The Interdependent Impacts of Regional COVID-19 Re-openings in the United States

Michael Zhao[1] & Sinan Aral[1]

[1]*MIT Sloan School of Management, 100 Main St, Cambridge, MA 02142, United States*

After limits on human mobility effectively reduced the spread of COVID-19 around the world [1,2], many countries began to reopen. When these reopenings began in the U.S., several COVID-19 hotspots emerged, causing some local governments to reimpose local shutdowns. However, we know little about the impacts of regional reopenings or subsequent shutdowns [3,4], and have no quantitative evidence on the direct impact of a region's reopening policies on its own population's mobility; the spillover effects of peer regions' policies on a focal region's mobility; the mediation of these effects by endogenous peer behavior across regions; or the impacts of origin and destination policies on cross-region travel. Here we show that individual states' ad hoc local reopening policies significantly influenced mobility across the entire U.S. due to inter-state travel and social influence. When all peer states locked down, focal county mobility dropped by 15-20% but increased by 19-32% once peer states reopened. When an origin county was subject to a statewide shelter-in-place order, travel to counties yet to impose lockdowns increased by 52-65%. If the origin reopened, but the destination remained closed, travel to destination counties was suppressed by 9-17% for nearby counties and 21-27% for distant counties. But, when a destination reopened while an origin remained closed, people from the closed origins flooded into the destination by 11-12% from nearby

counties and 24% from distant counties. Our findings demonstrate how reopenings contribute to the emergence of new hotspots and, counterintuitively, how the re imposition of shutdown orders can increase mobility as citizens flee to open peer regions. The research highlights the risks of ad hoc local reopenings and the urgent need to coordinate COVID-19 reopenings across regions.

Population-scale digital trace data[5] has been useful for studying the impacts of social distancing policies and what makes them successful[6]. Researchers have shown, for example, that demographic attributes[7], political partisanship[8,9], broadband access[10], belief in science[11], and information exposure[12,13] moderate compliance with social distancing policies. Holtz et al[14] found strong evidence of cross-county spillovers from shelter-in-place policies, underscoring the importance of governmental coordination to reduce a potential "loss from anarchy" in piecemeal implementations of closure policies across regions. Unfortunately, Holtz et al only analyzed data from March and April, 2020, before shelter-in-place policies began to lift, and little systematic research investigates the effects of subsequent reopening policies across regions.

If reopenings create substantial mobility and exhibit strong regional spillover effects, countries that reopen without national coordination could face significant difficulty in controlling the resurgent spread of the coronavirus. We therefore combined data on the mobility of over 22 million mobile devices, daily data on state-level closure and reopening policies, social media connections among 220 million U.S. Facebook users, temperature and precipitation data from 62,000 weather stations, and county-level census data to measure the direct impact of a focal state's COVID-19

closure and reopening policies on its own population's mobility patterns; the spillover effects of alter states' closure and reopening policies on a focal state's mobility patterns; and the mediation of these effects by endogenous peer behavior across state and county borders from January 1st to July 1st, 2020. We further investigated the impacts of both origin- and destination-county closure and reopening policies on cross-county mobility, capturing the travel related spillover effects created by uncoordinated policies implemented across states and counties.

Our measures of human mobility are constructed using data provided by Safegraph. For each county, we track the daily average number of locations visited by mobile devices, the proportion of devices traveling more than 2km, the proportion of devices that spend over an hour away from home, and the proportion of devices leaving home. Time series trends of these four measures are plotted in Fig. 1A. For each origin-destination county pair, we tracked both the number and proportion of devices moving from an origin county to a destination county. Differences in the log number of devices traveling to specific counties before and after lockdowns and reopenings are plotted in Fig. 1B.

We analyze state-level data on closure and reopening policies from the COVID-19 US State Policy (CUSP) Database [15]. Due to the different approaches taken by various states across the closure and reopening policy space, we simplify our analysis to three consolidated "policy periods:" the "initial policy period" (ip), which covers the period from when a state implements its first closure policy of any kind until it implements a stay-at-home order; the "stay-at-home period" (sh), which covers the duration of a statewide stay-at-home order or until the state starts to reopen; and

the "reopening period" (ro), which starts after a state begins its reopening plan (Fig. 1C).

To construct peer policy and mobility measures, we rely on Facebook's "Social Connectedness Index[16]" (SCI), which provides a measure of the intensity of Facebook connectedness between geographic locations, generated from an anonymized snapshot of the entire Facebook friendship network in the U.S. (Fig. 1D). These data are further supplemented with temperature and precipitation data from the Global Historical Climatology Network[17] (GHCN) and county-level census population estimates.

Analysis from a "no spillovers" difference-in-differences model indicate that statewide shelter-in-place orders reduced mobility within a state by 5-6% on average (Fig. 2A). Once a state reopened, mobility increased by 3-5% on average, returning to levels statistically indistinguishable from pre-pandemic levels. When accounting for alter state policy spillovers, these ego state policy estimates, while lower, are not significantly different from the base model estimates. Consistent with Holtz et al[14], we also find strong evidence of spillover effects in social distancing policies and closures, and extend the analysis to reopenings. When all alter states begin implementing social distancing policies, ego county mobility drops by 4-7%. When all alter states impose a lockdown, ego county mobility drops by an additional 15-20%. However, after all alter states begin reopening, an ego county's mobility increases by 19-32%. All of these effects account for ego counties' policies and estimate the additional effects of alter states' policies on ego county mobility.

When we further account for endogenous alter state mobility behavior using an instrumental variables analysis, the coefficients of alter-state closures and reopenings become statistically indis-

