity to explore within-subject differences related to measurement modality.

4) Sample representativeness. We tailored the survey recruitment to balance logistical feasibility with the goal of enrolling a representative sample. Recruitment was limited to the United States to optimize sampling density across geographic areas given funding constraints and researcher sensitivity to societal context. Recognizing that an exact representative sample was unattainable, we conducted a thorough audit of biases in our sample and calculated 'raking' weights, a post-stratification method to 'correct' for biases in sample demographics.

5) Rigor, Transparency, and Open Science. We strongly value scientific rigor, openness, and transparency. Extensive data quality metrics obtained during and after data collection are provided alongside the raw data. All materials pertaining to the study are archived with the study preregistration (<https://osf.io/sb6qx>). Because urgency rendered full preregistration of specific hypotheses impossible, we implemented a formal system to propose and register analyses prior to data access (see <https://osf.io/sb6qx>). Finally, after a 1-year period of embargo, these data will be publicly available.

The resulting COVID-Dynamic dataset is a unique, exceptionally rich, and inclusive longitudinal inventory (Figure 2) of the psychological, emotional, moral, attitudinal, and behavioral changes, as well as the personal experiences, of a large (1000+) cohort of U.S. residents from all 50 states during this unparalleled time. Here we describe the study design and methodology, detail the demographic composition of the sample, assess data quality, and characterize participant loss. We also showcase example data to demonstrate the promise of the dataset (Figure 4; and <http://covidynamicdash.caltech.edu/shiny/coviddash/>). These data will greatly improve our understanding of how traits, states, attitudes, beliefs, and behaviors are influenced by a year of profound individual, societal, and global events, and will be a valuable resource for researchers, clinicians, and policymakers.

Figure 1: Timeline of Real-World Events and COVID-Dynamic Wave Administrations. Visualization of the COVID-Dynamic data collection schedule in the context of the events of January 2020 to January 2021. Orange triangles denote each wave administration (black tick marks depict weekly intervals). The gray curve indicates the daily 7-day average of new, confirmed COVID-19 cases in the U.S.⁴, black encircled X's on top of the curve mark grim U.S. COVID-19-related death milestones (100,000 to 400,000 dead). The green line shows the monthly unemployment rate². The upper gradient (yellow-red) indicates the daily count of states with active stay-at-home restrictions (peak=41)³. The

lower gradient (blue-purple) shows the daily count of U.S. anti-racism crowd events⁵. Colored triangles below the gradients indicate local maxima for the various measures. All these external data (aligned to our data collection) are included in the dataset. Events of interest are indicated with vertical blue lines.

Figure 2: Summary of 376 COVID-Related Psychological Studies. Snapshot of COVID-19-related psychological studies publicly available as of August 11th, 2020 (see Methods for details on study identification) along 5 dimensions: (1) average follow-up frequency (y-axis; ‘cross-sectional’ denotes studies with a single time-point), (2) number of domains assessed (x-axis), (3) sample size (circle size; log-scale), (4) study location (circle color; within US in orange, non-US based in grey), and (5) inclusion of task-based (as opposed to survey-only) measures (circle opacity; opaque = tasks included, transparent = no tasks). Each circle represents an individual study; the COVID-Dynamic study is highlighted with a star.

Note that in contrast to the COVID-Dynamic study consisting of 16 discrete waves spread across 10 months, the vast majority of all summarized studies were conducted cross sectional, i.e., include only a single time point, or with very few repetitions, e.g., half or quarter-yearly.

A dynamic interactive figure with detailed study information is available at https://covidynamic.caltech.edu/documents/16453/summary_jittered.html.

1 Methods

2 *Participants*

3 Participant inclusion criteria: participants were recruited and screened through Prolific
 4 (www.prolific.co) based on the following inclusion criteria: aged 18-100, English fluency, United
 5 States residency (50 states), Prolific approval rating of 98% or higher, and minimum of 5 Prolific
 6 studies completed. To avoid drawing heavily from one geographic location, recruitment from the
 7 50 states was spread equally across 3 areas (East Coast, Middle US, West Coast). To counteract
 8 a bias towards younger adults, present in online studies, a portion of the recruitment specifically
 9 targeted individuals aged 45-100. In total, 1797 participants completed Wave 1 of the study (of
 10 the 1831 study openings made available on Prolific).

11 Participant exclusion criteria: To ensure study efficiency and data quality, for Waves 1-5 those
 12 participants who showed low levels of commitment and/or for whom data were deemed of poor
 13 quality, were excluded from continued recruitment following each wave. The exclusion criteria
 14 were as follows: unable to fully complete the experimental tasks in Wave 1 (27 participants),
 15 flagged on quality measures provided by Prolific (6 participants), missed >1 attention-check
 16 questions (15 participants); responded with less than 90% commitment to participation in future
 17 waves (94 participants; queried at beginning and end of each wave); and overall completion time
 18 below our minimum threshold (25, 30 or 35 minutes as determined for each wave through visual
 19 inspection of the study duration distribution and consensus among the COVID-Dynamic group)
 20 (94 participants). Two participants requested to be excluded from further data collection after
 21 Wave 3. From Wave 5 on, no participants were excluded. Application of these criteria resulted in
 22 the reduction of the study invitation list from 1797 to 1576 by the end of Wave 5.

23 Core participants: To provide an overview of the data from the 16 waves presented here, a core
 24 sample was created containing only those participants who had completed at least 50% of the
 25 waves and who had not met any of the exclusion criteria described above. This core sample
 26 consisted of 1177 participants (51.2 % female, median age = 39.4, age range = 18-77). Detailed
 27 demographic information for the Core sample is presented in Figure 3. Information about
 28 participants excluded from the Core Sample (i.e., the *excluded* and *low-completion rate* groups)
 29 is presented in the Results section. Sub-sample specific counts (core, exclude and low completion
 30 rate sample) are shown in Figure SI-1.

31 *Measures*

32 Self-Report Questionnaires

33 The full battery of questionnaires and tasks administered across the 16 waves is presented in
 34 Table 1 along with frequency of administration (see Table SI-1 for detailed descriptions of each
 35 measure). The self-report questionnaire battery included commonly used, published
 36 psychological assessment instruments and race-related surveys in addition to 2 publicly-available
 37 COVID19-specific instruments (Epidemic - Pandemic Impacts Inventory⁶, Scales from the Social
 38 Psychological Measurements of COVID-19⁷). We also administered a variety of self-report
 39 questionnaires created specifically for this study to characterize experiences and attitudes related

to COVID-19 and the BLM protests (Table 1, COVID & Protest Questionnaires), as well as demographic surveys and experimental questionnaires related to new and ongoing research (Table 1, Experimental Measures). A sample survey (including a randomly selected subset of items from the questionnaires and tasks) is also available for exploration (https://emotioncaltech.co1.qualtrics.com/jfe/form/SV_ctBqoufCi2MKnbv). Table SI-1 contains detailed descriptions of each measure and Table SI-2 contains a data dictionary of all survey items including possible responses for each item and scores generated for each measure. A continuously updated data dictionary for this study and an overview of measures included in ongoing data collection are presented on our website (<https://covidynamic.caltech.edu/research-projects>).

Tasks

In addition to questionnaires, participants completed a series of computer-based tasks designed to assess implicit social attitudes towards race (specifically Black/White/Asian) and pro-sociality, as well as estimations of COVID-related trustworthiness, and social decisions related to pro-sociality, group-cohesion, and altruism. The specific tasks included were:

- Pro-social Implicit Association Task (IAT): a version of the standard implicit association test⁸ that assesses implicit associations between self-identity (ME/THEY words) and pro-sociality (SERVICE/PROFIT words).
- Multi-category Brief Implicit Association Task (BIAT): a brief form of the implicit association test⁹ that assesses implicit associations in 3 subtests (administered in random order): 1) self-identity (ME/THEY words) and pro-sociality (SERVICE/PROFIT words); 2) evaluative (GOOD/BAD words) and White/Black People (WHITE/BLACK faces); and 3) evaluative (GOOD/BAD words) and White/Asian People (WHITE/ASIAN faces)
- Multi-race Affective Misattribution Procedure (AMP): this adaptation of the Multi-race Affect Misattribution Procedure¹⁰ assesses evaluative (PLEASANT/UNPLEASANT) implicit associations with different races (Asian, Black, and White).
- COVID-related Trustworthiness Rating Task: in this adaptation of the Trustworthiness Rating Task¹¹, participants are shown pictures of Asian/Black/White/other race faces and are asked to rate “How much do you trust this person to act responsibly with respect to the COVID-19 pandemic?” on a scale from 1 (not-at-all) to 9 (completely).
- Altruism Task: in this adaptation of the Altruism Task¹², participants make monetary decisions to distribute money equally or unequally (generous/selfish) amongst themselves and partners who vary with regard to age, race, occupation, and political identity.
- Public Goods Game¹³: a standard task from behavioral economics that assesses the tension between self-interest and group cohesion in an anonymous group fund-sharing scenario. Instructions for the public goods game were adapted from Wills et al¹⁴.

	WAVES	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Date	April 4	April 11	April 18	April 25	May 9	May 23	June 6	June 27	July 18	Aug 8	Aug 29	Sept 19	Oct 10	Nov 7	Dec 5	Jan 23
	Participant Count	1797	1498	1462	1391	1282	1178	1165	1100	1046	1029	987	932	942	908	900	876
Published Measures	Alcohol, Smoking, Substance Involvement Screening Test ¹⁵																
	Beck Depression Inventory - II ¹⁶																
	Connor-Davidson Resilience Scale ¹⁷																
	Disgust Scale – Revised ^{18,19}																
	External/Internal Motivation to Respond without Prejudice: Black & Asian ²⁰																
	Epidemic – Pandemic Impacts Inventory ⁶																
	Everyday Discrimination Scale ²¹																
	Free Will Inventory ²²																
	Humanitarianism-Egalitarianism ²³																
	Iowa Personality Disorder Screen ^{24,25}																
	Life Events Checklist for DSM-5 ²⁶																
	Monthly Substance Use Questionnaire ¹⁵																
	Major Experiences of Discrimination - Abbreviated ²⁷																
	NEO-FFI ^{28,29}																
	NIH toolbox: Emotional Support & Loneliness ³⁰																
	Perceived Stress Scale ³¹																
	Positive and Negative Affect Scales ³²																
	Posttraumatic Stress Disorder Checklist for DSM-5 ^{26,33}																
	Primary Care-PTSD Screen ³⁴																
	Scale of Ethnocultural Empathy ³⁵																
	Scales from the Social Psychological Measurements of COVID-19 ⁷																
	Social Dominance Orientation ³⁶																
	SNI Extended (2019) ³⁷																
	SNI Extended (current) ³⁷																
	State Trait Anxiety Inventory (state) ³⁸																
	State Trait Anxiety Inventory (trait) ³⁸																
	Very Short Authoritarianism Scale ³⁹																

	COVID & Protest Questionnaires	Experimental Measures	Tasks
	COVID Disgust & Fear		
	COVID Expectation Bias		
	COVID Experiences		
	COVID Explicit Norms		
	COVID Mask Questions		
	COVID Public Policy Knowledge		
	COVID Theories		
	COVID Vaccinations		
	COVID Vignettes		
	Post-Election 2020		
	Protest Surveys		
	Protest Survey – Election Related January 2021		
	Personal background – intake		
	Personal background - weekly & monthly updates		
	Consensus Task		
	Egocentric-Allocentric Task		
	Emotion Space		
	Race Thermometer		
	Responsibility Vignettes		
	Restriction Severity		
	Self-Continuity ⁴⁰		
	Verbal Self-Fluency ^{41,42}		
	Multicategory Brief Implicit Association Task ^{9,43}		
	Prosocial Implicit Association Task ^{8,44}		
	Multi-race Affective Misattribution Procedure ¹⁰		
	COVID-related Trust Rating Task ¹¹		
	Public Goods Gam ^{13,14}		
	Altruism Task ¹²		

Table 1: Prolific Testing Waves = W1-W16; 2 weeks = 2 dark lines between columns; 3 weeks = 3 dark lines between column; 4 weeks = wide dark line between columns; 7 weeks = wavy line between columns; Measurement domains: ● daily life impact of COVID pandemic, ● implicit attitude task, ● explicit attitudes and beliefs, ● emotion and mental health, ● social behaviors, ● decision making task, ● physical health, and ● participant demographics and background.

