Appendix for 'The COVID-19 Pandemic Revives Traditional Values in Japan'

Plamen Akaliyski, Naoko Taniguchi, Joonha Park, Stefan Gehrig, Raül Tormos

Contents

A Additional methods

Here, we describe in more detail the statistical methods used in the analyses that involve both World Values Survey wave 7 (WVS7) and Values in Crisis survey wave 1 and 2 (VIC1 and VIC2) in order to assess Hypotheses 1 and 2 on national-level change and prefecture-level variation in change, respectively. In particular, we describe in more detail the weighting procedure and give an illustrative example.

A.1 Inverse probability of treatment weighting

Inverse probability of treatment weighting (IPTW) adjusts for observed sample differences by upweighting (downweighting) those observations in the VIC1 sample that are similar (dissimilar) in their characteristics to the WVS7 sample. The procedure requires to estimate a propensity score that expresses the probability of being surveyed in WVS7 rather than VIC1, given individuals' characteristics. Propensity scores were estimated via logistic regression including nine characteristics: age, number of children, household size (all continuous), gender, marital status, education, prefecture (there are 47 prefectures in Japan, nested within seven regions), town size and prior religious services attendance (all categorical; find propensity model results in Appendix C). Post-stratification weights (explained below) were included when estimating in the propensity model.

Propensity scores were then used to construct weights. The propensity weight for all respondents in WVS7 was set to 1, while for respondents in VIC1 it was calculated as $w_i = \pi_i/(1 - \pi_i)$, where π_i is the propensity score (i.e., the conditional probability of being surveyed of WVS7, estimated in the logistic regression model) for respondent i. This weight definition implies that the interest lies in the average effect of the pandemic among the population surveyed in WVS7. Technically, this corresponds to a weighting scheme to estimate an "average treatment effect on the treated", where treatment here means being surveyed in WVS7 (Austin and Stuart 2015). In that sense, the distribution of covariates in the WVS7 survey sample serves as the "gold standard" and the VIC1 sample is re-weighted to come close to this standard. This choice was made as we expect that WVS7, with its design and implementation building on many years of experience and planning in national surveying, exhibits less sample bias.

Propensity scores show a large region of common support between WVS7 and VIC1 (main text, Fig. 3a) and weighting is strongly improving the balance between samples (main text, Fig. 3b and Appendix, Table S1). For valid IPTW, balance diagnostics are the essential yardstick (Austin and Stuart 2015) and we judge the standardized mean differences between survey samples after re-weighting (main text, Fig. 3b) to indicate an appropriate propensity model. Importantly, weighting needs to be understood as only a pre-processing step and further adjustment is achieved by including the same characteristics used for weighting (with one exception: region instead of prefecture dummies are used in the outcome regression due to the large number of prefectures) as predictors in regression models which we use to infer the pandemic's effects, with age included as both linear and squared term. Combining IPTW with adjustment variables in the outcome model is an instance of a "doubly robust" approach to causal inference, since one of the models is allowed to be misspecified (e.g., Hernán and Robins 2020).

A.2 Post-stratification

The goal of the methods described above is to create comparability between the WVS7 and VIC samples with respect to the most important (observed) confounding factors – an aspect of internal validity of our study. External validity, in contrast, refers to the ability to draw conclusions about the Japanese population as a whole based on the statistical results. In principle, the chances to correctly generalize from the present data are good: all data come from nationally representative survey efforts and we calculate the propensity weights in a way which fixes the probably most representative survey (WVS7) as the weighting target. Yet, some of the propensity-weighted sample frequencies (e.g., frequency of female respondents) might still deviate from the Japanese population frequencies. Therefore, we also created post-stratification weights for all unique strata defined by gender (male, female), age group (18-35 years, 36-50 years, older than 50 years) and prefecture (47 prefectures) relying on the freely accessible 2020 Japanese census (Statistics Bureau of Japan 2021) for the true population frequencies in the adult population ($\geq$ 18 years). Post-stratification is standard procedure in survey statistics to de-bias estimates from non-representative samples after the completion of data collection (Levy and Lemeshow 2008, chap. 15.6).