Figure 1: **Data on Mobility Behaviors, State Policy Timing, and Social Connectedness.**
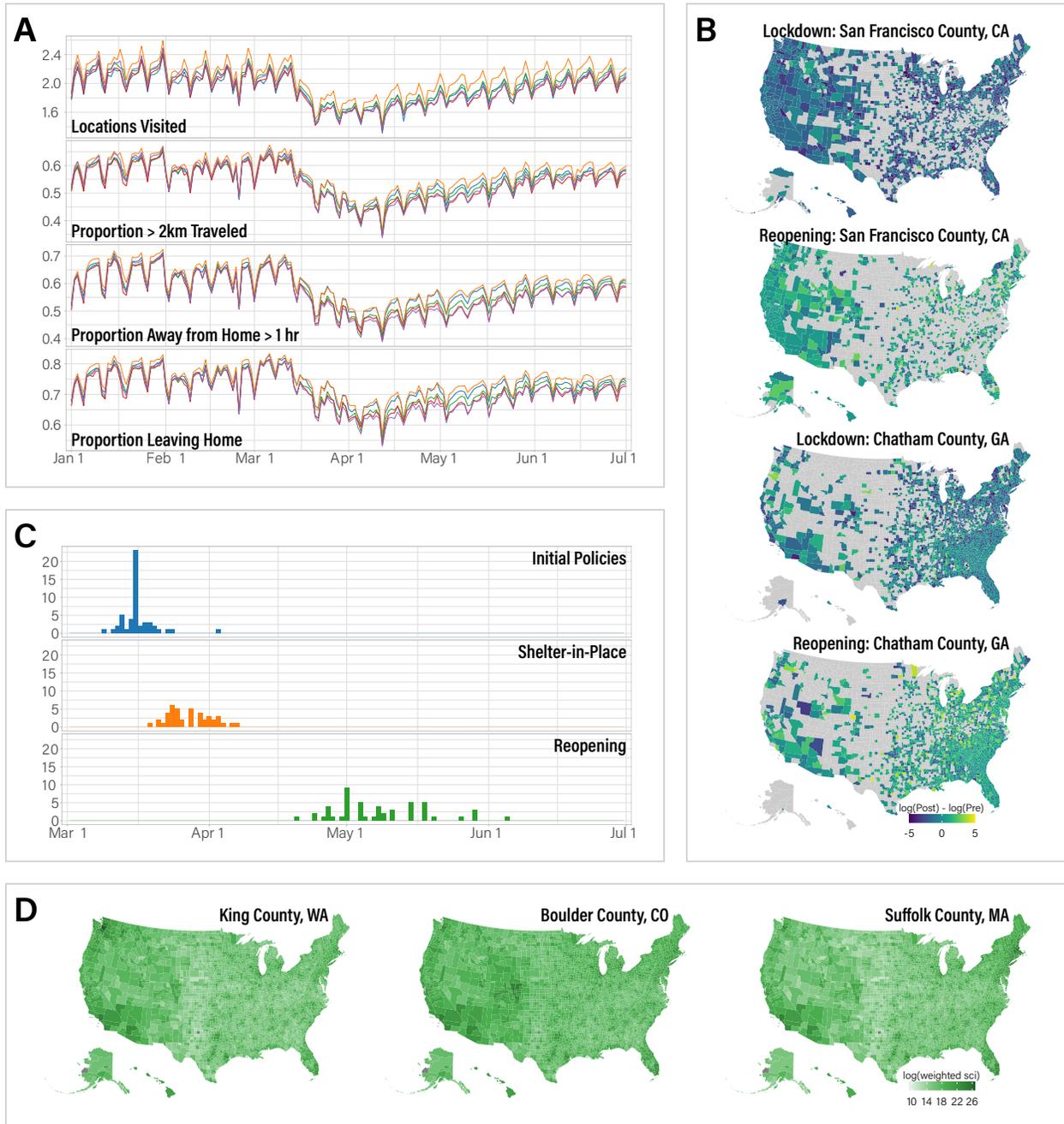
(A) shows the time series trends for the number of locations visited per device, the proportion of devices traveling more than 2 km, the proportion of devices the spend more than an hour away from home, and the proportion of devices leaving home. Each color represents the averages across clusters of 10 states grouped by how soon they reopened. (B) shows examples of the difference in travel to a destination county for the 3 weeks before and after a lockdown or reopening. (C) plots the count of states that enter into a particular policy period on each day. (D) shows examples of the population-weighted social connectedness index used to construct socially weighted measures of alter state policies and behavior.

tinguishable from zero, indicating that alter state policy effects on ego counties are largely mediated by endogenous peer mobility behavior in alter states. The estimate of the peer effect coefficient itself is also quite significant, ranging from 1.8-2.2, meaning that a 1% change in out-of-state peer mobility causes between a 1.8-2.2% change in ego county mobility (Fig. 2B).

Our cross-state mobility analysis also shows clear evidence of policy spillovers (Fig. 3). We find that destination counties under statewide shelter-in-place orders receive 8-14% less cross-state traffic compared to pre-pandemic levels. These estimates also exhibit notable heterogeneity as travel from "distant" counties decreased by 13-18%, while there is no measurable impact on travel from "nearby" counties. As expected, reopenings boost travel to destination counties, by 12-13%, with no detectable differences between effects on nearby or distant counties.

When we expand our this analysis to include all possible origin-destination policy interactions (Fig. 4), several more key findings emerge. When origin counties were in their initial policy period (having implemented social distancing policies but before shelter-in-place orders were imposed), destination policies did not measurably impact cross state mobility. But, when origin counties entered the stay-at-home period (having implemented statewide shelter-in-place orders), travel to distant counties decreased by 10% while travel to nearby counties increased by 52-65%, if those counties had not yet implemented a shelter-in-place order. Once destinations implemented shelter-in-place orders, distant cross-state travel decreased by 14-16%, with no detectable effect on nearby cross-state travel. When a destination reopened and an origin was still locked down, people from origin counties not in that state flooded into the destination, by 11-12% from nearby counties