1 External Measures

2 In addition to data collected from participants, we have aggregated data from a variety of external
3 sources that can be used to understand the participant data in context. This includes:

- 4 1. State-wide coronavirus-related legal restrictions imposed to mitigate the spread of Covid-
5 19. These included stay-at-home or shelter-in-place orders, limits on gathering size, mask
6 mandates, restrictions on business or retail, and closures of public parks and playgrounds.
7 We gathered this data by evaluating the language in executive orders, public health
8 orders, and official statements by parks departments in place during each wave.
9 Specifically, we measured whether the restrictions were mandatory (often indicated by the
10 imposition of fines or punishments or language stating that citizens “shall” or “must”
11 comply) or suggested (often indicated by language stating that the restrictions were
12 “recommended”). All were coded at the state level. If different counties imposed different
13 restrictions, we noted the predominant restrictions in the state for that wave. Where these
14 statements were uninformative, we supplemented with contemporaneous news headlines
15 and by checking live resources maintained by the *New York Times*⁴⁵ or National Public
16 Radio⁴⁶. Masks and stay-at-home orders were coded as mandatory if an official executive
17 or public health order required them or imposed a fine for violation and recommended if
18 that state’s governor or public health department encouraged compliance, either in an
19 official order or in a press conference or online but did not formally impose the restriction.
20 Size restrictions of gatherings were coded on a scale of 1 to 11, borrowing aspects of
21 MultiState’s methodology⁴⁷. We assigned larger scores to states that recommended
22 rather than mandated the restriction, as assessed by language used in the executive
23 order; instituted larger actual gathering sizes; or opted to delegate gathering restrictions
24 to municipalities rather than setting a statewide limit. Parks were coded in one of three
25 options, all based on official statements from governors, public health departments, or
26 park departments: either fully closed to all visitors, partially closed -- for instance, to
27 camping or other activities requiring public facilities -- or open. Playgrounds were coded
28 separately as either open or closed, as this was largely left to municipalities to regulate
29 and official park statements often neglected playgrounds.
- 30 2. Public US-state-wise 7-day Covid-19 statistics, specifically the averages (over the week
31 preceding each wave’s release date) of cumulative and new daily confirmed Covid-19
32 cases and deaths, day-to-day changes (7-average of day-to-day differences) in new cases
33 and deaths extracted from the daily updated public New-York Times repository⁴
34 (<https://github.com/nytimes/covid-19-data>).
- 35 3. Anti-racism political crowd data from the Crowd Counting Consortium (CCC), which
36 collates publicly reported occurrences of political crowds in the United States (e.g.,
37 marches, protests, riots, or demonstrations)⁵. The CCC provides pre-sorted anti-racism
38 crowd data from May 2020 through July 2020. Additionally, we extracted anti-racism crowd
39 data for August 2020 through December 2020: Crowds whose description contained ‘raci’
40 were taken from the raw crowd data from the CCC database for these months and then
41 counter-protest crowds with ‘against anti-racism’ in their descriptions (22 crowds) were
42 removed.

Figure 1 provides an overview of COVID cases, key events in the US, stay-at-home restrictions, anti-racist crowd data, and other pertinent information.

Procedures

The study procedures were [preregistered](https://osf.io/sb6qx) on the Open Science Framework website (<https://osf.io/sb6qx>) to ensure transparency. All procedures were reviewed and deemed exempt (apart from the ASSIST) by the Internal Review Board of the California Institute of Technology (protocol 20-0992). The ASSIST was deemed not exempt and was approved following review by the Internal Review Board of the California Institute of Technology under a separate protocol (protocol 20-1014).

Recruitment through Prolific.co began on April 4, 2020. Invitations were sent to 1830 individuals, 1797 of whom completed Wave 1. Prolific randomly invites participants from their pool of eligible participants. To ensure geographic and age representation in the sample better reflected that of the US population, we stratified the distribution of invitations by these factors in the following manner (all invitees for each group participated unless indicated otherwise): all US states - ages 18-100 (30 invites), eastern US states - ages 18-100 (450 invites), eastern US states - ages 40-100 (150 invites), central US states - ages 18-100 (450 invites), central US states - ages 40-100 (150 invites), western US states - ages 18-100 (450 invites), western US states and ages 40-100 (110 out of 125 invites); Hawaii and Alaska and ages 18-100 (2 out of 25 invites). 5 additional participants are not included in any recruitment group. They completed the study via a direct survey link due to issues with administering the survey through Prolific. Each of the subsequent waves was posted on Prolific on a Saturday morning at 10am EDT and participants had 48 hours (until Monday at 10am EDT) to begin completing the week's survey. Each wave was designed to be completed in approximately 60 minutes and participants were allowed 140 minutes to complete it from the moment they accepted the task on Prolific.

Participants were paid at a standard rate of approximately \$10/h. Additionally, to reduce attrition, from Wave 5 onwards, 50 participants were randomly selected to receive a bonus of \$10 times the proportion of total waves they had completed. After each completed month, all participants were entered into a lottery for ten \$50 bonuses. Odds were determined by their choices in the public goods game and the altruism task.

A researcher monitored the data collection in real time on Saturdays and Sundays between 10am EDT and midnight EDT and Mondays from 8am EDT until the closing of the wave. Monitors provided feedback to participants through Prolific's anonymous messaging system and addressed any technical difficulties as they arose. From Wave 3 on, participants received two reminders to complete the wave during data collection (morning and afternoon on Sunday).

Waves 1 through 4 were conducted weekly (April 4, 11, 18, and 25 2020), followed by bi-weekly administration of Waves 5 through 7 (May 9 and 23, June 6, 2020) and Wave 8, 3-weeks later (June 27, 2020) (see Table 1). Subsequent waves have been administered at tri-weekly intervals and then once every 4 weeks.

Testing was administered using a combination of the Qualtrics survey platform (questionnaires) and Pavlovia⁴⁸ (tasks). While the specific content varied from wave to wave (see Table 1), the general sequence remained consistent. Participants first consented to participate and endorsed a level of commitment to the ongoing study. Next, they completed questionnaires regarding background, COVID experiences & mask use, experiences related to protests and racial discrimination, and the Emotion-Space experimental measure. This was followed by the COVID Vignettes (implicit assessment of COVID norms), then the experimental tasks (in randomized order with the BIAT always last), the Responsibility Vignettes, and then various standardized and experimental questionnaires in randomized order. Toward the end, the participants answered 3 questions about explicit norms related to COVID-19. In some waves, they also answered questions about how and where they believed COVID-19 originated. Finally, they were presented with the debriefing and a list of resources to aid with mental health issues, acute mental health crises and food shortages. Participants were also provided with a PDF file containing a resource list at the beginning of the survey.

To provide a comparison between our COVID-DYNAMIC project and other COVID-related psychological studies, we compiled information (Table SI-3) and created a [visualization](#) (Figure 2) of relevant studies from the following sources: a) longitudinal studies open for recruitment on the COVID-MINDS website as of August 11th, 2020⁴⁹, b) empirical papers related to COVID-19 on the COVID-MINDS website as of July 13⁵⁰, c) COVID-19 international social science research project tracker as of July 13⁵¹ and d) COVID-19 mental health preprints as of July 2⁵². We identified 565 psychological studies that collected, or are collecting, data from human participants. Table SI-3 provides a summary of the 376 studies which met these criteria: a) an English-language website was available with additional information, b) sample size was clear, and c) follow-up frequency was clear. The table includes the following information for each study, if applicable: study name; authors and affiliations; short summary; country of data collection; sample size (we tried to distinguish between planned sample size, sample size currently tested and core sample size across sessions for longitudinal studies whenever possible); characteristics about the population; psychological measures (which were further divided into several domains: demographics and background information; daily life impact; attitudes and beliefs; emotion and mental health; physical health; social behaviors; decision making tasks; implicit attitude tasks and other tasks); timeline of data collection; duration/follow up frequency; recruitment method; nature of participation (voluntary/paid/existing cohort); data availability and additional notes if not covered already. The visualization together with the summary list allows researchers to quickly identify and locate studies of interest.

Analysis

All data processing, analysis and visualization was conducted in Python 3.7^{53,54} and R⁵⁵.

Data Quality

To characterize the data quality, 15 quality indices (described below; 16 in Wave 1) were calculated for each wave and each participant and then compared to index-specific thresholds. A summary score for each participant of the percentage of quality indices deemed 'passed' was then calculated. Results for each individual index (both index value and the binary pass/fail

outcome; 1=fail) are provided in the data release. Note, these quality metrics were not used for participant exclusion. Participant exclusion was based solely on the criteria detailed in *Participants* and applied only to the first 5 waves. All quality metrics described below are provided with the data and researchers can apply those metrics for exclusion at their own discretion.

Several data quality indices were based on the mean number of consecutive identical responses in questionnaires and tasks (i.e., the mean response-string length). Each participant's mean response-string length was computed for each questionnaire/task in each wave. Participants with mean response-string length outside the range defined by the 1st quartile-3*IQR and the 3rd quartile+3*IQR were deemed outliers. The same criterion was applied to identify outliers in response consistency (see data quality index 9: "Response consistency" below).

Data quality indices for each participant and each wave included:

1. [Wave 1 only] Mismatch between the age the participant provided to Prolific when registering with the platform (continuously updated based on year of birth) and the age they indicated in Wave 1 of our study.
2. Completed less than 50% of waves (i.e., <8 waves).
3. Failed 2 or more attention questions.
4. Repetitive response behavior:
 - a. Mean response-string length deemed an outlier in greater than 50% of all questionnaires with more than 4 items
 - b. Mean response-string length deemed an outlier in greater than 50% of "core" questionnaires: Positive and Negative Affect Scales (PANAS), State Trait Anxiety Inventory (STAI, state and trait), NEO Five-Factor Personality Inventory (NEO-FFI), Very Short Authoritarianism (VSA), and Scale of Ethnocultural Empathy (SEE)
 - c. Mean response-string length deemed an outlier in the Multi-race Affective Misattribution Procedure
 - d. Mean response-string length was deemed an outlier in the Altruism Task (from Wave 4 on, prior to Wave 4 response key associations were not randomized and long response strings were considered plausible behavior).
5. Response speed:
 - a. More than 10% of RTs < 300ms in COVID-related Trust Rating Task
 - b. More than 10% of RTs < 300ms in Altruism Task
 - c. More than 10% of RTs < 300ms in Multi-race Affective Misattribution Procedure
6. Task data quality:
 - a. Excluded from Pro-social Implicit Association Task based on recommended IAT analysis criteria⁵⁶
 - b. Excluded from Multi-category Brief Implicit Association Task based on recommended BIAT analysis criteria⁴³

7. Free text responses: Valid free text responses were defined as responses that included at least one verb or noun per response. The number of nouns and verbs was automatically extracted using the Natural Language Toolkit⁵⁷.
 - a. Valid responses present in less than 50% of the questions asking participants to list stressors in their life.
 - b. Valid responses present in less than 50% of the questions asking participants to list memorable news items.
8. Frequency of missing (i.e., NA) responses greater than 50% in questions that include NA response options (e.g., “does not apply”, “prefer not to disclose”, or similar).
9. Response consistency: Response consistency was quantified via the mean difference score between regular and reverse scored items in STAI, SEE, and the Disgust Scale-Revised (DS-R) and mean score differences between PANAS positive and negative items and PANAS positive and PSS scores. Participants that were deemed an outlier (defined via the IQR, see above) in less than 50% of included comparisons are considered to pass this quality measure.

Survey responses, calculated scores, task performance indices, and data quality measures are available for data request.

Identification of Distinct Individual Participants for Visualization:

A key strength of this longitudinal dataset is that it facilitates studying individual differences in temporal variation of responses. To illustrate this potential, data from 5 individual participants from the core sample were selected for visualization in Figure 4. The particular set of 5 participants was selected using a maximum variation sampling procedure⁵⁸, which sampled participants by maximizing the sum of Euclidean distances between their scores on the five NEO-FFI subscales reflecting their personality traits. Specifically, the first participant in a set was one from the core sample. Subsequent participants were selected so that each new participant maximized the Euclidean distances from the previously selected participants in the 5-dimensional personality vector space. We repeated the sampling process for all possible initializations and selected the specific sample with the maximum sum of Euclidean distances. The patterns of each of the selected 5 participants’ NEO-FFI scores are illustrated in Figure 4A.

Sample Loss

To understand the factors that contributed to sample loss, we characterized participants that did not continue with the longitudinal study by comparing three participant sub-samples: (1) *excluded* participants - those who were excluded from further participation (Waves 1-5 only) based on the exclusion criteria listed in the participants section above; (2) participants with a *low-completion rate* - individuals that were not excluded, but completed less than 8 out of 16 waves; and (3) the *core* sample of participants - those who do not fall within the excluded or low-completion rate categories.

To assess if the participants in the *excluded* ($N = 221$) and *low-completion rate* ($N = 399$) groups differed in meaningful ways from the *core* sample ($N = 1117$), we compared the two omitted groups to the core sample on a number of demographic and psychological variables.

Demographic measures include age, birth sex, political party self-identification, highest education, weekly income bracket, and race and ethnicity. Psychological variables comprise STAI-trait, Beck Depression Inventory-II (BDI-II) and NEO-FFI scores.

For each of the omitted groups (*excluded* and *low-completion rate*), we quantified the difference between that group and the core sample by calculating the Euclidean distance between the samples using vectors representing the categories of each demographic factor and the difference in mean scores for each psychological factor. The true Euclidean distances and mean difference scores between the omitted groups and the core sample were compared against bootstrapped distributions of distances and mean scores of sub-samples drawn from the core sample (N samples = 10,000, N participants in bootstrapped samples [sampled without replacement] = N participants in excluded sample/ sample with low-completion rate). Results of these analyses are presented in Supplemental Materials (Figure SI-2 and Figure SI-3).