The resulting post-stratification weights were used already during estimation of the propensity score model (Ridgeway et al. 2015). We then combined the resulting propensity weights with the post-stratification weights via multiplication in order to include them as weights in the outcome regressions. This procedure is recommended to simultaneously account for aspects of confounding and survey design in observational studies (Ridgeway et al. 2015; Dong et al. 2020). Indeed, including the post-stratification weights brings the survey samples close to the population census in age, gender and region (Table S1). Weights are normalized to have mean 1 before entering the regressions and are re-estimated for each analysis and subset of the data used, always following the outlined procedure.

A.3 Example

As an illustration, we exemplify the weighting approach for a single respondent surveyed in VIC1 (subject ID `VIC706` in the uploaded data set). The respondent is a 65-year old woman in Hyogo prefecture (we ignore other characteristics for simplicity). Based on the propensity score model shown in Appendix C and her socio-demographic characteristic, she is assigned a predicted probability of being surveyed in WVS7 of $\widehat{\pi} = 0.577$, which is a higher probability than most other VIC1 respondents (main text, Fig. 3a). This implies that she is more similar to the sample of WVS7 respondents than most of her VIC1 co-respondents. Her propensity weight is $\widehat{w} = 0.577/(1 - 0.577) = 1.36$ and her response to the emancipative and secular values questions is therefore upweighted relative to VIC1 respondents with lower $\widehat{\pi}$. In the VIC1 sample, her stratum (female, older than 50 years, Hyogo prefecture) is represented with 0.65%. The census tells, however, that her stratum makes up 1.31% of the Japanese adult population. Therefore, she is assigned a post-stratification weight of $1.31\%/0.65\% = 2.02$ to make up for the under-representation of the stratum in the VIC1 sample compared to the population. The observation's final weight is obtained by multiplying the propensity and post-stratification weight, so here it is $1.36 \times 2.02 = 2.75$.[1]

[1] For WVS7 respondents, the propensity weight is always 1, so the combined weight is equal to the post-stratification weight.

B Distributions of covariates

Figure S1 and Table S1 show distributions for covariates which were used for adjustment. All statistics are based on the samples that remained after listwise deletion of cases which had missing values in any essential variables, i.e., either in the covariates themselves or in *both* outcome variables EVI and SVI.