6

Figure 2: **Empirical Estimates of Ego State Policy, Alter State Policy, and Endogenous Peer Behavior on County-Level Mobility.**
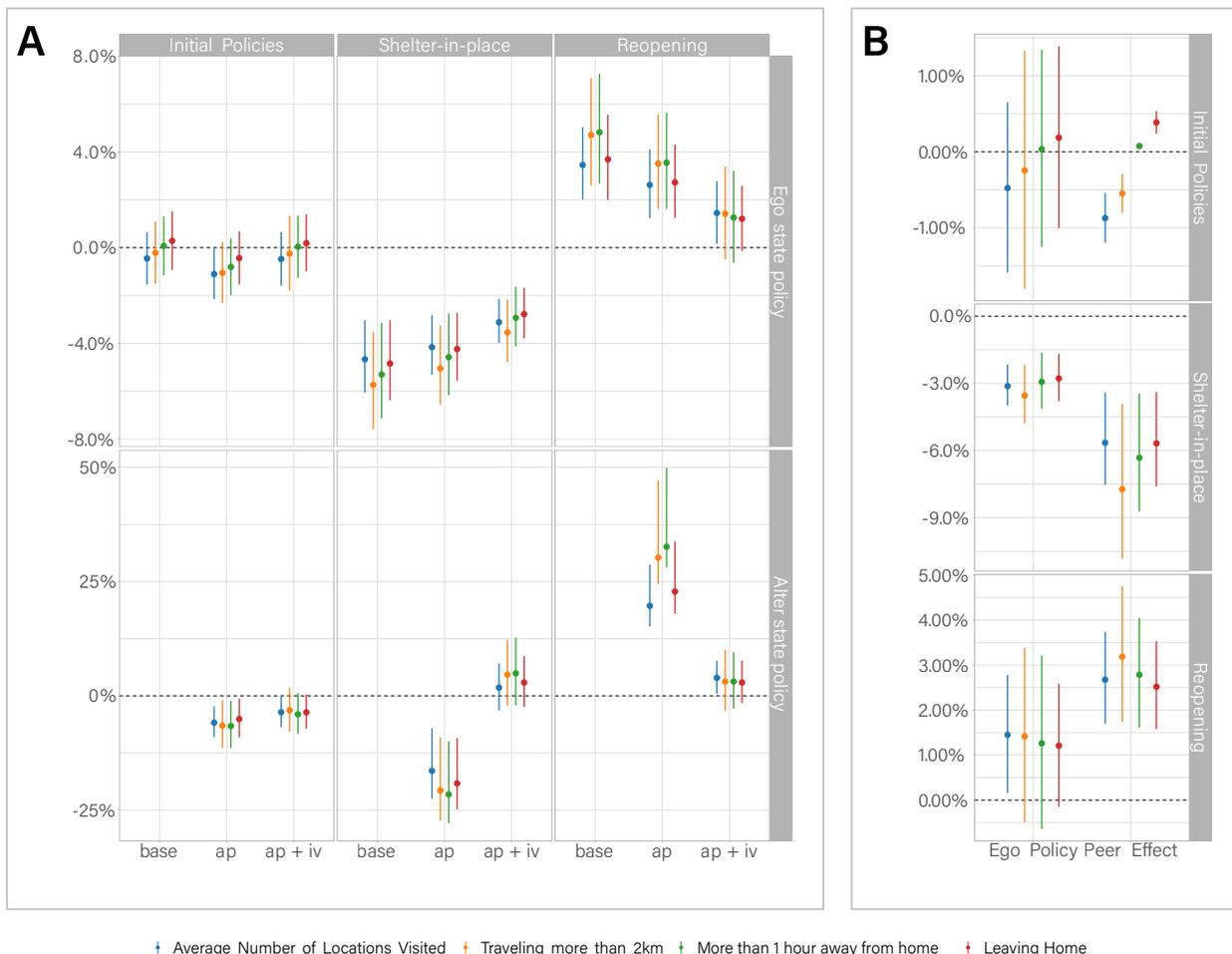
Figure 3: (A) plots the estimated impact of both ego and alter state policies. The top row corresponds to ego state's policy periods and the bottom row corresponds to the alter states' policy periods. The x-axis denotes the the model specification used to generate the estimates: "base" which corresponds to DiD without spillovers, "ap" which corresponds to DiD with alter state policies, and "ap + iv" which corresponds DiD with alter state policies and endogenous peer behavior. The "base" and "ap" estimates are produced using weighted least squares, with weights determined by county population. The "ap + iv" estimates are produced using two stage weighted least squares, where peer behavior is instrumented for with using alter state weather. (B) plots the coefficient estimates of ego state policy and endogenous peer behavior. The magnitudes of the endogenous peer effects are scaled to point estimates of the ego state policy.

7

Figure 4: **Empirical Estimates of Origin and Destination Policies on Cross-State Travel.**
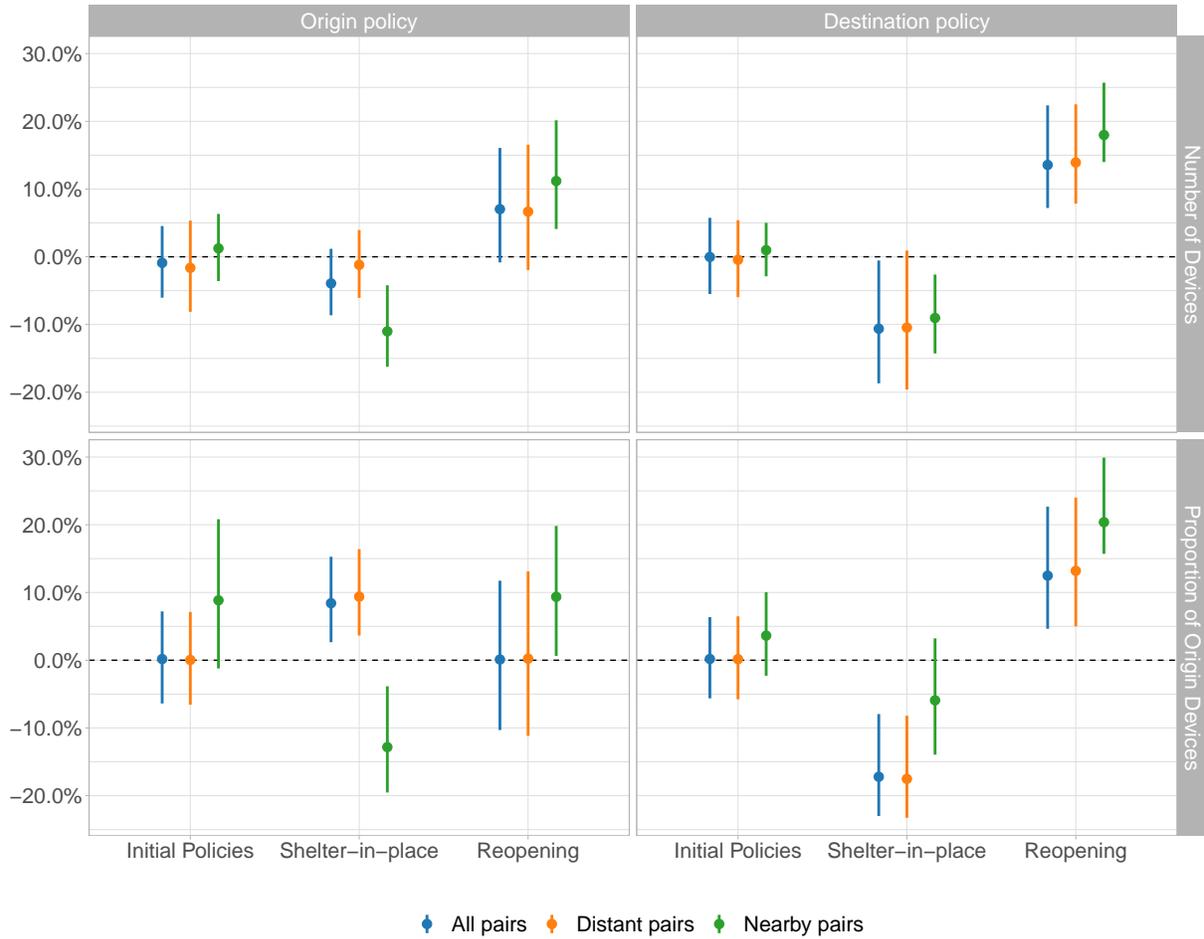
In this figure, the left block corresponds to the origin policy periods, while the the right block corresponds destination policy periods, denoted across the x-axis as initial policies, shelter-in-place, and reopening. The top row reflects policy effects on the number of devices moving from an origin to a destination, estimated using OLS and the bottom row reflects policy effects on the proportion of origin devices moving from an origin to a destination estimated using WLS with weights proportional to origin county population. Colors correspond to estimates produced using all pairs in blue, distant pairs (¿ 100km) in orange, and nearby pairs (¡ 100km) in green. 2-way origin and destination state clustered standard errors are used to compute 95% confidence intervals.

and 24% from distant counties. If the origin had reopened, but the destination was still closed, travel to both nearby and distant destination counties was suppressed by 9-17% and 21-27% respectively. Once both origin and destination counties reopened, there was a 14-19% increase in travel from distant origins, though there was no change in travel from nearby origins.