Data Raking

Data raking is a common post-stratification method used by survey researchers to “correct” for biases in sample demographics to make the sample more representative of the relevant population—the US population in this case. For researchers who wish to correct the sample in this manner, we provide raking weights for sex, age bracket, race and ethnicity for the core sample (N = 1177) according to their marginal distributions as given in the American Community Survey 5-Year Data⁵⁹. Weights were computed in R⁵⁵ using the *anesrake* package⁶⁰ with a 5% discrepancy limit and no trimming cap. Following the recommendations of Battaglia et al.⁶¹, we collapsed variable categories with less than 5% in the sample (race: American Indian/Alaska Native (0.09%), Native Hawaiian or Other Pacific Islander (0.09%), Multiracial (3.14%), Other (0.42%) and Prefer not to disclose (0.33%); age: 18-19 (3.2%; collapsed with 20-24)¹) and all age brackets ≥ 65 (summing to 5.3%). Further, we imputed missing ethnicity data (0.99% Prefer not to disclose) by randomly sampling from the two variable categories (Hispanic or Latinx and Not Hispanic or Latinx).

Results

Overall Data Quality

To summarize overall data quality, we calculated the number of participants that passed increasing percentages of our 15 data quality indices (16 indices in Wave 1; see Methods) per wave (Table 2). In all waves, the majority of participants (mean 70.59%; min=60.01%) passed all data quality metrics. Pass rates increased as the percentage threshold decreased, plateauing between 80% and 70% thresholds. Across all waves all participants passed at least 50% of all data quality metrics. Participant- and wave- wise scores for all quality measures as well as flags for missing task data are provided in the data release.

¹ Note that the US-census age data is provided in 5-year age brackets. That means our youngest participants (18-19) fall within the 15-19 bracket but cover only a small subset of that bracket. To account for the large mismatch between target and sample variable we used the census 20-24 bracket as target for the collapsed 18-24 age bracket in the sample.

Wave	pass 100%	pass >95%	pass >90%	pass >85%	pass >80%	pass >75%	pass >70%	pass >65%	pass >60%	pass >55%	pass >50%	total N
1	68.61	68.61	93.43	93.54	98.27	99.67	99.67	99.94	99.94	99.94	100	1797
2	60.01	60.01	87.45	87.52	95.66	98.13	98.13	99.53	99.93	99.93	100	1498
3	67.65	67.65	92.54	92.68	98.36	98.77	99.66	99.79	100	100	100	1462
4	70.45	70.45	94.68	94.82	98.63	99.35	99.78	100	100	100	100	1391
5	69.11	69.11	90.56	90.56	96.57	99.45	99.45	99.92	99.92	100	100	1282
6	80.22	80.22	95.16	95.16	98.73	99.58	99.58	100	100	100	100	1178
7	62.75	62.75	83.86	92.88	93.22	97.17	99.48	99.48	99.91	100	100	1165
8	72.09	72.09	94.27	94.36	98.36	99.64	99.64	100	100	100	100	1100
9	78.30	78.30	94.36	94.55	98.28	99.90	99.90	99.90	99.90	100	100	1046
10	71.33	71.33	93.00	93.00	98.25	99.32	99.32	100	100	100	100	1029
11	70.82	70.82	93.41	93.41	98.38	99.39	99.39	99.90	99.90	100	100	987
12	72.85	72.85	94.21	94.21	98.07	99.57	99.57	100	100	100	100	932
13	72.08	72.08	93.21	93.31	98.09	99.47	99.47	99.89	99.89	100	100	942
14	70.48	70.48	93.39	93.50	97.58	99.34	99.34	99.78	99.78	99.89	100	908
15	70.56	70.56	92.78	92.78	98.56	99.67	99.67	99.89	100	100	100	900
16	72.15	72.15	94.41	94.86	98.06	99.66	99.66	99.89	99.89	100	100	876

Table 2: Summary of Participant Performance on Data Quality Metrics. For each wave (table rows) the count of participants that completed that wave is displayed in the “Total Completed” column. The remaining columns show the percentage of participants in that wave that passed increasing percentages (50% to 100% in increments of 5%) of our 15 data quality indices. In each wave, more than half of all participants passed all quality measures and all participants passed at least 50% of our quality measures. From Wave 1 to 4, the proportion of participants that contributed high quality data monotonically increased. This can be attributed to the exclusion of participants based on three control measures in the first five waves (see Methods).

Results: Core Sample

The core sample of 1177 participants (51.23 % female, median age = 39.38; age range = 18-77) includes participants from all 50 states. Figure 3 compares core sample demographic distributions to the overall US population^{59,62}, including: US state of residence at Wave 1 (top left), distribution of age and birth sex at Wave 1 (top right), political self-identification (bottom left), race (bottom middle), and ethnicity (bottom right) at Wave 1. The distribution of participants’ state of residence

1 was reflective of the overall US population with a few notable instances of over-representation
2 (California, Oregon, and Washington) counterbalanced by more distributed under-representation.
3 The participant distribution of sex at birth reflected that of the US population (Table SI-7). As
4 expected, likely due to age-related differences in internet usage, older participants (>45) are
5 systematically under-represented in our sample, despite our efforts during recruitment (see
6 Methods) (Table SI-8). With regards to political party affiliation, our sample contains an over-
7 representation of Democrats with corresponding under-representation of Republicans and
8 Independents. Political leanings (liberal/conservative) were skewed according to party affiliation,
9 i.e., Democrats leaned more liberal and Republicans more conservative, with Independents
10 endorsing a more balanced range of leanings. Finally, regarding race and ethnicity, White, Asian,
11 and non-Hispanic/Latinx participants were over-represented in our core sample relative to the US
12 population, whereas Black, Multiracial and Hispanic/Latinx participants were under-represented
13 (Table SI-9 & 10).

Figure 3: Core Sample Demographics Compared Against US Population. Core sample demographics collected at Wave 1 compared to the overall US population (census data). At the top left, the state-wise population proportion is shown sorted in descending order according to the 2018 American Community Survey (ACS) estimate (blue) together with the proportions of state representation in the core sample (orange). On the map inset, zip-code data (collected in Wave 8) are displayed which, due to attrition, comprise only 77% of the core sample. At the top right, the age distribution for males and females in the core sample is displayed along with that of the US population. Age brackets follow the ACS brackets (Note: all study participants were over 18). On the bottom row, pie charts of Political Affiliation and Political Leaning (left), Race (middle), and Ethnicity (right) are displayed for the US population (inner circle) and for our core sample (outer ring; except for bottom left panel, where the outermost ring designates political leaning in our core sample in grayscale).

Figure 4: Group and Individual Temporal Trends from an Example Variable in Each Domain Assessed A) NEO-FFI scores (five-factor personality inventory - N: Neuroticism, E: Extraversion, O: openness to experience, A: Agreeableness, C: Conscientiousness) from five participants with maximally distant personality profiles (see Methods). B-H) Results at the individual- (5 exemplary participants; left) and group- (entire participant sample; right) levels for a single measure from each of the study domains. Measure scores are represented on the y-axis, time since beginning of data collection (in weeks) is represented on the x-axis). Orange triangles indicate weeks in which data were collected. In “Individuals’ Scores” graphs, shaded violin plots show the distribution of all 1000+ participant’s responses in each wave, overlaid colored lines indicate the individual scores from each of the five participants described in panel A. In “Average Scores” graphs, the black line indicates the mean response across all 1000+ participants in each wave, with 95% confidence intervals denoted by the shaded region. Panels B-H present one example measure from each domain. In each panel description below, the meaning of higher values for the presented score is given in parentheses: B) implicit attitudes, Asian/White + Good/Bad brief implicit association test (IAT), IAT D score (higher anti-Asian attitudes); C) daily life impact, COVID-19 related changes in daily behaviors, sum total of 9 items (greater changes in behavior); D) explicit attitudes and beliefs, very short authoritarianism (VSA) scale, VSA raw summary score (higher authoritarian tendencies); E) emotion and mental health, State-Trait Anxiety Inventory state scale (STAI-state), STAI-state raw summary score (higher state anxiety); F) social behaviors - National institutes of Health Loneliness scale, T-score (increased loneliness); G) decision making, Public Goods Game group-investments, tokens (amount invested with the group);

H) physical health - Epidemic-Pandemic Inventory (EPII) physical health total, sum of 8 items (higher number of non-COVID health problems endorsed).

To illustrate the richness of these data, and in particular the breadth of included measures and the variation over time and across individuals, we selected a subset of representative measures for visualization. Figure 4 depicts results for the entire sample of 1000+ participants and for five participants with maximally divergent personality profiles (see Methods) for one representative measure (i.e., item or summary score) from each measurement domain assessed in the study (i.e. daily life impact, attitudes and beliefs, emotion and mental health, physical health, social behaviors, decision making tasks, and implicit attitudes; note: demographics are omitted as these are visualized in detail in Figure 3). For each measure (Figure 4B-4H) and wave (x-axis), the distribution of individual scores (left panels) is depicted in a violin plot with the individual scores from the five exemplary participants superimposed (NEO-FFI profiles for each of the five individuals are shown in Figure 4A). The group-level plot for each measure (right panels) shows the mean response across all participants in each wave. Note: not all measures were collected every wave. Collectively, the data in Figure 4 demonstrate there exist substantial and meaningful variations in the data across the waves, both on average across participants and for individual participants. Hence, each measure has a clear pattern of evolution on average across participants over the time period displayed. Moreover, individual participants with distinct personalities and life situations had distinct trajectories within each measure.

Shiny App

This online visualization tool (<http://covidynamicdash.caltech.edu/shiny/coviddash/>) aims to provide researchers with a firsthand glimpse at the dataset. This tool was built using the Shiny-app framework⁶³ in R. It. Users have the option to filter the sample using demographic variables, visualize group mean temporal trends on a subset of measures included in this study, and examine within-wave Pearson correlations of any two measures in a selected wave. We encourage all users of the data to explicitly detail any data-exploration using this tool preceding formal analyses and hypothesis testing in their pre-registrations and publications.

Characterization of Excluded and Low-Completion Rate Participants

To understand what might distinguish the excluded and low-completion rate groups from the core sample, we examined their demographic (i.e., age, birth sex, race/ethnicity, political affiliation, education, and weekly income) and psychological (STAI-Trait and NEO-FFI acquired in Wave 1; BDI-II acquired in Wave 2) characteristics (see Figures SI-1, SI-2 and Tables SI-4 and SI-5 for detailed results). The analyses of the demographic data (see Methods for details) revealed that both male and female participants in the excluded and the low-completion group differed significantly from the core sample in age, with the core sample including significantly less participants in the lowest age brackets (18-29) than the other two groups. Excluded participants further differed from the core sample in racial and ethnical composition. The excluded group contained more Asian and American Indian/ Alaskan Native individuals than the core sample. Participants that completed less than 8 out of 16 waves further differed from the core sample in highest education level and weekly income: The proportion of participants with *some college*

education was significantly higher than in the core sample; and more participants from the low completion group fell in the lowest income bracket (\$250-499 per week) while in the core sample a larger proportion of participants fell in the second lowest income bracket (\$500-999 per week).

With respect to psychological characteristics, the low completion group scored higher values of depressive symptoms (BDI-II) than the core sample, as well as lower levels of conscientiousness and agreeableness, and higher levels of Neuroticism in the NEO-FFI. The excluded group endorsed less openness to experience, conscientiousness and agreeableness, and higher levels of neuroticism and on the NEO-FFI than the core sample, but no evidence of significant difference was found on any of the other tested psychological variables (see Figure SI-3 and Table SI-5).

Data Raking

For comparison with US population parameters and to facilitate post-hoc stratification, we determined summary raking weights for birth sex, age bracket, race, and ethnicity of the core sample according to their marginal proportions as given in the American Community Survey 5-Year Data⁵⁹ (see Methods). In addition to enabling us to account for sampling bias by ‘adjusting’ our data, this analysis also allowed us to quantify the areas in which representation in our core sample deviated most from the American population (Table SI-6 to SI-9). Raking weights revealed considerable age differences between our core sample and the US population. While 45 to 49-year-old individuals are most overrepresented in the sample (+11.8%), over 60-year-old residents are most underrepresented (-17.8%). Across races, sample-Census differences were largest in the Asian (-5.4%) and Black (+5.5%) groups. Ethnicity (Hispanic or Latinx/ not Hispanic pr Latinx) was misrepresented by (+/-7.9%). The US-population’s female/male proportion was well represented in the sample (+/-1.7% female/male). Participant-specific raking weights (mean = 1; SD = 0.89; min = 0.3; max = 5) are provided with the data release for those researchers who wish to utilize this.