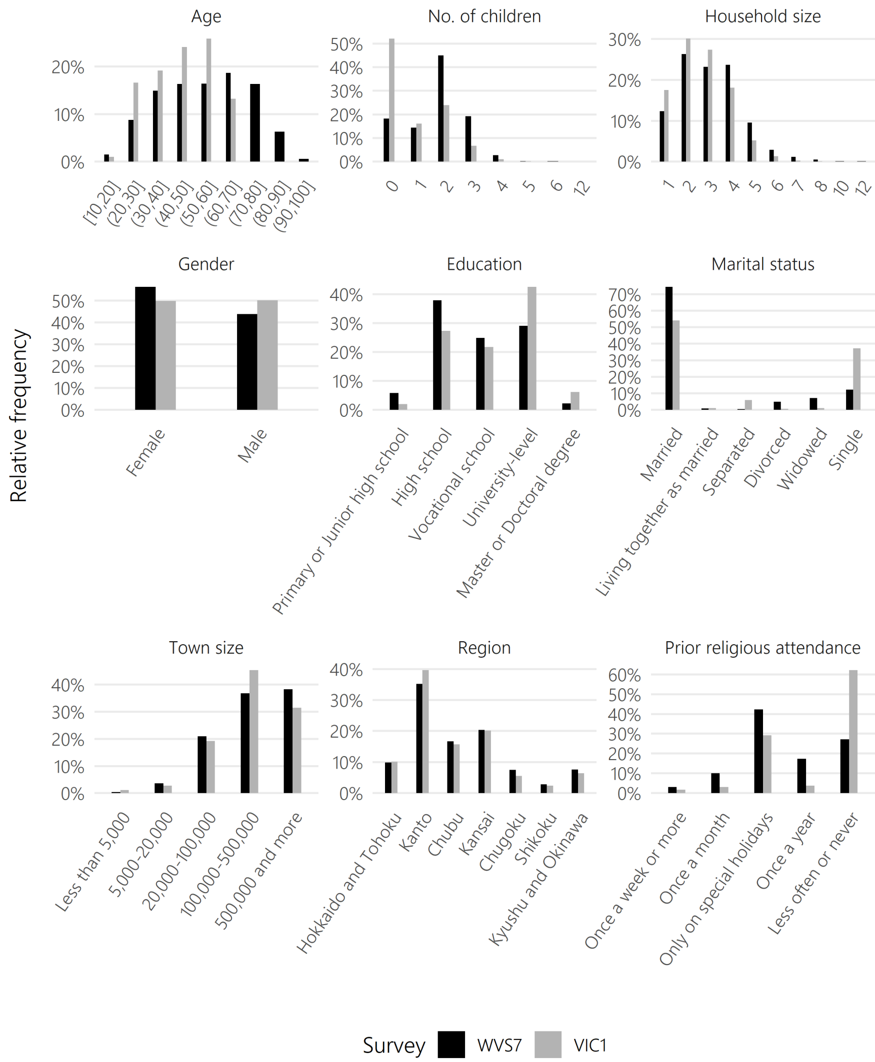

Figure S1: Relative frequencies of characteristics in the WVS7 and VIC1 samples. Raw data prior to weighting is shown.

Table S1: Summary statistics for respondent characteristics in WVS7, VIC1 and VIC2. Distribution for the weighted samples (propensity weights combined with post-stratification weights) for WVS7 and VIC1 are also presented. For some variables, we show statistics for the adult population from the Japanese population census 2020 for comparison.

Characteristic	WVS7	VIC1	VIC2	WVS7 (weighted)	VIC1 (weighted)	Census 2020
N	1138	2920	2827			
Age	54.8 (17.8)	45.5 (12.8)	45.6 (12.9)	53.8 (18.4)	51.1 (11.9)	53.6
No. of children	1.8 (1.1)	0.9 (1.1)	0.8 (1.1)	1.7 (1.1)	1.7 (1.1)	
Household size	3.1 (1.4)	2.7 (1.2)	2.6 (1.2)	3.1 (1.4)	3.2 (1.3)	
Gender						
Female	640 (56%)	1,456 (50%)	1,404 (50%)	53%	53%	52%
Male	498 (44%)	1,464 (50%)	1,423 (50%)	47%	47%	48%
Education						
Primary or Junior high school	67 (6%)	59 (2%)	54 (2%)	6%	10%	
High school	432 (38%)	800 (27%)	763 (27%)	37%	36%	
Vocational school	283 (25%)	636 (22%)	599 (21%)	24%	21%	
University-level	331 (29%)	1,243 (43%)	1,258 (44%)	30%	31%	
Master or Doctoral degree	25 (2%)	182 (6%)	153 (5%)	2%	3%	
Marital status						
Married	847 (74%)	1,580 (54%)	1,472 (52%)	73%	70%	
Living together as married	10 (1%)	29 (1%)	30 (1%)	1%	1%	
Separated	4 (0%)	174 (6%)	162 (6%)	0%	0%	
Divorced	57 (5%)	17 (1%)	9 (0%)	4%	7%	
Widowed	81 (7%)	32 (1%)	34 (1%)	7%	8%	
Single	139 (12%)	1,088 (37%)	1,120 (40%)	15%	14%	
Town size						
Less than 5,000	4 (0%)	33 (1%)	27 (1%)	0%	0%	
5,000-20,000	41 (4%)	82 (3%)	56 (2%)	3%	3%	
20,000-100,000	238 (21%)	561 (19%)	505 (18%)	22%	21%	
100,000-500,000	419 (37%)	1,324 (45%)	1,329 (47%)	34%	31%	
500,000 and more	436 (38%)	920 (32%)	910 (32%)	41%	45%	
Region						
Hokkaido and Tohoku	112 (10%)	294 (10%)	262 (9%)	11%	10%	11%
Kanto	401 (35%)	1,157 (40%)	1,136 (40%)	38%	38%	35%
Chubu	190 (17%)	458 (16%)	449 (16%)	16%	17%	17%
Kansai	232 (20%)	588 (20%)	603 (21%)	19%	21%	18%
Chugoku	85 (7%)	163 (6%)	152 (5%)	5%	5%	6%
Shikoku	32 (3%)	72 (2%)	68 (2%)	2%	2%	3%
Kyushu and Okinawa	86 (8%)	188 (6%)	157 (6%)	9%	8%	11%
Prior religious attendance						
Once a week or more	35 (3%)	50 (2%)	40 (1%)	3%	2%	
Once a month	114 (10%)	93 (3%)	83 (3%)	9%	7%	
Only on special holidays	481 (42%)	854 (29%)	932 (33%)	42%	42%	
Once a year	198 (17%)	109 (4%)	115 (4%)	18%	22%	
Less often or never	310 (27%)	1,814 (62%)	1,657 (59%)	28%	27%	
Big Five: Extraversion		2.6 (0.9)	2.6 (0.9)		2.7 (0.8)	
Big Five: Agreeableness		3.1 (0.7)	3.1 (0.7)		3.1 (0.7)	
Big Five: Conscientiousness		2.8 (0.7)	2.8 (0.7)		2.8 (0.7)	
Big Five: Neuroticism		3.2 (0.7)	3.3 (0.8)		3.2 (0.7)	
Big Five: Openness		3.0 (0.8)	3.0 (0.8)		3.0 (0.8)	
Psychological distress		1.7 (0.8)	1.7 (0.8)		1.7 (0.8)	