To our knowledge, ours is the first large-scale study of the impacts of reopenings on mobility, explicitly estimating cross-state spillovers and the mediation of cross-state policy effects by peer behaviors. But, this work is not without its limitations. First, while the Safegraph panel is sufficiently large to minimize concerns about sampling error, it may exhibit sampling bias as mobile device ownership significantly varies by age and income.[1] While Safegraph has shown their panel is geographically consistent with US Census population estimates,[2] it is not clear if certain demographics are over- or under-represented as no device-level demographic data is collected by Safegraph. Though it is reassuring that we find similar results when using mobility data provided by Facebook as a robustness check, there are also concerns about the representativeness of Facebook's data as well. Second, while we investigated potential anticipatory or lagging behaviors, our analysis does not explore the impacts of discrete reopening policies (e.g. resuming restaurant dine-in service or lifting gathering restrictions) and instead measures average changes in mobility behavior across "policy periods." Third, our analysis only captures variation in state-level closures and reopenings and ignores the relatively few instances in which local- or county-level policy differs from the state. We encourage such analysis for reopenings in future work because disputes between states and localities may further thwart efforts to reduce mobility and control the

---

[1]See https://www.pewresearch.org/internet/fact-sheet/mobile/. Accessed August 2020.
[2]See https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset. Accessed August 2020.

Figure 5: **Empirical Estimates of Origin and Destination Policy Interactions on Cross-State Travel.**
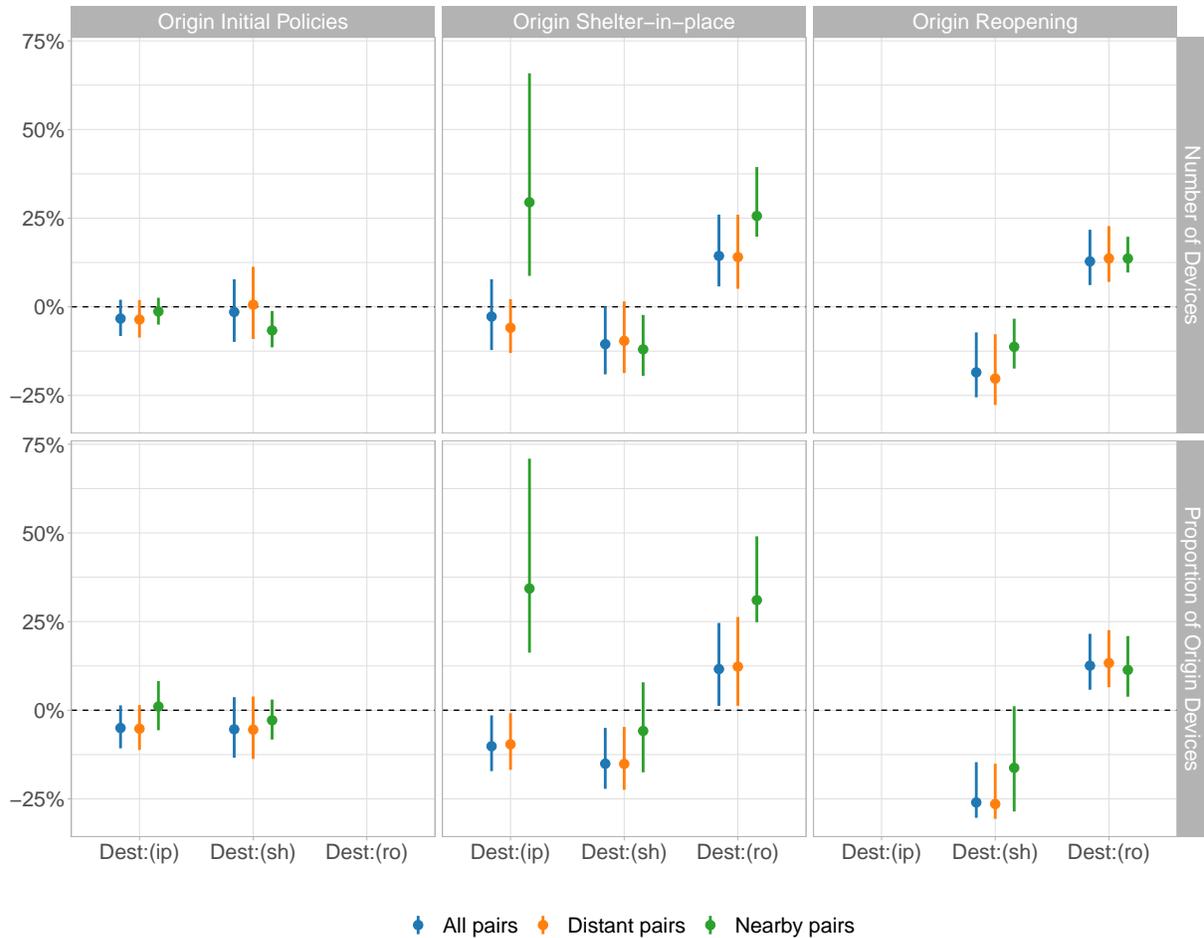
Figure 6: Each block in the figure corresponds to different origin policy periods: initial policies, shelter-in-place, and reopening from left to right. The top row reflects policy effects on the number of devices moving from an origin to a destination, estimated using OLS and the bottom row reflects policy effects on the proportion of origin devices moving from an origin to a destination estimated using WLS with weights proportional to origin county population. Different values along the x-axis of each column correspond to different destination policy periods: destination initial policies Dest:(ip), destination shelter-in-place Dest:(sh), and destination reopening Dest:(ro). Within each column, the marginal effects of each destination policy given the origin policy. Colors correspond to estimates produced using all pairs in blue, distant pairs (¿ 100km) in orange, and nearby pairs (¡ 100km) in green. 2-way origin and destination state clustered standard errors are used to compute 95% confidence intervals.

pandemic. Fourth, our analysis is restricted to mobility outcomes and purposefully avoids extrapolating to health outcomes like morbidity and mortality. While it is widely believed that reductions in mobility drive reductions in new infections and their associated deaths[18], rigorously establishing the causal chain from cross-state spillovers to infection rates and deaths is beyond the scope of this paper.