Discussion

By mid-March of 2020, epidemiological predictions indicated that the COVID-19 pandemic would alter—if not define—our lives for years to come. Realizing that the psychological effects of this would be complex and far-reaching, but also highly dynamic and long-lasting, we launched a large, within-participant longitudinal study of the human psychological response to COVID-19. Many other studies have queried the psychological effects of the pandemic but, to our knowledge, none have captured a comparable breadth, sample size, testing frequency, and duration as the COVID-Dynamic study (see Figure 2, Tables 1 & SI-3). As such, the COVID-Dynamic dataset constitutes a unique resource for probing individual variation across many psychological constructs and relating that variation to a trove of information, ranging from individuals’ demographic and psychological contexts to their personal- and societal-level pandemic experiences.

The pandemic was also not the sole stressor of 2020. Sharp racial and ethnic disparities in pandemic-related health and economic outcomes^{64,65}, the horrific killing of George Floyd⁶⁶, and increased Anti-Asian harassment⁶⁷, both highlighted and likely exacerbated long-existing societal

fissures, fueling large anti-racism protests. Partisanship greatly increased during the 2020 presidential election culminating in an attack on the US capitol. From its inception the COVID-Dynamic study cast a wide net, assessing political beliefs, racial attitudes, social and moral norms, and social biases. The dataset documents individual variation in these highly societally relevant domains not only as the pandemic evolved, but also as racial and political tensions mounted across 2020.

A good illustration of the dataset's potential for relating real world events to temporal variation in psychological measures can be found in mean state anxiety (STAI-state; Figure 4E right panel). State anxiety peaked around Wave 2, shortly after the initial peak of new United States COVID-19 case-counts (see Figure 1). However, despite consistently elevated COVID-19 rates, mean anxiety scores decreased from Wave 2 to 5, perhaps reflecting overall psychological adjustment (e.g., coping and/or numbing). A second spike comes at Wave 7, following the killing of George Floyd. The qualitative relationship between this exemplary psychological variable and concurrent societal events speaks to the general value of this dataset. Second, on top of sample-level temporal variation, the dataset reveals large inter-individual differences, as illustrated across seven exemplary measures across five distinct individuals (Figure 4). Differences across measures in initial value and response trajectories from these five participants indicate substantial variation in their subjective experience of 2020. Once again state anxiety scores (STAI-state; Figure 4E left panel) reveal distinct individual patterns across time, with none resembling the sample trajectories. These individual patterns suggest that, in addition to societal-level events, individuals were influenced by a multitude of personal factors. Given the sample size and breadth of measures in this longitudinal dataset, the inclusion of self-reported personal experiences at each wave, as well as external data on collective events (e.g., COVID-case-rates, unemployment), researchers can explore these inter-individual differences and relate them to both personal and collective experiences.

The COVID-Dynamic project has prioritized scientific rigor throughout study design and data collection and has engaged multiple practices to ensure statistically sound and transparent analysis (see Methods and Results, SI-4 to SI-9, and <https://osf.io/uenc5/>). The data, procedures and preprocessing protocols have been carefully documented to facilitate data-sharing, and a formal proposal-based system for requesting access to subsets of the data was implemented (osf.io/uenc5). To-date more than 16 distinct projects have been approved by the COVID-Dynamic Team including: a published law essay on anti-masking laws and their contribution to racial disparities in usage of personal protective equipment⁶⁸, a longitudinal assessment of emotional experience across the pandemic (osf.io/t9p25), a characterization of change in attitudes towards Black- and Asian- Americans across 2020 (osf.io/q5jpb), a basic research study of the characteristics of states versus traits (osf.io/fhqp2), and an investigation of substance use trajectories and their relation to mental health (osf.io/5hbws).

While the strengths of these data are numerous, we also note limitations. The dataset does not contain within-participant pre-pandemic baseline measures apart from retrospective self-reports. Thus, some initial effects (e.g., acute initial increases in anxiety) are likely not captured by these data. However, the period of coverage includes multiple devastating local and national peaks of

COVID-19 cases, deaths, and restrictions, as well as many national events (e.g., BLM protests, US-presidential elections), and extensive documentation of personal experiences, yielding ample subjective variance to probe. Second, due to practical constraints, the data were sampled at irregular expanding intervals and a number of measures were not sampled every wave, posing unique analysis challenges. There are multiple methods for addressing these issues (e.g., data imputation^{69,70} and advanced linear modeling methods^{71,72}) which should be applied cautiously by individual researchers analyzing the data.

Additional limitations, particularly prevalent in anonymous online and longitudinal studies, include poor data-quality and spurious results driven by inattentive or fraudulent responders^{73,74}, as well as recruitment and retention biases. To ensure high data-quality, participants who were obviously inattentive or responded fraudulently in early waves were expelled. Further, we conducted detailed data quality assessments, provided with the dataset (see Methods and Results). Additional biases resulted from non-random, self-selected sampling and attrition (see Methods and Supplement). While these limit the generalizability of the COVID-Dynamic data to a broader population they have less impact on analyses addressing within-participant variation over time and relations between variables and psychological constructs. Nevertheless, we took steps to reduce recruitment and retention biases, including pseudo-stratification of recruitment (age and location), above-average study compensation, bonuses for continued participation, and continuous participant support during data-collection. Additionally, we compared our sample to the US census (see Tables SI-6 and SI-9) identifying recruitment biases in location, age, racial, ethnic, and political composition of the sample (see Figure 3 and Results), as well as retention biases with regards to age and race. For researchers wishing to conduct post-stratification adjustments on the sample we provide “raking” weights. We encourage researchers to take these factors into consideration during data analysis and interpretation.

Finally, as the declared goal of this study was to track psychological change across the pandemic, it is important to consider psychological biases in the sample. In accordance with previous research⁷⁵, we found the low completion rate group to be more anxious and depressed than the core sample. Furthermore, excluded and low completion rate samples were less conscientious and agreeable, and more neurotic than the core sample (Figure SI-3, Table SI-5). However, it is noteworthy that, relative to the community-based norms, the NEO personality factors²⁹, trait anxiety (STAI³⁸) and depression index (BDI-II¹⁶) averages for all groups (attrition and core) were within the normal range. By providing this detailed characterization of our sample, we aim to constrain interpretation as well as to situate our study in context and thereby provide a basis on which to integrate it with other valuable COVID-19-related psychological datasets (e.g. the [APS Global Collaboration on COVID-19](#)⁷⁶).

In the face of the extraordinary events of 2020, it is crucial that we, as a society, document the objective and *subjective* experiences of those affected—both to understand what happened and to plan for the future. The COVID-Dynamic dataset facilitates exactly that and is uniquely positioned to tackle myriad exciting and unanswered questions across many distinct psychological and social domains. Moreover, the insights gained—both across our participant population and within individual participants—may provide a knowledge base upon which to form

policy and organize action. Hence, on top of its contributions to basic psychological science, this dataset may also contribute to two clear goals of applied psychological research: mitigation of the impact of trauma and promotion of resilience in the face of adversity.

	Tessa Rusch ^h	Yanling Han ^b	Dehua Liang ^c	Amber Hopkins ^e	Caroline Lawrence ^d	Uri Maoz ^{b,c}	Lynn Paul ^a	Damian Stanley ^{a,a}	Ralph Adolphs ^a	R. Michael Alvarez ^a	Isabella Campionson ^o	Laura Harrison ^a	Denise Hien ^f	Tian Lan ^{a,c}	Chujun Lin ^{g,a}	Teresa Lopez-Castro [†]	Marie-Christine Nizzi ^u	Allison Rabin Golden ^c	Iman Wahle ^a	Gideon Yaffe [‡]
Conceptualization: Initial Inception of Covid-Dynamic Study	3	3	0	0	0	3	2	3	3	0	0	2	0	0	1	1	0	0	0	1
Conceptualization: Composition and Scope of Article	3	3	0	0	0	1	3	3	0	0	0	0	0	0	0	0	0	0	0	0
Data Curation: Cleaning/Preprocessing	1	1	3	0	2	1	3	2	0	1	0	0	0	0	0	0	0	0	0	0
Data Curation: Validation	3	0	2	0	0	1	1	2	0	1	0	0	0	0	0	0	0	0	0	0
Formal Analysis	2	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Funding Acquisition	0	0	0	0	0	3	0	1	3	1	0	0	1	0	0	1	0	0	0	2
Investigation: Data Collection	3	3	0	1	1	0	3	3	1	1	1	1	0	1	0	0	0	2	0	0
Investigation: Scraping External Data	1	2	1	2	2	0	0	1	0	3	0	0	0	0	0	0	0	0	1	0
Methodology: Questionnaire and Task Selection/Development	3	3	0	0	0	2	2	3	0	3	0	2	1	0	0	2	1	0	0	1
Methodology: Questionnaire and Task Implementation	3	2	0	0	0	0	3	3	0	1	1	1	0	1	0	0	1	0	0	0
Project Administration	3	3	0	2	0	0	3	3	1	1	0	0	1	1	0	1	0	1	0	0
Supervision	2	2	0	0	0	2	2	3	1	0	0	0	0	0	0	1	1	0	0	0
Software: Online Data Tool Visualizations	0	2	3	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Writing: Original draft	3	2	2	1	1	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0
Writing: Review & Editing	3	2	2	1	0	2	3	3	1	0	0	0	0	0	0	0	1	0	0	0

Figure 5: Contribution Table. The main authors' contributions are displayed in the left section of the table in the same order as the author list. The contributions of COVID-Dynamic-team members that are not listed individually in the author list are displayed in the right section of the table in alphabetical order. Contribution strengths are indicated numerically: 0 (no contribution), 1 (support), 2 (medium), 3 (lead); and color-coded with darker colors indicating stronger contributions.

Affiliations: ^a Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA; ^b Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA; ^c Institute for Interdisciplinary Brain and Behavioral Sciences, Crean College of Health and Behavioral Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, USA; ^d Yale Law School, New Haven, CT, USA; ^e Derner School of Psychology, Adelphi University, Garden City, NY, USA; ^f Center of Alcohol & Substance Use Studies, Graduate School of Applied and Professional Psychology, Rutgers University-New Brunswick, Piscataway, NJ, USA; ^g Department of Psychology and Brain Sciences, Dartmouth College, Hanover, NH, USA; United States of America; ^h Psychology Department, The City College of New York, New York, NY; ⁱ Arts and Science Faculty, Dartmouth College, Hanover, NH, USA; United States of America.

Resources

- [Web-based data explorer](http://coviddynamicdash.caltech.edu/shiny/coviddash/): <http://coviddynamicdash.caltech.edu/shiny/coviddash/>
- [Information about accessing data](https://coviddynamic.caltech.edu/data-sharing): <https://coviddynamic.caltech.edu/data-sharing>
- [Summary of COVID-19 Psychological Studies](https://coviddynamic.caltech.edu/resources/other-covid-studies): <https://coviddynamic.caltech.edu/resources/other-covid-studies>

d) [2020 Timeline](https://covidynamic.caltech.edu/resources/timeline-2020-world-events): <https://covidynamic.caltech.edu/resources/timeline-2020-world-events>

Acknowledgment

This dataset is the result of a collaborative effort by the entire COVID-Dynamic team (in alphabetical order): Ralph Adolphs, R. Michael Alvarez, Ruby Basyouni, Isabella Camplison, Yanting Han, Denise Hien, Amber Hopkins, Jonathan Katz, Tian Lan, Caroline Lawrence, Dehua Liang, Chujun Lin, Teresa Lopez-Castro, Uri Maoz, Marie-Christine Nizzi, Lynn K. Paul, Allison Rabkin Golden, Tessa Rusch, Damian A. Stanley, Iman Wahle, and Gideon Yaffe. A detailed characterization of each team member's contribution is given in the contribution table (Figure 5).

This work was supported by grants from the National Institute of Mental Health (2P50MH094258), the Caltech Chen Neuroscience Institute, and the Caltech Merkin Institute (RA), by the Oscar M. Ruebhausen Fund at Yale Law School (GY), the Rutgers Center of Alcohol & Substance Use Studies (DH), the John Templeton Foundation and the Kay Family COVID-19 Rapid Response Research Awards at Chapman University (UM), and the Colin Powell School for Civic and Global Leadership, The City College of New York (TLC). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

Bibliography

1. Shiller, R. Why We Can't Foresee the Pandemic's Long-Term Effects. *The New York Times* <https://www.nytimes.com/2020/05/29/business/coronavirus-economic-forecast-shiller.html>.
2. Bureau of Labor and Statistics. <https://www.bls.gov/data/> (2020).
3. Eisenberg, I. W. *et al.* Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat. Commun.* 10, 2319 (2019).
4. Times, T. N. Y. Coronavirus in the U.S.: Latest Map and Case Count. *The New York Times* (2020).
5. Crowd Counting Consortium. <https://sites.google.com/view/crowdcountingconsortium/view-download-the-data>.
6. Grasso, D. J., Briggs-Gowan, M. J., Ford, J. D. & Carter, A. S. *The Epidemic—Pandemic Impacts Inventory (EPII)*. (2020).
7. Conway, I. I. I., Lucian G., Woodard, S. R. & Zubrod, A. Social Psychological Measurements of COVID-19: Coronavirus Perceived Threat, Government Response, Impacts, and Experiences Questionnaires. (2020).
8. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480 (1998).
9. Sriram, N. & Greenwald, A. G. The Brief Implicit Association Test. *Exp Psychol* 56, 283–294 (2009).
10. Payne, B. K., Cheng, C. M., Govorun, O. & Stewart, B. D. An inkblot for attitudes: Affect misattribution as implicit measurement. *J. Pers. Soc. Psychol.* 89, 277–293 (2005).