C Propensity score regression model

Table S2 shows coefficient estimates and their standard errors from the logistic regression model used for estimation of propensity scores. The outcome variable was whether a respondent was surveyed in WVS7 (coded as 1) or in VIC1 (coded as 0) and the preidcted probabilities were used to contruct inverse probability of treatment weights for the WVS7-VIC1 comparison. Note that the size of estimates for two prefectures (Miyazaki, Tottori) and their uncertainty are extremely large, which is because the WVS7 sample does not include respondents from these two prefectures, which are rather small in terms of population. Hence, the model predicts low propensity scores for the few respondents from those prefectures in the VIC1 sample.

Table S2: Estimated coefficients in the logistic regression model for propensity scores.

Term	Estimate	Standard error
Intercept	-1.429	0.421
Age	0.008	0.004
No. of children	0.346	0.055
Household size	0.243	0.040
Gender		
Female	-	
Male	0.071	0.089
Education		
High school	-	
Master or Doctoral degree	-1.175	0.251
Primary or Junior high school	1.183	0.245
University-level education	-0.577	0.105
Vocational school/University-preparator	-0.132	0.114
Marital status		
Divorced	-	
Living together as married	-1.808	0.529
Married	-2.032	0.302
Separated	-5.185	0.648
Single	-2.091	0.319
Widowed	-0.482	0.376
Town size		
100,000-500,000	-	
20,000-100,000	0.158	0.118
5,000-20,000	0.129	0.263
500,000 and more	0.583	0.107
Less than 5,000	-1.098	0.593
Prefecture		
Aichi	-	
Akita	-0.472	0.499
Aomori	-0.426	0.563
Chiba	0.291	0.244
Ehime	-0.396	0.460
Fukui	-0.765	0.930
Fukuoka	0.116	0.264
Fukushima	-0.069	0.381
Gifu	0.145	0.357
Gunma	0.209	0.408
Hiroshima	0.131	0.315
Hokkaido	0.064	0.260
Hyogo	0.088	0.251
Ibaragi	0.177	0.319
Ishikawa	-0.455	0.555
Iwate	-0.083	0.522
Kagawa	0.491	0.503
Kagoshima	0.371	0.401

Kanagawa	0.043	0.222
Kochi	0.311	0.651
Kumamoto	-0.681	0.407
Kyoto	-0.455	0.360
Mie	-0.114	0.391
Miyagi	-0.541	0.361
Miyazaki	-14.809	288.800
Nagano	0.205	0.367
Nagasaki	-1.163	0.669
Nara	0.158	0.419
Niigata	0.385	0.340
Okayama	-0.032	0.367
Okinawa	-1.583	0.878
Ooita	0.513	0.508
Osaka	0.212	0.226
Saga	-0.397	0.561
Saitama	0.328	0.233
Shiga	-0.152	0.427
Shimane	-0.798	0.806
Shizuoka	-0.087	0.294
Tochigi	0.168	0.372
Tokushima	-1.389	0.696
Tokyo	0.227	0.207
Tottori	-13.799	364.492
Toyama	-0.004	0.494
Wakayama	0.439	0.506
Yamagata	-0.402	0.538
Yamaguchi	0.243	0.451
Yamanashi	-0.724	0.703
Prior religious attendance		
Less often or never	-	
Once a month	1.720	0.182
Once a week or more	1.117	0.263
Once a year	2.311	0.151
Only on special holidays	1.036	0.094
No. Obs.	4058	
Log-likelihood	-1932.139	
Deviance	3571.257	
Residual df	3990	
Null deviance	4648.071	
Null df	4057	
AUC	0.826	

D Variation between prefectures

Variation between prefectures in pandemic severity at the time of the first VIC survey in May 2020, defined as number of COVID-19 infections per 100,000 people, is shown in Figure S2.