Despite these limitations, our work provides critical information for policy makers, especially given the importance of restricting mobility to prevent the spread of COVID-19. Though many countries have seemingly reopened safely, new hotspots may yet emerge, forcing governments to reimpose mobility restrictions. Our results show that it is crucially important to take spillover effects into account when formulating national policy and for national and local policies to coordinate across regions where spillovers are strong. Our results suggest that reimposing local social distancing or shelter-in-place orders may be far less effective than policy makers would hope when peer states and counties remain reopened, due to travel and peer influence. In fact, such closure policies may actually be counterproductive[19,20] as they can encourage those in locked down regions to flee to reopened regions, potentially causing new hotspots to emerge. Our analysis demonstrates that such travel spillovers are not only systematic and predictable, but also large and thus meaningful to our public health.

Methods

Empirical Methodology Our empirical methodology is grounded in a reduced form econometric approach called difference-in-differences (DiD), a widely used approach across economics, po-

litical science, and public health for policy evaluation. We begin with a basic model that only

incorporates each county's own state policy specified as follows:

$$\log(Y_{it}) = \delta_{(ip)} D_{it}^{(ip)} + \delta_{(sh)} D_{it}^{(sh)} + \delta_{(ro)} D_{it}^{(ro)} + f(W_{it}) + \alpha_i + \tau_t + \epsilon_{it}, \tag{1}$$

where $\log(Y_{it})$ are the log transformed mobility outcomes. The policy variables, $D_{it}^{(ip)}$, $D_{it}^{(sh)}$, and

$D_{it}^{(ro)}$ are binary indicators that take the value 1 once county i is subject to a statewide closure

policy of any sort (ip), a stay-at-home order (sh), and reopening (ro) respectively. The associated

parameters $\delta_{(ip)}$, $\delta_{(sh)}$, and $\delta_{(ro)}$ estimate the marginal mobility effects of moving into the corre-

sponding policy periods. $f(W_{it})$ flexibly controls for the potential non-linear effects of weather

using a "double machine learning" approach [21], while α_i and τ_t denote a set of county and time

fixed effects and ϵ_{it} denotes the error term.

We extend this base specification to capture spillover effects with the following specifica-

tions:

$$\log(Y_{it}) = \delta_{(ip)} D_{it}^{(ip)} + \delta_{(sh)} D_{it}^{(sh)} + \delta_{(ro)} D_{it}^{(ro)} +$$
$$\gamma_{(ip)} D_{-it}^{(ip)} + \gamma_{(sh)} D_{-it}^{(sh)} + \gamma_{(ro)} D_{-it}^{(ro)} + f(W_{it}) + \alpha_i + \tau_t + \epsilon_{it} \tag{2}$$

$$\log(Y_{it}) = \beta \log(Y_{-it}) + \delta_{(ip)} D_{it}^{(ip)} + \delta_{(sh)} D_{it}^{(sh)} + \delta_{(ro)} D_{it}^{(ro)} +$$
$$\gamma_{(ip)} D_{-it}^{(ip)} + \gamma_{(sh)} D_{-it}^{(sh)} + \gamma_{(ro)} D_{-it}^{(ro)} + f(W_{it}) + \alpha_i + \tau_t + \epsilon_{it}, \tag{3}$$

where $D_{-it}^{(ip)}$, $D_{-it}^{(sh)}$, and $D_{-it}^{(ro)}$ denote the socially weighted average of alter states' policies, weighted

by Facebook connectedness, and where the cross-state policy spillovers are captured by the terms

$\gamma_{(ip)}$, $\gamma_{(sh)}$, and $\gamma_{(ro)}$ respectively. Endogenous peer behavior is captured by $\log(Y_{-it})$, which is the

log transformed socially weighted average of alter states' mobility behavior. As estimation of peer effects is generically confounded in observational data [22], we employ an instrumental variables (IV) approach where we leverage alter county weather as a source of exogenous variation to properly identify the endogenous peer effect β[14,23–25]. For these specifications, we limit our analysis to the 2683 counties with a daily mean device count of at least 500 to minimize measurement error induced by the Laplacian noise introduced by Safegraph's differential privacy algorithm.

To measure the impact of policy on cross-county mobility, we employ the following specifications:

$$\log(Y_{o \to d,t}) = \sum_m \lambda_m D_{ot}^m + \sum_n \psi_n D_{dt}^n + \alpha_{o \to d} + \tau_t + \epsilon_{o \to d,t} \tag{4}$$

$$\log(Y_{o \to d,t}) = \sum_m \lambda_m D_{ot}^m + \sum_n \psi_n D_{dt}^n + \sum_m \sum_n \pi_{m,n}(D_{ot}^m * D_{dt}^n) + \alpha_{o \to d} + \tau_t + \epsilon_{o \to d,t} \tag{5}$$

Here, $\log(Y_{o \to d,t})$ refers to the log transformed cross-state mobility from an origin county o moving to a destination county in a different state d on date t. Origin and destination policies are denoted by D_{ot}^m and D_{dt}^n respectively, where $m, n \in \{(ip), (sh), (ro)\}$; $\alpha_{o \to d}$ and τ_t correspond to directed dyad and time fixed effects; and $\epsilon_{o \to d,t}$ represents the error term. Equation 4 models the impacts of origin and destination policies linearly whereas Equation 5 includes all possible interactions between origin and destination policies.

Data Availability Access to Safegraph's COVID-19 mobility data can be requested here. Access to Facebook's Social Connectedness Index can be requested by emailing sci_data@fb.com. State-level social distancing and reopening data is openly available here. Weather data is openly available

1. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature* 1–5 (2020).

2. Hsiang, S. *et al.* The effect of large-scale anti-contagion policies on the covid-19 pandemic. *Nature* 1–9 (2020).

3. Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P.-Y. & Colizza, V. Expected impact of reopening schools after lockdown on covid-19 epidemic in île-de-france. *medRxiv* (2020).

4. Glaeser, E. L., Jin, G. Z., Leyden, B. T. & Luca, M. Learning from deregulation: The asymmetric impact of lockdown and reopening on risky behavior during covid-19. Working Paper 27650, National Bureau of Economic Research (2020). URL http://www.nber.org/papers/w27650.

5. Buckee, C. O. *et al.* Aggregated mobility data could help fight COVID-19. *Science* (2020).

6. Oliver, N. *et al.* Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle (2020).

7. Olsen, A. L. & Hjorth, F. Willingness to distance in the COVID-19 pandemic. Tech. Rep., University of Copenhagen (2020).