11. Stanley, D. A., Sokol-Hessner, P., Banaji, M. R. & Phelps, E. A. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proc Natl Acad Sci U A* 108, 7710–7715 (2011).
12. Hutcherson, C. A., Bushong, B. & Rangel, A. A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron* 87, 451–462 (2015).
13. Ledyard, J. O. Public Goods: A Survey of Experimental Research. 101 (1994).
14. Wills, J. *et al.* Dissociable contributions of the prefrontal cortex in group-based cooperation. *Soc Cogn Affect Neurosci* 13, 349–356 (2018).
15. Wolff, N. & Shi, J. Screening for Substance Use Disorder Among Incarcerated Men with the Alcohol, Smoking, Substance Involvement Screening Test (ASSIST): A Comparative Analysis of Computer-Administered and Interviewer-Administered Modalities. *J Subst Abuse Treat* 53, 22–32 (2015).
16. Beck, A. T., Steer, R. A. & Brown, G. K. *BDI-II, Beck Depression Inventory: Manual*. (Psychological Corporation, 1996).
17. Connor, K. M. & Davidson, J. R. T. Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depress Anxiety* 18, 76–82 (2003).
18. Haidt, J., McCauley, C. & Rozin, P. Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Individ Dif* 16, 701–713 (1994).
19. Olatunji, B. O. *et al.* The Disgust Scale: item analysis, factor structure, and suggestions for refinement. *Psychol Assess* 19, 281–297 (2007).
20. Plant, E. A. & Devine, P. G. Internal and external motivation to respond without prejudice. *J Soc Psychol* 75, 811–832 (1998).
21. Williams, D. R., Yan Yu, Jackson, J. S. & Anderson, N. B. Racial Differences in Physical and Mental Health: Socio-economic Status, Stress and Discrimination. *J Health Psychol* 2, 335–351 (1997).
22. Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C. & Ross, L. T. The free will inventory: measuring beliefs about agency and responsibility. *Conscious Cogn* 25, 27–41 (2014).
23. Katz, I. & Hass, R. G. Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *J Soc Psychol* 55, 893–905 (1988).
24. Langbehn, D. R. *et al.* The Iowa Personality Disorder Screen: development and preliminary validation of a brief screening interview. *J Disord* 13, 75–89 (1999).
25. Trull, T. J. & Amdur, M. Diagnostic efficiency of the Iowa personality disorder screen items in a nonclinical sample. *J Disord* 15, 351–357 (2001).
26. Weathers, F. W. *et al.* *The life events checklist for DSM-5 (LEC-5)*. (Instrument available from the National Center for PTSD at www.ptsd.va.gov, 2013).
27. Sternthal, M. J., Slopen, N. & Williams, D. R. RACIAL DISPARITIES IN HEALTH. *Bois Rev. Soc. Sci. Res. Race* 8, 95–113 (2011).
28. McCrae, R. R. & Costa, P. T. The NEO-PI/NEO-FFI manual supplement. *Odessa FL Psychol. Assess. Resour. Google Sch.* (1989).
29. NEO-FFI subcomponents: Norms and scoring. <https://pages.uoregon.edu/gsaucier/NEO-FFI%20subcomponent%20norms%20and%20scoring.htm>.

30. Gershon, R. C. *et al.* NIH toolbox for assessment of neurological and behavioral function. *Neurology* 80, S2-6 (2013).
31. Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *J. Health Soc. Behav.* 24, 385–396 (1983).
32. Watson, D., Clark, L. A. & Carey, G. Positive and negative affectivity and their relation to anxiety and depressive disorders. *J Abnorm Psychol* 97, 346–353 (1988).
33. Weathers, F. W. *et al.* The PTSD checklist for DSM-5 (PCL-5). *Scale Available Natl. Cent. PTSD Www Ptsd Va Gov* 10, (2013).
34. Prins, A. *et al.* The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): Development and Evaluation Within a Veteran Primary Care Sample. *J Gen Intern Med* 31, 1206–1211 (2016).
35. Wang, Y.-W. *et al.* *The Scale of Ethnocultural Empathy: Development, validation, and reliability.* *Journal of Counseling Psychology* vol. 50 (2003).
36. Ho, A. K. *et al.* The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO_r scale. *J. Pers. Soc. Psychol.* 109, 1003–1028 (2015).
37. Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S. & Gwaltney, J. M., Jr. Social ties and susceptibility to the common cold. *JAMA* 277, 1940–1944 (1997).
38. Spielberger, C. D. *Manual for the State-trait Anxiety Inventory (form Y) (“self-evaluation Questionnaire”).* (Consulting Psychologists Press, 1983).
39. Bizumic, B. & Duckitt, J. Investigating right wing authoritarianism with a very short authoritarianism scale. *J Soc Polit Psych* 6, 129–150 (2018).
40. Nizzi, M.-C. *et al.* From armchair to wheelchair: how patients with a locked-in syndrome integrate bodily changes in experienced identity. *Conscious. Cogn.* 21, 431–437 (2012).
41. Bellet, B. W. *et al.* Identity confusion in complicated grief: A closer look. *J. Abnorm. Psychol.* 129, 397–407 (2020).
42. Nizzi, M.-C. Assessing the sense of self (Unpublished doctoral dissertation). *Dep. Psychol. Harv. Univ. Camb. MA* (2018).
43. Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J. & Greenwald, A. G. Understanding and Using the Brief Implicit Association Test: Recommended Scoring Procedures. *PLOS ONE* 9, e110938 (2014).
44. Marvel, J. D. & Resh, W. D. An Unconscious Drive to Help Others? Using the Implicit Association Test to Measure Prosocial Motivation. *Int. Public Manag. J.* 22, 29–70 (2019).
45. See Coronavirus Restrictions and Mask Mandates for All 50 States - The New York Times. <https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html>.
46. Treisman, R. West: Coronavirus-Related Restrictions By State. *NPR.org* <https://www.npr.org/2020/05/01/847416108/west-coronavirus-related-restrictions-by-state> (2020).
47. MultiState. COVID-19 State Reopening Guide. *MultiState* <https://www.multistate.us/issues/covid-19-state-reopening-guide>.
48. Pavlovia. <https://pavlovia.org/>.
49. COVID Minds Longitudinal Studies. *COVID-MINDS* <https://www.covidminds.org/longitudinalstudies>.
50. COVID Minds Empirical Papers. <https://www.covidminds.org/empiricalpapers>.

- 1 51. COVID-19 international social science research project tracker.
2 [https://docs.google.com/spreadsheets/d/1DuY8VLV2yG8TAOZbtyNtrhHERmzUjSaXpAo0_59z_](https://docs.google.com/spreadsheets/d/1DuY8VLV2yG8TAOZbtyNtrhHERmzUjSaXpAo0_59z_XU/edit?ts=5e73b15d#gid=60477339)
3 [XU/edit?ts=5e73b15d#gid=60477339](https://docs.google.com/spreadsheets/d/1DuY8VLV2yG8TAOZbtyNtrhHERmzUjSaXpAo0_59z_XU/edit?ts=5e73b15d#gid=60477339).
- 4 52. COVID-19 mental health preprints. *medRxiv: The Preprint Server for Health Science*
5 <https://www.medrxiv.org/search/mental%252Bhealth%252BCOVID>.
- 6 53. Van Rossum, G. & Drake, F. L. *Python 3 Reference Manual: (Python Documentation*
7 *Manual Part 2)*. (CreateSpace Independent Publishing Platform, 2009).
- 8 54. Van Rossum, G. & Drake, F. L. *Introduction to Python 3: (Python Documentation Manual*
9 *Part 1)*. (CreateSpace Independent Publishing Platform, 2009).
- 10 55. Team, R. C. *R: A language and environment for statistical computing*. (Vienna, Austria,
11 2013).
- 12 56. Lane, K. A., Banaji, M. R., Nosek, B. A. & Greenwald, A. G. Understanding and Using
13 the Implicit Association Test: IV: What We Know (So Far) about the Method. in *Implicit*
14 *measures of attitudes* 59–102 (The Guilford Press, 2007).
- 15 57. Bird, S., Klein, E. & Loper, E. *Natural Language Processing with Python: Analyzing Text*
16 *with the Natural Language Toolkit*. (“O’Reilly Media, Inc.”, 2009).
- 17 58. Qualitative Research & Evaluation Methods. *SAGE Publications Inc*
18 <https://us.sagepub.com/en-us/nam/qualitative-research-evaluation-methods/book232962>
19 (2021).
- 20 59. Bureau, U. S. C. Selected housing characteristics: 2013–2017 American community
21 survey 5-year estimates. (2017).
- 22 60. Pasek, J. anesrake: ANES raking implementation. R package version 0. 80. (2019).
- 23 61. Battaglia, M. P., Hoaglin, D. C. & Frankel, M. R. Practical Considerations in Raking
24 Survey Data. *Surv. Pract.* 2, 2953 (2009).
- 25 62. Gallup. Gallup Historical Trends: Party Affiliation. *Gallup.com*
26 <https://news.gallup.com/poll/15370/Party-Affiliation.aspx> (2007).
- 27 63. R Studio. Shiny. <https://shiny.rstudio.com/>.
- 28 64. Raifman, M. A. & Raifman, J. R. Disparities in the Population at Risk of Severe Illness
29 From COVID-19 by Race/Ethnicity and Income. *Am. J. Prev. Med.* 59, 137–139 (2020).
- 30 65. Health Equity Considerations & Racial & Ethnic Minority Groups. *Centers for Disease*
31 *Control and Prevention* [https://www.cdc.gov/coronavirus/2019-ncov/community/health-](https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html)
32 [equity/race-ethnicity.html](https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html) (2020).
- 33 66. How George Floyd Died, and What Happened Next - The New York Times.
34 <https://www.nytimes.com/article/george-floyd.html>.
- 35 67. FACT SHEET: Anti-Asian Prejudice March 2021 & Center for the Study of Hate &
36 Extremism. FACT SHEET: Anti-Asian Prejudice March 2021.
37 [https://www.csusb.edu/sites/default/files/FACT%20SHEET-%20Anti-](https://www.csusb.edu/sites/default/files/FACT%20SHEET-%20Anti-Asian%20Hate%202020%20rev%203.21.21.pdf)
38 [Asian%20Hate%202020%20rev%203.21.21.pdf](https://www.csusb.edu/sites/default/files/FACT%20SHEET-%20Anti-Asian%20Hate%202020%20rev%203.21.21.pdf) (2021).
- 39 68. Lawrence, C. *Masking Up: A COVID-19 Face-off between Anti-Mask Laws and*
40 *Mandatory Mask Orders for Black Americans*. <https://papers.ssrn.com/abstract=3695257> (2020)
41 doi:10.2139/ssrn.3695257.
- 42 69. Murray, J. S. Multiple Imputation: A Review of Practical and Theoretical Findings.
43 *ArXiv180104058 Stat* (2018).

70. Huque, M. H., Carlin, J. B., Simpson, J. A. & Lee, K. J. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Med. Res. Methodol.* 18, 168 (2018).
71. Liu, X. *Methods and applications of longitudinal data analysis*. (Elsevier, 2015).
72. Fine, K. L., Suk, H. W. & Grimm, K. J. An Examination of a Functional Mixed-Effects Modeling Approach to the Analysis of Longitudinal Data. *Multivar. Behav. Res.* 54, 475–491 (2019).
73. Zorowitz, S., Niv, Y. & Bennett, D. Inattentive responding can induce spurious associations between task behavior and symptom measures. (2021) doi:10.31234/osf.io/rynhk.
74. Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G. & Stoughton, J. W. Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Comput. Hum. Behav.* 76, 417–430 (2017).
75. McCredie, M. N. & Morey, L. C. Who Are the Turkers? A Characterization of MTurk Workers Using the Personality Assessment Inventory. *Assessment* 26, 759–766 (2019).
76. APS Global Collaboration on COVID-19. *Association for Psychological Science - APS* <https://www.psychologicalscience.org/covid-initiative>.

Supplemental Information

Participants

Figure SI-1: Sample Composition across Waves.

The figure summarizes the count of participants *invited* to complete each wave via Prolific, and the count of participants included in the *core*, *low completion rate*, and *excluded* sample that completed each wave.