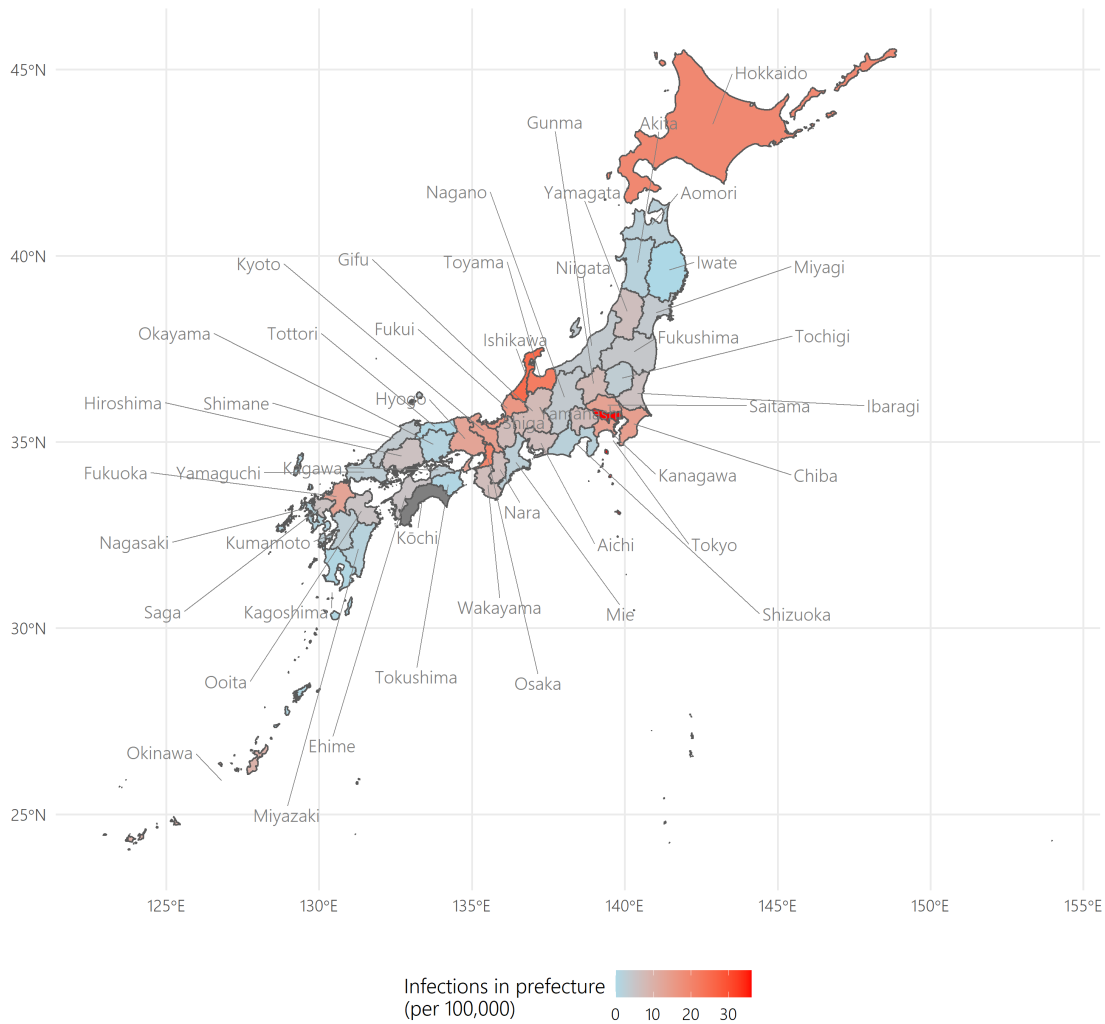

Figure S2: Cumulative COVID-19 infections in Japanese prefectures per 100,000 until initial VIC1 survey date (May 15, 2020).