8. Painter, M. & Qiu, T. Political beliefs affect compliance with COVID-19 social distancing orders. Tech. Rep. (2020). Available at SSRN 3569098.

9. Allcott, H. *et al.* Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. Tech. Rep. 26946, National Bureau of Economic Research (2020).

10. Chiou, L. & Tucker, C. Social distancing, internet access and inequality. Tech. Rep. 26982, National Bureau of Economic Research (2020).

11. Brzezinski, A., Kecht, V., Van Dijcke, D. & Wright, A. L. Belief in science influences physical distancing in response to covid-19 lockdown policies. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2020).

12. Simonov, A., Sacher, S. K., Dubé, J.-P. H. & Biswas, S. The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic. Tech. Rep., National Bureau of Economic Research (2020).

13. Ash, E., Galletta, S., Hangartner, D., Margalit, Y. & Pinna, M. The effect of fox news on health behavior during covid-19. *Available at SSRN 3636762* (2020).

14. Holtz, D. *et al.* Interdependence and the cost of uncoordinated responses to covid-19. *Proceedings of the National Academy of Sciences* (2020). URL https://www.pnas.org/content/early/2020/07/29/2009522117. https://www.pnas.org/content/early/2020/07/29/2009522117.full.pdf.

15. Raifman, J. *et al.* Covid-19 us state policy database (2020).

16. Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives* **32**, 259–80 (2018).

17. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology* **29**, 897–910 (2012).

18. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in china. *Science* (2020).

19. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* (2020).

20. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).

21. Chernozhukov, V. *et al.* Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* **21** (2018).

22. Manski, C. F. Identification of endogenous social effects: The reflection problem. *The review of economic studies* **60**, 531–542 (1993).

23. Coviello, L. *et al.* Detecting emotional contagion in massive social networks. *PloS one* **9** (2014).

24. Aral, S. & Nicolaides, C. Exercise contagion in a global social network. *Nature Communications* **8**, 1–8 (2017).

25. Aral, S. & Zhao, M. Social spillovers in online news consumption. *Available at SSRN 3328864* (2020).

Supplementary Note 1: Data and Data Processing Procedures

Safegraph Data Our primary measures of human mobility are constructed from data provided by Safegraph[1], a San Francisco-based company that sells data related to points of interest that are relevant to businesses. Safegraph collects anonymized human geo-spatial data from a number of partner mobile applications that need to obtain affirmative opt-in consent from device users. We specifically make use of Safegraph's "Social Distancing Metrics" dataset which provides daily measures of mobility behavior aggregated at the census block group level starting from January 1, 2020. In this data, each device's "home" location is assigned by determining its common nighttime location across a period of 6 weeks at a geohash-7 granularity (approx. 153m × 153m). We specifically make use of the following fields[2]:

- `origin_census_block_group`: The unique 12-digit FIPS code for the Census Block Group. Please note that some CBGs have leading zeros.

- `date_range_start`: Start time for measurement period in ISO 8601 format of YYYY-MM-DDTHH:mm:SS±hh:mm (local time with offset from GMT). The start time will be 12 a.m. of any day.

- `device_count`: Number of devices seen in our panel during the date range whose home is in this census block group. Home is defined as the common nighttime location for the device over a 6 week period where nighttime is 6 pm - 7 am. Note that we do not include any census block groups where the count <5.

---

[1]https://www.safegraph.com
[2]Descriptions copied directly from https://docs.safegraph.com/docs/social-distancing-metrics

- `bucketed_distance_traveled`: Key is range of meters (from geohash-7 of home) and value is device count. If a device made multiple trips, we use the median distance for the device.

- `completely_home_device_count`: Out of the device count, the number of devices which did not leave the geohash-7 in which their home is located during the time period.

- `bucketed_home_dwell_time`: Key is range of minutes and value is device count of devices that dwelled at geohash-7 of home for the given time period. For each device, we summed the observed minutes at home across the day (whether or not these were contiguous) to get the total minutes for each device this day. Then we count how many devices are in each bucket. Beginning in v2, we include the portion of any stop within the time range regardless of whether the stop start time was in the time period.

- `destination_cbgs`: Key is a destination census block group and value is the number of devices with a home in census block group that stopped in the given destination census block group for >1 minute during the time period. Destination census block group will also include the origin census block group in order to see if there are any devices that originate from the origin census block group but are staying completely outside of it.

To further preserve privacy, Safegraph applies a differential privacy algorithm[1] to all metrics that it computes other than `device_count`. We use this data to construct 4 measures of county level mobility and 2 measures of cross-county dyadic travel.

From this data, we construct the following 4 county-level measures of mobility: mean census block group visted (`mcgbv`), proportion of devices with greater than 2km traveled

(`pgt2kmt`), proportion of devices spending more than 1 hour away from home (`pgt1hafh`) and proportion of not completely home devices (`pnchd`). We construct `mcgbv` by first summing across the number of non home census block groups in the `destination_cbgs` field. We aggregate this count to the county level and simply divide by the count-level sum of `device_count`. To build `pgt2kmt`, we first appropriately sum across the `bucketed_distance_traveled` field at the census block group level, aggregate to the county level, and then divide by `device_count`. `pgt1hafh` is constructed in a similar manner, except that we instead sum across the `bucketed_home_dwell_time` field. `pnchd` is simply defined as 1 minus `completely_home_device_count` divided by `device_count` aggregated to the county level. In our analysis, we use the log transformations of each of these measures: `log_mcgbv`, `log_pgt2kmt`, `log_pgt1hafh`, and `log_pnchd`. To minimize the impact of the Laplacian noise introduced by the differential privacy algorithm, we limit our analysis to counties with a mean `device_count` greater than 500.

We also create the following 2 measures of cross-county dyadic travel: number of devices moving from origin to destination `ndotd` and proportion of origin devices moving from origin to destination `pdotd`. To construct both these measures, we first build the a directed dyad list by keeping track of the `origin_census_block_group` and unrolling the `destination_cbgs`. We then aggregate by summing values to the origin county / destination county for each day to build `ndotd`. To get to `pdotd`, we simply divide `ndotd` by each day's county-level `device_count`. This data is quite sparse as there is little travel between most county pairs on most days. For our analysis, we limit ourselves to directed dyads with at least some travel between them for each day in our dataset.