Measures

Published Measures
Alcohol, Smoking, Substance Involvement Screening Test (ASSIST) ¹⁵ - a self-report, culturally neutral instrument that measures degree of personal risk related to use of the following substances: tobacco products, alcohol, cannabis, cocaine, amphetamine-type stimulants (ATS), sedatives and sleeping pills (benzodiazepines), hallucinogens, inhalants, opioids, and 'other' drugs.
Beck Depression Inventory – II (BDI-II) ¹⁶ - a 12-item self-report questionnaire that examines depressive symptomatology over the prior 2 weeks. Total scores indicate general level of depression: none, mild, moderate, severe.
Connor-Davidson Resilience Scale - 10 Item (CD-RISC-10) ¹⁷ - a self-report index of resilience based on Likert-scale ratings of coping responses during the prior month.
Disgust Scale - Revised (DS-R) ^{18,19} - a 27-item self-report questionnaire that measures an individual's sensitivity to disgust. It provides an index of core disgust (including food, animals, and body products), animal-reminder disgust (addressing death and organism violations) and contamination disgust (concerns about interpersonal transmission of materials).
External/Internal Motivation to Respond without Prejudice Scale (EMS / IMS): Black & Asian ²⁰ - a brief self-report questionnaire, 10 items, that assesses the degree to which an individual's attitudes and prejudice toward a designated racial group are motivated by internal values or external social pressures. The instrument was administered twice, once focused on blacks and once on Asians.
Epidemic – Pandemic Impacts Inventory (EPII) ⁶ - a 92-item self-report instrument used to identify how the current pandemic has impacted various aspects of one's life including: work / employment, education / training, home life, social activities, economic, emotional health and well-being, physical health, physical distancing / quarantine, infection history, and positive change.
Everyday Discrimination Scale ²¹ - a self-report measure of how frequently one experiences 9 types of discrimination in day-to-day life. For each type of discrimination that occurs "a few times a year" or more frequently, the individual indicates the main trait that is eliciting the discrimination.
Free Will Inventory (FWI) ²² - a self-report questionnaire in which individuals report how strongly they believe in free will, determinism, and dualism (15 items that produce a score for each of the 3 categories), as well as beliefs regarding relationships between free will, choice, responsibility, punishment, scientific prediction, and dualism (14 items).
Humanitarianism-Egalitarianism ²³ - a 10-item self-report questionnaire that assesses the extent to which an individual adheres to ideals of equality, social justice, and concern for the well-being of others, i.e., humanitarian-egalitarian values).
Iowa Personality Disorder Screen (IPDS) ^{24,25} - an 11-item screening questionnaire used to identify the presence of personality disorder in nonclinical samples. A score of 2 or more on items 1 and 3-8 provides optimal sensitivity and specificity for screening.
Life Events Checklist for DSM-5 (LEC-5) ²⁶ - a self-report questionnaire which asks about an individual's lifetime experiences of 17 types of trauma.
Major Experiences of Discrimination – Abbreviated ²⁷ - a self-report, lifetime frequency index of experiencing 6 types of major discrimination and the individual trait that elicited discrimination.
NEO Five-Factor Personality Inventory (NEO-FFI) ^{28,29} - a self-report questionnaire in which an individual rates how well each of the 60-items describes his/her behaviors, feelings, experiences and beliefs. In combination, these ratings indicate the degree to which the individual exhibits each of these 5 personality traits: extraversion, neuroticism, openness, conscientiousness, and agreeableness. Scoring algorithm and norms used for z-score conversion at http://www.uoregon.edu/~gsaucier/NEO-FFI%20subcomponent%20norms%20and%20scoring.htm .
NIH toolbox: Emotional Support & Loneliness ³⁰ - brief self-report indices of how often an individual had access to emotional support (8 items) and felt lonely or alone in the past month (5 items). These measures are part of the NIH Toolbox and outcomes are reported as T-scores normed to the general population (based on 2010 US Census).
Perceived Stress Scale (PSS) ³¹ - a ten-item self-response questionnaire that measures the extent to which a participant perceives personal life events in the previous month as stressful.

Positive and Negative Affect Scales (PANAS) ³² - a 20-item self-report measure designed to assess the subjects' current affective state.
Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) ^{26,33} - a self-report questionnaire which asks about the degree to which 20 posttraumatic stress disorder symptoms were experienced during the prior month.
Primary Care-PTSD Screen (PC-PTSD) ³⁴ - a brief self-report questionnaire in which participants are asked to report the presence of five post-traumatic stress disorder symptoms during the past month. Scores of 4 or 5 indicate probable PTSD.
Scale of Ethnocultural Empathy (SEE) ³⁵ - a 32-item, self-report measure of how strongly one empathizes with people of other racial and ethnic backgrounds. It provides 4 indices: Empathic Feeling and Expression, Empathic Perspective Taking, Acceptance of Cultural Differences, and Empathic Awareness.
Scales from the Social Psychological Measurements of COVID-19: City, State and Federal Government Response to Coronavirus Questionnaire ⁷ - self-report surveys of beliefs and opinions regarding how city, state, and federal government is, and should be, addressing the COVID-19 pandemic. Includes items related to personal restrictions, punishment for violating restrictions, emotional reactions, advocating for research, economic stimulus, and distrust of public information.
Social Dominance Orientation ³⁶ - a 16-item self-report index of preference for social inequality vs. hierarchy attenuation.
Social Network Index (SNI) Extended (current & 2019 version) ³⁷ - a self-report questionnaire used to quantify the extent of one's social connections during a specific timeframe. Outcome variables include: a) Number of High-Contact Roles / Network Diversity (number of social roles in which the respondent has contact with one person or more at least once every 2 weeks; maximum is 12 including spouse, parent, child, child-in-law, close relative, close friend, church/temple member, student, employee, neighbor, volunteer, and group member), b) Number of People in Social Network (measures the total number of people which whom respondent maintains contact at least once every 2 weeks -- reflecting overall network size), and c) Number of Embedded Networks (measures the number of different groups these contacts belong to, reflecting network complexity; maximum is 8., including family, friends, church/temple, school, work, neighbors, volunteering, and groups).
State Trait Anxiety Inventory (STAI, state & trait) ³⁸ - a self-report questionnaire that differentiates between the temporary condition of "state anxiety" and the more general and long-standing quality of "trait anxiety." State anxiety is characterized by feelings of apprehension, tension, nervousness, and worry. Elevated trait anxiety is characterized by an overall pattern of increased arousal response to physical danger and psychological stress. Trait anxiety is typically high in people with depression or other psychiatric conditions.
Very Short Authoritarianism Scale (VSA) ³⁹ - a six-item self-report questionnaire that assesses right-wing authoritarianism. VSA scores have been shown to predict other factors including nationalism, ethnocentrism, political orientation, political party/candidate support, attitudes towards ingroups or outgroups and anti-minority bias. The measure addresses three facets of authoritarianism: authoritarian submission, authoritarian aggression and conventionalism.
Measures Created by COVID-Dynamic Team
COVID & Protest Questionnaires
COVID Disgust & Fear - brief questionnaires regarding strength of current disgust (9 items) and fear (7 items) responses when faced with various COVID-related events.
COVID Expectation Bias - instrument that aims to identify bias in predicting 3 COVID-related outcomes (diagnosis of COVID-19, severe illness from COVID-19 and severe financial distress). For each outcome, the participant is asked to rate how likely this is for a randomly selected group. Groups are designated according to race/ethnicity, socioeconomic status, and other factors (elderly, immigrants, healthcare workers, etc ...).
COVID Experiences - questionnaire used to identify COVID-related experiences (e.g., diagnosis, illness, treatment), impact of pandemic on daily activities, engagement with news media, and judgment of leadership responses to the pandemic.
COVID Explicit Norms - participants rate the importance of 3 specific COVID=19 prevention tactics (social isolation, social distancing and use of personal protective equipment).

COVID Mask Questions - self-reported description of mask-usage before and during the pandemic, as well as beliefs regarding mask-usage to address the pandemic.
COVID Public Policy Knowledge - a 12-item questionnaire that assesses the participants' knowledge of current state-enacted COVID-related policies.
COVID Theories - very brief questionnaire about beliefs regarding the existence and origins of COVID-19.
COVID Vaccination - survey created to assess experiences, attitudes, and beliefs related to COVID-19 vaccines.
COVID Vignettes – Implicit Norms – implicit assessment of how participants attribute responsibility / blame regarding the spread of COVID-19. Participants read one out of 9 brief vignettes describing how the behavior of two people may contribute to transmitting COVID-19 to a pharmacist. then answer questions regarding responsibility of the people in the vignette.
Post-Election 2020 - survey created to characterize voting behavior in November 2020 election, as well as beliefs and attitudes regarding voting and the election outcome.
Protest Surveys (Race) - surveys created to assess experiences, attitudes, and beliefs related to the racial protests that occurred in June 2020. Peaceful protests, looting, and violence are assessed independently.
Protest Survey (Election) - survey created to assess experiences, attitudes, and beliefs related to the election-related protests that occurred in January 2021.
<i>Experimental Measures</i>
Personal background at enrollment - demographic questionnaire administered at enrollment.
Personal background updates - questions regarding relationship status, work changes, finances, political activities / affiliations, psychological treatment, religion, and current stressors.
Consensus Task - questionnaire that assesses individual values, beliefs, and preferences across a wide range of topics including COVID-19, social issues, political policies, and hobbies, among others.
Emotion Space - self-report questionnaire regarding current emotional experience.
Egocentric-Allocentric Task - perspective tasking task in which participant responds to spatial questions regarding a photo of a political leader.
Monthly Substance Use Questionnaire - survey created to assess substance-use frequency during the prior month.
Race Thermometer - participants use a slider to indicate how they feel toward 6 racial/ethnic groups: Black/African Americans, Asian Americans, Hispanic/Latino Americans, White/Caucasian Americans, Chinese People, and European People. The scale ranges from 0 (Very Cold or Unfavorable) to 100 (Very Warm or Favorable).
Responsibility Vignettes - implicit assessment of how participants attribute responsibility / blame in a common, low-risk social interaction. Participants read one out of 3 brief vignettes describing how the behavior of two people in a reception area inconveniences an important client that arrives after them, then answer questions regarding responsibility of the people in the vignette.
Restriction Severity - scale that assesses how severe an individual finds various restrictions that may be enforced or encouraged during the pandemic.
Self-Continuity ⁴⁰ - this measure asks individuals to rate changes in self-perception, behavior, and emotional experiences resulting from events during 2020. Participants retrospectively rated changes in these 3 domains as of March, June and September 2020.
Verbal Self-Fluency ^{41,42} - in this measure, individuals are asked to provide up to 30 statements that “describe you as a unique individual.” These free text responses are evaluated for
<i>Tasks</i>
Pro-social Implicit Association Task (IAT) ^{8,44} - a version of the standard implicit association test ¹⁵ that assesses implicit associations between self-identity (ME/THEY words) and pro-sociality (SERVICE/PROFIT words) ¹⁶

Multi-category Brief Implicit Association Task (BIAT) ^{9,43} - a brief form of the implicit association test ¹⁷ that assesses implicit associations in 3 subtests (administered in random order): 1) self-identity (ME/THEY words) and pro-sociality (SERVICE/PROFIT words); 2) evaluative (GOOD/BAD words) and White/Black People (WHITE/BLACK faces); and 3) evaluative (GOOD/BAD words) and White/Asian People (WHITE/ASIAN faces)
Multi-race Affective Misattribution Procedure (AMP) ¹⁰ - this adaptation of the Multi-race Affect Misattribution Procedure ¹⁸ assesses evaluative (PLEASANT/UNPLEASANT) implicit associations with different races (Asian, Black, and White).
COVID-related Trustworthiness Rating Task ¹¹ - in this adaptation of the Trustworthiness Rating Task ¹⁸ , participants are shown pictures of Asian/Black/White/other race faces and are asked to rate "How much do you trust this person to act responsibly with respect to the COVID-19 pandemic?" on a scale from 1 (not-at-all) to 9 (completely).
Altruism Task ¹² - in this adaptation of the Altruism Task, participants make monetary decisions to distribute money equally or unequally (generous/selfish) amongst themselves and partners who vary with regard to age, race, occupation, and political identity.
Public Goods Game ^{13,14} - a standard task from behavioral economics that assesses cooperation and self-interest and group cohesion in an anonymous group fund-sharing scenario. Instructions for the public goods game were adapted from Wills et al.

Table SI-1: COVID-DYNAMIC Test Battery.

[COVID-Dynamic Data Dictionary](#) (download link)

Table SI-2: Complete Data Dictionary. Data dictionary of all survey items including possible responses for each item and scores generated for each measure. The dictionary is too large for display here. A continuously updated data dictionary for this study, an overview of all measures, and pdf-documents of tests as administered are presented on our website (<https://covidynamic.caltech.edu/research-projects>).

Table SI-3: Summary of 376 COVID-Related Psychological Studies. The table summarizing details of all 376 studies is too large for display here. See <https://covidynamic.caltech.edu/resources/other-covid-studies> for a link to the complete table.