In Figure S3, we visualize the estimated effect of pandemic severity on change in societal values. In particular, we show how for an 'typical individual' from WVS7, defined by having the (marginal) median or mode in all covariates, the number of cumulative COVID-19 infections in the prefecture affects change in EVI and SVI, based on our outcome regressions shown in the main text. The difference in slopes displays the interaction effect of both variables (see Fig. 6 in the main text for estimated coefficients of the interaction effect). As should be expected under a well-specified model, the estimated number of infections that a prefecture *will* have at the point in time when the VIC1 survey is conducted (May 2020) is not related to societal values at the point in time when the WVS7 survey is conducted (September 2019), given the variables included in the model. This is indicated by the flat solid lines.

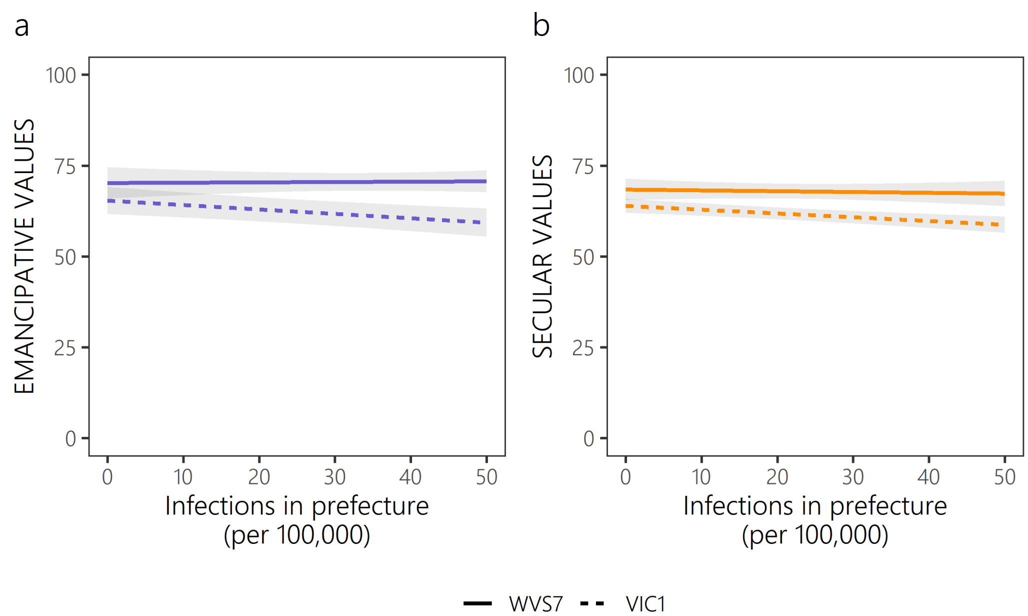

Figure S3: Predicted values for a typical individual as a function of COVID-19 pandemic severity (cumulative infections in the prefecture per 100,000 until the VIC1 survey date) and survey time point. Pointwise 95% confidence intervals are shown.

E Fixed-effect regressions for psychological distress

As part of the analysis on individual-level variation in psychological distress, we provide results from fixed-effect panel models in Figure S4. In the main text, Fig. 7 presents estimates for the effect of psychological distress on emancipative and secular values based on comparisons *between* individuals. Those between-individual estimates are adjusted for wide range of between-individual variation in, e.g., household income or personality type. Here, instead, we present estimates based on comparisons *within* individuals over time by using linear regressions with fixed effects for individual subjects and a dummy predictor variable for survey wave. No further adjustment variables are included here due to their collinearity with the fixed effects. Whereas the between-individual analysis is based only on a single observation per respondent in either VIC1 or VIC2, the within-individual analysis shown here is based only on respondents who completed both survey waves (sample sizes in Figure S4 refer to number of observations, which is double the number of individual subjects).