COVID-19 US State Policy Database (CUSP) Our policy data comes from the COVID-19 US State Policy Database[2] assembled by researchers at the Boston University School of

Public Health. This database tracks all state-wide wide directives and mandates, but not recommendations. It keeps track of state policies like gathering bans, entertainment closures, business closures, shelter-in-place orders, reopenings, and more for all 50 states plus Washington DC. As of this writing, the latest update to the database was made on Aug. 5, 2020. As mentioned in the main paper, we avoid quantifying the impact of each policy individually, as there is simply not enough data to generate reliable estimates for such a high-dimensional policy space. We instead consolidate our analysis down to 3 main policy periods: the period from the first statewide social distancing policy of any kind until a shelter-in-place order takes effect or the "initial policies" period (ip); the period in which a shelter-in-place order is in effect (or until reopening begins) or the "stay home" period (sh); and the period after reopening begins or the "reopening" period (ro).

Facebook Social Connectedness Index (SCI) The Social Connectedness Index[3]—released as a part of Facebook's Data for Good[3] Initiative—constructs a measure of "connectedness" between two counties (or NUTS3 regions outside the US) based on the friendship ties between them. It is constructed from an anonymized snapshot of the global Facebook friendship graph of over 2.45 billion users. Specifically, the `sci` between two counties is computed as:

$$\text{sci}_{ij} = \frac{\text{fb\_connections}_{ij}}{\text{fb\_users}_i \times \text{fb\_users}_j} \tag{S1}$$

The numerator, fb_connections$_{ij}$ is just the number of friendship ties that are empirically observed between users in county i and j, while the denominator is simply the product between the number of Facebook users that reside in county i (fb_users$_i$) and county j (fb_users$_j$). Therefore sci$_{ij}$ can be interpreted as the probability that a friendship link exists between a random user that resides in i and a random user that resides in j. The SCI reports a `scaled_sci` measure that divides each of the values by the maximum, multiplies by

---

[3]https://dataforgood.fb.com/

1,000,000,000, and then rounding to the nearest integer.

The version of the SCI we use for our analysis is based on snapshot take on December 31, 2019. Rather than using the `scaled_sci` measure directly, we instead weight it by the population of the friend county (using 2018 estimates from the US Census) to capture the relative differences in the number of ties coming from alter counties. For example, the `scaled_sci` between New York City and Boulder, Colorado is relatively low due to the population of Facebook users in NYC. However, the number of friendship links between the two counties is relatively high, again due to the fact of NYC's population. Several examples of this population weighted sci measure can be seen in Fig. 1d of main text.

Weather Data We use weather data from the National Oceanic and Atmospheric Association's (NOAA) Global Historical Climatology Network (GHCN). This data records daily observations of maximum temperature, precipitation, and other weather metrics for roughly 62,000 weather stations in the United States (see Menne et al[4] for more details). In order to construct measures of county level-weather, we begin by filtering out any weather stations that are missing maximum temperature or precipitation measures entirely. We use the geographic coordinates of each weather station, along with shapefiles specifying county borders to determine which weather stations are contained in which counties. For counties that contain three or more weather stations, we simply generate county precipitation and max temperature by its weather stations.

However, out of 3,233 counties in the US, 243 have no weather stations and 967 have fewer than three weather stations. For each of these counties, we assign the nearest three stations within 100 kilometers of the county's centroid. To generate county-level measures, we again just take the average of these assigned weather stations. Though this procedure assigns weather stations to nearly every county, there are still missing values for either pre-

cipitation or max temperature for some county-day pairs. We fill in these missing values by averaging across the nearest 3 stations without missing data. After this procedure, we achieve 99.9% coverage of all county-days in our sample. We provide some visualizations of county-level maximum temperature and precipitation across the United States in Supplementary Figures 1 and 2.

Supplementary Note 2: Difference-in-Difference Models

As mentioned in the main text, our base empirical approach to measuring the causal effects of COVID-19 policies and reopenings is difference-in-differences (DiD), much like several other papers[5–7]. The most basic version of DiD requires multiple observations over time of at two groups, where one of the groups is exposed to some treatment or intervention at some point. DiD works by comparing the change over time in the outcomes for this "treatment" group, relative to the change over time for the "control group." The key assumption of DiD is "parallel trends" or the idea that the trends in outcome variable(s) would have been the same had the treatment not occurred.

Base Model For our analysis, our base model is a commonly used adaptation of the basic DiD model that simply employs unit and time fixed effects that allows for arbitrary or staggered variation in treatment timing across different units[4]. Our base model specification is as follows:

$$\log(Y_{it}) = \delta_{(ip)}D_{it}^{(ip)} + \delta_{(sh)}D_{it}^{(sh)} + \delta_{(ro)}D_{it}^{(ro)} + f(W_{it}) + \alpha_i + \tau_t + \epsilon_{it} \tag{S2}$$

Here, $\log(Y_{it})$ refers one of the our four main mobility outcomes defined in Supplementary Note 1, `log_mcbgv`, `log_pgt2kmt`, `log_pgt1hafh`, and `log_pnchd`, indexed by county i on date t. $D_{it}^{(ip)}$ is a binary indicator that takes the value of 1 once county i's state has adopted some kind of social distancing policy. Similarly, $D_{it}^{(sh)}$ and $D_{it}^{(ro)}$ switch to 1 once i is subject to a shelter-in-pace order or once i's state starts to reopen, respectively. Due to the way these binary indicators are coded (once they switch to 1, they do not switch back to 0), the associated parameters $\delta_{(ip)}$, $\delta_{(sh)}$, and $\delta_{(ro)}$ capture the marginal effects conditional

---

[48] has shown that the estimand of such staggered DiD models decomposes into a weighted average of all possible two-group/two-period DiD estimators in the data.

on the previous policies. More concretely, $\delta_{(ip)}$ captures the average difference in mobility between pre-pandemic levels and while counties have just started implementing social distancing policies; $\delta_{(sh)}$ is then the average difference in mobility during the "stay home" period compared to the $\delta_{(ip)}$; and $\delta_{(ro)}$ is then difference in mobility compared to $\delta_{(ip)} + \delta_{(sh)}$. Put differently, to estimate the differences in mobility from pre-pandemic levels and the reopening, we would need to sum $\delta_{(ip)}$, $\delta_{(sh)}$, and $\delta_{(ro)}$. $f(W_{it})$ is a term that captures the effect of local weather, which may be highly nonlinear, using a "double machine learning" (DML) approach[9]. This procedure is explained in much greater detail in Supplementary Note 6. α_i and τ_t denote a set of county and time fixed effects respectively, while ϵ_{it} represents the error term.