Sample Loss

To characterize the factors that contributed to sample loss we compared three groups: (1) participants *excluded* due to poor data quality, (2) participants with a *low-completion rate*, and (3) the *core* sample of participants (see Methods section *Analysis - Sample Loss* for details). This section illustrates and summarizes the difference between the excluded and the core sample, and the low-completion-rate and the core sample, respectively.

Figure SI-2: Demographic Comparison of Excluded and Low-Completion Rate Samples to the Core Sample. Two groups of participants were not included in the analyses of the Core Sample: individuals excluded (*excluded*) due to data quality indices in Waves 1-5 (NOTE: after Wave 5 no more participants were excluded), and participants that did not return for repeated testing (*low-completion rate*, <8 out of 16 waves). To test for sample selection biases in age and sex (A, B), political affiliation (C, D), highest education (E, F), weekly income (G, H), and race and ethnicity (I, J), we compared the excluded (left column) and low-completion rate (right column) groups to the core sample. Histograms show the Euclidean distance between bootstrapped excluded and low completion samples and the core sample. Black vertical dotted lines mark the 2.5th and 97.5th percentile of the differences between the bootstrapped samples and the core sample, respectively. Vertical red lines mark true Euclidean

distances between the excluded or low-completion sample and the core sample, respectively. Butterfly charts superimpose age and sex in the core sample with the excluded (A) and low-completion rate (B) samples. Nested pie charts show all other demographic variables (C-J) with the core sample's proportions presented in the center and the comparison group (excluded on left, low-completion-rate on right) in the outer circles. For readability, labels are only displayed for the excluded and low-completion rate samples and values <1% are dropped from the pie charts.

A

age female	prop. core	prop. excl.	prop. low compl.	abs. diff. core - excl.	abs. diff. core - low compl.	z-value (core/excl.)	p-value (core/excl.)	z-value (core/low compl.)	p-value (core/low compl.)
18-19	0.033	0.108	0.103	0.075	0.069	-3.507	>0.001*	-3.861	>0.001*
20-24	0.148	0.162	0.154	0.015	0.006	-0.395	0.693	-0.213	0.831
25-29	0.106	0.198	0.200	0.092	0.094	-2.739	0.006	-3.398	0.001*
30-34	0.116	0.09	0.149	0.026	0.033	0.798	0.425	-1.202	0.23
35-39	0.088	0.072	0.056	0.016	0.031	0.548	0.584	1.407	0.159
40-44	0.121	0.09	0.123	0.031	0.002	0.936	0.349	-0.075	0.94
45-49	0.109	0.054	0.077	0.055	0.033	1.781	0.075	1.308	0.191
50-54	0.075	0.108	0.062	0.033	0.013	-1.196	0.232	0.617	0.537
55-59	0.08	0.054	0.041	0.026	0.039	0.936	0.35	1.833	0.067
60-64	0.071	0.027	0.015	0.044	0.056	1.746	0.081	2.913	0.004
65-69	0.045	0.027	0.01	0.018	0.035	0.857	0.392	2.239	0.025
70-74	0.008	0.009	0.01	0.001	0.002	-0.076	0.939	-0.256	0.798
75-79	NA	NA	NA	NA	NA	NA	NA	NA	NA
80-84	NA	NA	NA	NA	NA	NA	NA	NA	NA
age male	prop. core	prop. excl.	prop. low compl.	abs. diff. core - excl.	abs. diff. core - low compl.	z-value (core/excl.)	p-value (core/excl.)	z-value (core/low compl.)	p-value (core/low compl.)
18-19	0.031	0.164	0.098	0.132	0.067	-5.692	>0.001*	-3.795	>0.001*
20-24	0.111	0.164	0.235	0.052	0.124	-1.542	0.123	-4.326	>0.001*
25-29	0.185	0.191	0.211	0.006	0.026	-0.154	0.877	-0.814	0.416
30-34	0.145	0.173	0.123	0.028	0.022	-0.759	0.448	0.782	0.434
35-39	0.131	0.082	0.093	0.049	0.038	1.43	0.153	1.412	0.158
40-44	0.094	0.045	0.078	0.049	0.016	1.664	0.096	0.671	0.502

COVID-Dynamic

45-49	0.084	0.055	0.059	0.029	0.025	1.036	0.3	1.14	0.254
50-54	0.064	0.073	0.044	0.008	0.02	-0.32	0.749	1.058	0.29
55-59	0.064	0.027	0.029	0.037	0.035	1.523	0.128	1.882	0.06
60-64	0.038	0.018	0.02	0.02	0.019	1.052	0.293	1.278	0.201
65-69	0.033	NA	0.005	0.033	0.028	NA	NA	NA	NA
70-74	0.014	0.009	NA	0.005	0.014	0.409	0.683	1.695	0.09
75-79	0.005	NA	NA	NA	NA	NA	NA	NA	NA
80-84	NA	NA	0.005	NA	NA	NA	NA	NA	NA

*p Bonferroni = 0.003

B

highest education level	prop. core	prop. excl..	prop. low compl.	abs. diff. core - excl.	abs. diff. core - low compl.	z-value (core/excl.)	p-value (core/excl.)	z-value (core/low compl.)	p-value (core/low compl.)
PhD	0.015	0.018	0.005	0.003	0.01	-0.308	0.758	1.585	0.113
master's degree	0.138	0.086	0.088	0.052	0.05	2.099	0.036	2.606	0.009
some graduate education	0.033	0.05	0.045	0.017	0.012	-1.222	0.222	-1.107	0.268
bachelor's degree	0.332	0.339	0.291	0.007	0.041	-0.207	0.836	1.533	0.125
professional degree	0.02	0.023	0.018	0.002	0.003	-0.214	0.831	0.354	0.723
associate degree	0.109	0.09	0.123	0.018	0.014	0.809	0.418	-0.768	0.442
some college	0.232	0.24	0.323	0.008	0.091	-0.254	0.8	-3.618	>0.001*
high school	0.111	0.136	0.095	0.024	0.016	-1.045	0.296	0.896	0.37
some high school	0.008	0.014	0.01	0.005	0.002	-0.722	0.47	-0.281	0.779
other	0.001	0.005	0.003	0.004	0.002	-1.326	0.185	-0.803	0.422

*p Bonferroni = 0.005

C

weekly income	prop. core	prop. excl..	prop. low compl.	abs. diff. core - excl.	abs. diff. core - low compl.	z-value (core/excl.)	p-value (core/excl.)	z-value (core/low compl.)	p-value (core/low compl.)
more than \$3000	0.034	0.027	0.02	0.007	0.014	0.523	0.601	1.4	0.162
\$1500 - \$2999	0.095	0.054	0.083	0.041	0.012	1.96	0.05	0.744	0.457
\$1000 - \$1499	0.128	0.122	0.09	0.006	0.038	0.25	0.802	2.032	0.042
\$500 - \$999	0.169	0.176	0.185	0.007	0.016	-0.268	0.788	-0.748	0.455
\$250 - \$499	0.15	0.154	0.083	0.004	0.067	-0.165	0.869	3.401	>0.001*

less than \$249	0.393	0.434	0.514	0.041	0.12	-1.142	0.253	-4.206	>0.001*
don't know	0.031	0.032	0.025	0.001	0.006	-0.086	0.932	0.566	0.571

*p Bonferroni = 0.007

D

race/ethnicity	prop. core	prop. excl..	prop. low compl.	abs.. diff. core - excl.	abs. diff. core - low compl.	z-value (core/excl.)	p-value (core/excl.)	z-value (core/low compl.)	p-value (core/low compl.)
American Indian/Alaska Native	0.001	0.014	0.005	0.013	0.004	-3.25	>0.001*	-1.649	0.099
Asian	0.101	0.167	0.088	0.066	0.013	-2.873	0.004*	0.778	0.436
Black or African American	0.071	0.054	0.083	0.017	0.011	0.921	0.357	-0.747	0.455
Hispanic/Latinx	0.094	0.072	0.11	0.022	0.016	1.04	0.298	-0.926	0.355
Multiracial	0.031	0.041	0.045	0.009	0.014	-0.71	0.478	-1.286	0.198
Native Hawaiian or Other Pacific Islander	0.001	0.005	0.003	0.004	0.002	-1.326	0.185	-0.803	0.422
other	0.004	0.014	0.003	0.009	0.002	-1.687	0.092	0.488	0.625
prefer not to disclose	0.003	0.018	0.003	0.015	0.001	-2.658	0.008	0.274	0.784
White	0.692	0.615	0.662	0.077	0.031	2.254	0.024	1.144	0.253

*p Bonferroni = 0.0006

Table SI-4: Proportion of Participants in (A) Age, (B) Education, (C) Income, and (D) Race and Ethnicity Brackets compared between Core and Excluded, and Core and Low-Completion Sample.

For all demographic features that differed significantly between the core and low completion or the core and excluded sample, respectively (see Figure SI-2), we compared the proportion of participants in each feature bracket between the groups using post-hoc two-proportion z-tests (critical p-value Bonferroni corrected for each feature). Each table lists the within group (core, excluded, low completion) distribution of participants across the respective feature brackets (proportion of sample per bracket), absolute proportion differences per bracket between the core and excluded, and core and low-completion group, and lastly the z- and p-scores for two-proportion z-tests comparing the core sample against excluded and low completion sample. Significant differences between samples are highlighted in bold. NOTE: prop. = proportion; excl. = excluded; compl. = completion; abs. diff. = absolute difference

Figure SI-3: Psychological Characteristic Comparison of Excluded and Low-Completion Rate Samples to Core Sample. The core sample was compared to the excluded and the low-completion rate samples using seven psychological variables: trait anxiety (STAI trait; top left), depression (BDI-II; top right), and personality (NEO-FFI subscales N=neuroticism, E=extraversion, O=openness, A=agreeableness, C=conscientiousness; bottom). Violin and strip plots show mean scores bootstrapped from the core sample for the low-completion rate (left, blue) and the excluded sample (right, green). NOTE: the only difference in bootstrapped distributions for the low completion rate and the excluded sample is the size of the bootstrapped samples. Black horizontal lines mark the 2.5th and 97.5th percentile of the bootstrapped distributions, respectively. Orange circles mark the true mean scores in the excluded and low-completion rate samples. Compared to the core sample, excluded participants did not differ from the core sample on trait anxiety scores or depression symptom

scores but differed in four personality subscales: lower on agreeableness, conscientiousness and openness to experience; higher on neuroticism. The low-completion-rate sample showed higher trait anxiety, depression symptoms, and neuroticism scores on the NEO and were lower on the agreeableness and conscientiousness personality scores than the core sample. Importantly, despite differing from the core sample on select factors, average personality scores for both comparison groups and the core sample were within the normal range (1 SD from the mean of a community sample on the NEO^{37,38}) and both STAI trait⁴⁷ and BDI-II²⁵ scores were below the cut-off for clinical significance. Some of the differences from the core group are not surprising (e.g., that those scoring low on conscientiousness and agreeableness were more likely to be excluded due to poor data quality or low completion rate).

	excl. mean	excl. bootstr. mean	excl. bootstr. 2.5th perc.	excl. bootstr. 97.5th perc.	low compl. mean	low compl. bootstr. mean	low compl. bootstr. 2.5th perc.	low compl. bootstr. 97.5th perc.
STAI trait T-score	60.353	59.404	57.718	61.091	61.198*	59.383	57.703	61.078
BDI-II	12.836	12.712	11.439	14.014	14.865*	12.723	11.452	14.018
NEO-O z-score	-0.455*	-0.229	-0.35	-0.111	-0.224	-0.228	-0.347	-0.108
NEO-C z-score	-0.586*	-0.238	-0.399	-0.078	-0.737*	-0.238	-0.399	-0.078
NEO-E z-score	-0.739	-0.871	-1.046	-0.697	-0.696	-0.871	-1.047	-0.696
NEO-A z-score	-0.585*	-0.075	-0.236	0.087	-0.38*	-0.075	-0.23	0.082
NEO-N z-score	0.576*	0.314	0.143	0.485	0.725*	0.315	0.141	0.489

*p=0.05

Table SI-5: Psychological Mean Test Scores and Bootstrapped Means and Percentiles from the Excluded and Low-Completion Rate Sample Groups.

Bold mean values are significantly different (*p < .05; two sided) from the bootstrapped null distribution. NOTE: excl. = excluded; bootstr. = bootstrap; perc. = percentile; compl. = completion; BDI-II = Beck Depression Inventory – II; STAI = State Trait Anxiety Inventory; NEO = NEO-Five Factor Inventory.