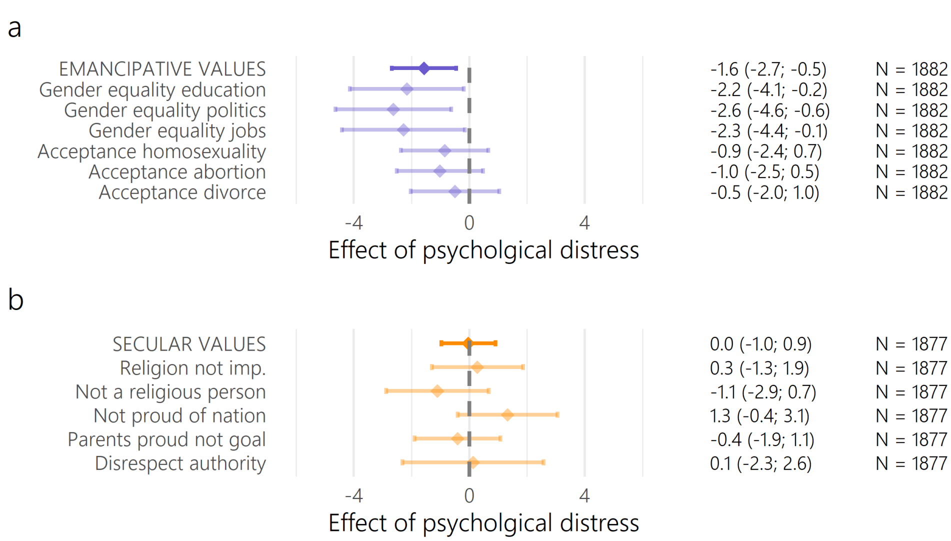

Figure S4: [...]

An advantage of the fixed effects approach is that time-invariant (but not time-variant) unobserved variation between individuals is controlled for, which eliminates further potential confounders - in particular, unobserved variables which differ between individuals, do not vary over time, are in a time-invariant way associated with both psychological distress and emancipative or secular values and whose effect is not yet captured by our other adjustment variables. A clear disadvantage in our setting is that the fixed effects approach can only exploit within-individual variation in distress between VIC1 and VIC2 (which is comparably low, see main text) and therefore discards all (co-)variation in distress and values that occurred during the initial onset of the COVID-19 pandemic (i.e., between WVS7 and VIC1). This fact limits its efficiency. Nevertheless, also in the fixed effect models we see a negative relationship between psychological distress and emancipative values (Figure S4): People whose distress decreased between May 2020 and April 2021 on average also report higher emancipative values in April 2021 than in May 2020, most strikingly in terms of gender equality.

For SVI and its items, there is no clear evidence for an association (Figure S4). This differs from the results of the between-individual analysis (main text, Fig. 7), which are more in line with all the other results presented in the main text. We cannot give definitive explanations for the deviating results in the fixed-effect analysis of secular values. For the outlined reasons of non-random drop-out[2], sample size and extent of variation in

[2] People with higher distress at VIC1 were more likely to subsequently drop out of the VIC survey (see main text) and there is also evidence

exposure, we generally favor the between-individual analysis presented in the main text, but acknowledge that results in Figure S4 cast some additional uncertainty on the strength and existence of individual-level effects of psychological distress on religious values.

---

for drop-outs being less secular at VIC1 than people who remained in the sample until VIC2 (65.2 vs. 66.5 points, $p = 0.026$). A similar relationship is not found with respect to emancipative values (61.5 vs. 61.6 points, $p = 0.84$).

References

Austin, Peter C., and Elizabeth A. Stuart. 2015. "Moving Towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies." *Statistics in Medicine* 34 (28): 3661–79. https://doi.org/10.1002/sim.6607.

Dong, Nianbo, Elizabeth A. Stuart, David Lenis, and Trang Quynh Nguyen. 2020. "Using Propensity Score Analysis of Survey Data to Estimate Population Average Treatment Effects: A Case Study Comparing Different Methods." *Evaluation Review* 44 (1): 84–108. https://doi.org/10.1177/0193841X20938497.

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Levy, Paul S, and Stanley Lemeshow. 2008. *Sampling of Populations: Methods and Applications*. 4th ed. John Wiley & Sons.

Ridgeway, Greg, Stephanie Ann Kovalchik, Beth Ann Griffin, and Mohammed U. Kabeto. 2015. "Propensity Score Analysis with Survey Weighted Data." *Journal of Causal Inference* 3 (2): 237–49. https://doi.org/10.1515/jci-2014-0039.

Statistics Bureau of Japan. 2021. "2020 Population Census." https://www.stat.go.jp/english/data/kokusei/index.html.