Data Raking

A common post-stratification technique to account for biases in sample demographics is data-raking. Here, we provide raking weight for sex, age, race, and ethnicity to make the core sample more comparable to the US-population. Summary weights for each demographic variable are presented in Tables SI-7 to Table SI-10.

	target prop.	unweighted N	unweighted prop.	wtd N	wtd prop.	change in prop.	resid. disc.	orig. disc.
female	0.508	603	0.512	623.027	0.529	0.017	-0.021	-0.004
male	0.492	574	0.488	553.973	0.471	-0.017	0.021	0.004
sum	1	1177	1	1177	1	0.034	0.043	0.009

Table SI-6: Summary Weights for Data Raking by Sex. NOTE: prop. = proportion; wtd = weighted; resid. disc. = residual discrepancy after weighting; orig. disc. = original discrepancy before weighting

age range	target prop.	unweighted N	unweighted prop.	wtd N	wtd prop.	change in prop.	resid. disc.	orig. disc.
18-24	0.093	191	0.162	109.011	0.093	-0.07	0	-0.07
25-29	0.095	170	0.144	112.17	0.095	-0.049	0	-0.049
30-34	0.09	119	0.101	105.851	0.09	-0.011	0	-0.011
35-39	0.086	34	0.029	101.111	0.086	0.057	0	0.057
40-44	0.083	128	0.109	97.952	0.083	-0.026	0	-0.026
45-49	0.087	241	0.205	102.691	0.087	-0.118	0	-0.118
50-54	0.09	82	0.07	105.851	0.09	0.02	0	0.02
55-59	0.09	85	0.072	105.851	0.09	0.018	0	0.018
60-100	0.286	127	0.108	336.511	0.286	0.178	0	0.178
sum	1	1177	1	1177	1	0.546	0	0.546

Table SI-7: Summary Weights for Data Raking by Age Range. NOTE: prop. = proportion; wtd = weighted; resid. disc. = residual discrepancy after weighting; orig. disc. = original discrepancy before weighting

	target prop.	unweighted N	unweighted prop.	wtd N	wtd Prop.	change in prop.	resid. disc.	orig. disc.
American Indian/Alaska Native, Native Hawaiian or Other Pacific Islander, Multiracial, other, prefer not to disclose	0.091	85	0.072	107.214	0.091	0.019	0	0.019
Asian	0.054	121	0.103	63.622	0.054	-0.049	0	-0.049
Black	0.127	85	0.072	149.629	0.127	0.055	0	0.055
White	0.728	886	0.753	856.536	0.728	-0.025	0	-0.025
sum	1	1177	1	1177	1	0.148	0	0.148

Table SI-8: Summary Weights for Data Raking by Race. NOTE: prop. = proportion; wtd = weighted; resid. disc = residual discrepancy after weighting; orig. disc. = original discrepancy before weighting

	target prop.	unweighted N	unweighted prop.	wtd N	wtd prop.	change in prop.	resid. disc.	orig. disc.
Hispanic or Latinx	0.178	114	0.097	209.506	0.178	0.081	0	0.081
Not Hispanic or Latinx	0.822	1063	0.903	967.494	0.822	-0.081	0	-0.081
sum	1	1177	1	1177	1	0.162	0	0.162

Table SI-9: Summary Weights for Data Raking by Ethnicity. NOTE: prop. = proportion; wtd = weighted; resid. disc. = residual discrepancy after weighting; orig. disc. = original discrepancy before weighting

Description of Conte Sample

Methods

Participants

Currently active Caltech Conte Center participants ($n = 105$) were notified of this study on April 14 via email and 62 expressed interest in participating; of those, 58 had completed at least one testing wave by the end of June 2020. Conte Center participants were 18-50 years old, fluent in English, and had Full Scale Intelligence Quotient (FSIQ) ≥ 90 . Exclusion criteria at Conte Center enrollment included a first-degree relative with schizophrenia or ASD, currently taking psychotropic medication, uncorrected vision or hearing impairment, and moderate-severe depression on the Beck Depression Inventory. Additional exclusionary criteria included history of any of the following: premature birth, epilepsy, major medical condition, metabolic disorder, chemotherapy or radiation, brain surgery, head injury, eating disorder, neurological condition, psychosis, bipolar disorder, autism, suicide attempt, substance dependence or abuse, alcoholism, color blindness or strabismus. In addition to the data collected in this study, Conte participants have been tested in a large variety of behavioral and fMRI studies. During the seven years prior to this study, all Conte Center participants completed a battery of psychological questionnaires (see Table SI-10 for full list). Previously collected Conte Data is available on request.

For the COVID-Dynamic study, the Conte Center sample completed 13 waves of testing (labeled A – M) which correspond to Waves 3 and 5-16 of the Prolific data. Unlike the Prolific sample, Conte participants completed only one wave during April (on the 15th). Table SI-10 provides a full list of measures administered each wave. The data reported here are from a core group of 45 participants who completed at least 50% (7 out of 13) of these waves. This Core Conte Sample is 50% female, with mean age = 33.7, and age range = 25-48. Details about demographics of the Core Conte Sample are presented in Table SI-11, along with demographics of the group of participants with low-completion rates ($n = 13$). All procedures were reviewed and approved by the Internal Review Board of the California Institute of Technology (protocol 16-0692B)

1

	WAVES	A	B	C	D	E	F	G	H	I	J	K	L	M
	Date	April 15	May 9	May 23	June 6	June 27	July 18	Aug 8	Aug 29	Sep 19	Oct 10	Nov 7	Dec 5	Jan 23
	Participant Count	45	49	49	49	45	49	43	33	35	43	38	42	42
Published Measures	Alcohol, Smoking, Substance Involvement Screening Test ¹⁵													
	Beck Depression Inventory - II ¹⁶													
	Connor-Davidson Resilience Scale ¹⁷													
	Disgust Scale – Revised ^{18,19}													
	External/Internal Motivation to Respond without Prejudice: Black & Asian ²⁰													
	Epidemic – Pandemic Impacts Inventory ⁶													
	Everyday Discrimination Scale ²¹													
	Free Will Inventory ²²													
	Humanitarianism-Egalitarianism ²³													
	Iowa Personality Disorder Screen ^{24,25}													
	Life Events Checklist for DSM-5 ²⁶													
	Monthly Substance Use Questionnaire ¹⁵													
	Major Experiences of Discrimination - Abbreviated ²⁷													
	NEO-FFI ^{28,29}													
	NIH toolbox: Emotional Support & Loneliness ³⁰													
	Perceived Stress Scale ³¹													
	Positive and Negative Affect Scales ³²													
	Posttraumatic Stress Disorder Checklist for DSM-5 ^{26,33}													
	Primary Care-PTSD Screen ³⁴													
	Scale of Ethnocultural Empathy ³⁵													
	Scales from the Social Psychological Measurements of COVID-19 ⁷													
	Social Dominance Orientation ³⁶													
	SNI Extended (2019) ³⁷													
	SNI Extended (current) ³⁷													
	State Trait Anxiety Inventory (state) ³⁸													
	State Trait Anxiety Inventory (trait) ³⁸													
	Very Short Authoritarianism Scale ³⁹													
COVID & Protest	COVID Disgust & Fear													
	COVID Expectation Bias													
	COVID Experiences													
	COVID Explicit Norms													
	COVID Mask Questions													
	COVID Public Policy Knowledge													
	COVID Theories													
	COVID Vaccinations													

[illegible]

Table SI-10: COVID-Dynamic Test Battery for Conte Waves A - M

WA - WM = Conte testing Waves; 2 weeks = 2 dark lines between columns; 3 weeks = 3 dark lines between column; 4 weeks = wide dark line between columns; 7 weeks = wavy line between columns; Measurement domains: daily life impact of COVID pandemic, implicit attitude task, explicit attitudes and beliefs, emotion and mental health, social behaviors, decision making task, physical health, and participant demographics and background

sex	core	low comp. rate
male	0.51	0.46
female	0.49	0.54

age	core	low comp. rate
18- 24	0.02	0
25- 29	0.044	0.06
30-34	0.08	0.08
35-39	0.043	0.02
40-44	0.03	0.02
45-49	0.04	0

race	core	low comp. rate
American Indian/Alaska Native	0.02	0.08
Asian	0.18	0
Black or African American	0.02	0.23
Multiracial	0.13	0.15
White	0.6	0.38
other	0.02	0.08
prefer not to disclose	0.02	0

ethnicity	core	low comp. rate
Hispanic or Latinx	0.73	0.54
Not Hispanic or Latinx	0.27	0.46

political party	core	low comp. rate
Democrat	0.6	0.38
Republican	0.04	0.08
Independent	0.27	0.46
Other	0.09	0

political leaning	core	low comp. rate
strongly liberal	0.29	0.23
moderately liberal	0.29	0.38
slightly liberal	0.09	0.08
neutral	0.18	0.23
slightly conservative	0.07	0
moderately conservative	0.02	0
strongly conservative	0.02	0

Table SI-11: Demographics of Participants Recruited through the Conte Center. No participants in the Conte sample were excluded due to poor data quality. NOTE: low comp. rate = low-completion rate group.

Results

To summarize overall data quality, we calculated the number of participants in the Conte Sample that passed increasing percentages of our 15 data quality indices per wave (Table SI-12). Overall, in all waves the majority of participants (mean=65%; min=51%) passed all data quality metrics. Pass rates increased as the percentage threshold decreased, plateauing between 80% and 70% thresholds. Except for Wave K, all participants passed at least 50% of all data quality metrics across all waves. Participant- and wave-wise scores for all quality measures, as well as flags for missing task data, are provided in the data release.

wave	pass=1	pass>0.95	pass>0.9	pass>0.85	pass>0.8	pass>0.75	pass>0.7	pass>0.65	pass>0.6	pass>0.55	pass>0.5	total N
A	53.33	53.33	80	82.22	91.11	95.56	97.78	100	100	100	100	45
B	71.43	71.43	91.84	91.84	97.96	100	100	100	100	100	100	49
C	75.51	75.51	93.88	93.88	97.96	97.96	97.96	97.96	100	100	100	49
D	57.14	57.14	81.63	83.67	89.80	91.84	91.84	93.88	97.96	97.96	100	49
E	77.78	77.78	93.33	93.33	95.56	95.56	100	100	100	100	100	45
F	79.59	79.59	93.88	93.88	97.96	97.96	100	100.00	100	100	100	49
G	65.12	65.12	83.72	86.05	97.67	97.67	97.67	97.67	100	100	100	43
H	60.61	60.61	72.73	78.79	90.91	90.91	96.97	96.97	100	100	100	33
I	57.14	57.14	71.43	74.29	94.29	94.29	97.14	97.14	100	100	100	35

J	69.77	69.77	83.72	88.37	100	100	100	100	100	100	100	43
K	63.16	63.16	78.95	84.21	92.11	92.11	94.74	94.74	94.74	97.37	97.37	38
L	61.90	61.90	80.95	83.33	97.62	97.62	100	100	100	100	100	42
M	61.90	61.90	76.19	80.95	97.62	97.62	100	100	100	100	100	42

Table SI-12: Core Conte Sample Data Quality Metrics for Each Wave. For each wave the table shows the percentage of Conte participants that passed different percentages (100% to 50% in decrements of 5%) of our 15 data quality indices. Apart from Wave K, all Conte participants passed >50% of all quality measures in all waves.

To illustrate the richness of these data, including variation over time and across individuals and similarity to the Prolific sample, Figure SI-4 depicts mean response for the core sample of Conte participants on the measures used in Figure 4: one representative measure (i.e. item or summary score) from each measurement domain assessed in the study other than demographics (i.e. daily life impact-A, emotion and mental health-B, decision making tasks-C, implicit attitudes-D, attitudes and beliefs-E, social behaviors-F, and physical health-G; note: not all measures were collected every wave). Collectively, the data in Figure SI-4 demonstrate that there exists substantial meaningful variation in the dataset both across and within waves and participants. In addition, each measure has a clear pattern of evolution over the time period displayed, which is similar to the pattern seen in the Prolific core sample (Figure 4). As this analysis is meant only to qualitatively showcase a subset of the data, we refrain from further reporting on group means and individual trajectories here.

Figure SI-4: Group Temporal Trends in the Conte Sample from a Sample Variable in Each Domain. The same seven measures as in Figure 4 are depicted here. Orange triangles indicate weeks in which data were collected. Small triangles indicate that in the respective wave we collected data from the Prolific sample only. Large triangles indicate data collection in the Prolific and the Conte sample. Prolific waves are designated 1-16 and Conte are A-M. Line plots indicate the mean response across subjects in each wave, with 95% confidence intervals denoted by the shaded region.

Discussion

Here, we include concurrently collected data from the 13 waves (A-M; April 15 to January 23, 2021) of a smaller (N=45), but extremely well-characterized, group of participants from the Caltech *Conte* Center. During the seven years prior to this study, these participants completed a battery of psychological questionnaires; many experimental tasks assessing social processing and decision-making; as well as structural, resting, and task-based MRI data. Many of the psychological measures previously acquired overlap with those collected in the current study, including personality and social measures (SNI), trait emotional measures (STAI-trait, BDI-II) and state emotional measures (PSS, STAI-state, PANAS). As such, these data can serve as valuable pre-COVID-19 baseline measures to which we can compare the distinct ways in which these individuals have responded to the pandemic and other events of 2020. In addition, the availability of brain imaging data for the individuals in this cohort adds another dimension to our data set, enabling us to link neurobiological mechanisms to 2020-related psychological outcomes.