Alter Policy (AP) Model In additional to parallel trends, DiD also assumes the that the stable unit treatment value assumption (SUTVA) holds. Put more plainly, SUTVA simply states that the effect of treatment does not "spillover" to the control groups. However, Holtz et al[10] finds strong evidence for spillover effects in social distancing policy. We extend our base model to account such effects as follows:

$$
\begin{aligned}
\log(Y_{it}) = {} & \delta_{(ip)} D_{it}^{(ip)} + \delta_{(sh)} D_{it}^{(sh)} + \delta_{(ro)} D_{it}^{(ro)} + \\
& \gamma_{(ip)} D_{-it}^{(ip)} + \gamma_{(sh)} D_{-it}^{(sh)} + \gamma_{(ro)} D_{-it}^{(ro)} + f(W_{it}) + \alpha_i + \tau_t + \epsilon_{it}
\end{aligned}
\tag{S3}
$$

The key differences here are the inclusion of $D^{(ip)}_{-it}$, $D^{(sh)}_{-it}$, and $D^{(op)}_{-it}$ which denote the socially weighted average of alter state policy period indicators. More formally:

$$D^{(ip)}_{-it} = \sum_j \omega_{j \to i} D^{(ip)}_{-it}$$

$$D^{(ip)}_{-it} = \sum_j \omega_{j \to i} D^{(ip)}_{-it}$$

$$D^{(ro)}_{-it} = \sum_j \omega_{j \to i} D^{(ro)}_{-it}$$

Since we are focused on differences in state policy, we purposefully set weights $\omega_{j \to i}$ to be equal to 0 if i and j belong to the same state. For counties belonging to different states however, the weights are defined as follows:

$$\omega_{j \to i} = \frac{\texttt{scaled\_sci}_{ij} * n_j}{\sum_k \texttt{scaled\_sci}_{ik} * n_k} : \text{state}_i \neq \text{state}_j, \text{state}_i \neq \text{state}_k$$

where n_j is the 2018 US Census estimated population of county j. Similar to the δ parameters, $\gamma_{(ip)}$, $\gamma_{(sh)}$, and $\gamma_{(ro)}$ are also considered marginal effects. One key difference however is that the alter policy variables $D^{(ip)}_{-it}$, $D^{(sh)}_{-it}$, and $D^{(op)}_{-it}$ are not binary indicators, meaning that each γ is interpreted as the marginal effect if all other states move onto that particular policy period.

Pre-Trends Before moving on to our results, we first show that parallel trends is satisified, at least in the pre-period. To start, in Figure 1A of the main text, it can be seen that the time series of the various state quintile groups by reopening data generally all follow the same trend, especially in the pre-pandemic period from Jan. 1, 2020 to Feb. 29, 2020. In Supplementary Figure 3, we plot the average residuals of our 4 main mobility dependent variables after partialing out county and date fixed effects limited only to the pre-pandemic period of

our data. Looking at the residuals, it is difficult to discern any systematic trend amongst the different groups. In fact, each series looks essentially like a random walk further supporting the assumption of identical pre-trends.

Results As the population sizes of various counties are quite heterogeneous, we estimate Equations 2 and 3 using weighted least squares, where observations are weighted according to their populations. This allows us to interpret our results as averages across human mobility rather than averages across county mobility[5]. The results of estimating Equations 2 and 3 are presented in Supplementary Tables 1 and 2 respectively.

Consistent with previous work, both models indicate significant decreases in mobility during statewide shelter-in-place orders. Specifically, both models' point estimates indicate mobility decreases of 4-6%, relative to the initial policy period (though this basically extends to pre-pandemic levels given that ego state initial policy coefficients are generally not statistically significant and close to 0). However, as we should expect, once a state begins reopening, mobility starts to increase. Again both models' here are quite quantitatively consistent, indicating a 3-5% increases in mobility. Overall, our estimates indicate that once reopened, mobility is slightly depressed compared to pre-pandemic levels (computed by summing the coefficients across all 3 policy periods), by 1-1.5% according to Table 1 and 2-2.5% according to 2[6]. Consistent with Holtz et al[10], our results indicate significant cross-state policy spillovers. Our estimates suggest that once all alter states initiate social distancing, county mobility drops by 4-5%. If all other states implement a shelter-in-place order, mobility will drop even further, by an additional 16-22%. In contrast, alter states reopenings

---

[5]Consider the following example of 2 counties, one with 1000 people and one with 9000 people. Suppose that a shelter-in-place order reduces mobility of county 1 by 10% and county 2 by 20% then the unweighted regression produce a shelter-in-place impact of 15%. In contrast, the weighted regression would produce a shelter-in-place impact of 19% which corresponds to the average decrease in mobility across the population.

[6]The base model estimates are not statistically distinguishable from pre-pandemic levels, but the ASPS estimates are.

have a generally larger but opposite effect, increasing county-level mobility by 20-32%.

Supplementary Note 3: Instrumental Variables Model

In addition to uncovering strong policy spillover effects, Holtz et al[10] showed that the spillovers were largely mediated by endogenous peer behavior. To investigate if such behavior extends to reopenings we estimate the following specification:

$$
\log(Y_{it}) = \beta \log(Y_{-it}) + \delta_{(ip)} D_{it}^{(ip)} + \delta_{(sh)} D_{it}^{(sh)} + \delta_{(ro)} D_{it}^{(ro)} +
$$
$$
\gamma_{(ip)} D_{-it}^{(ip)} + \gamma_{(sh)} D_{-it}^{(sh)} + \gamma_{(ro)} D_{-it}^{(ro)} + f(W_{it}) + \alpha_i + \tau_t + \epsilon_{it}
$$

(S4)

Compared to Equation 3 above, the only difference is the inclusion of $\log(Y_{-it})$ or the log transformed weighted average of alter state's mobility outcomes. As with the alter-policy variables above, $Y_{-it} = \sum_j \omega_{j \to i} Y_{jt}$. However, the DiD framework is theoretically insufficient at producing causal estimates of endogenous peer behavior due to challenges posed by simultaneity (aka the "reflection problem"[11]), correlated exposure to unobserved confounding factors, and homophily[12].

To address this issue, we shift our approach to instrumental variables (IV), an approach widely used in social sciences to address endogeneity concerns. To provide a basic overview of how IV functions, consider the following simple scenario where $Y = \beta_0 + \beta_1 X + \epsilon$, but X is correlated with the error term ϵ. In such a setting, simply regressing Y on X will produce a biased estimate of β_1. A third variable Z can be considered an "instrument" for X if it meaningful impacts X and is (conditionally) uncorrelated with the error term ϵ. These two requirements or restrictions are known as relevancy and exclusion respectively. If Z does indeed qualify as an instrument, consistent estimates of β_1 can be recovered via a 2-stage least squares (2SLS) procedure where X is first regressed on Z in to produce fitted values $\hat{X} = E[X|Z]$. In the second stage, Y is then regressed on these fitted values of X.

In our case, our IV strategy exploits exogenous shocks alters' mobility behavior stemming from variation in alters' weather. Similar weather IV approaches have been used to measure emotional contagion[13], peer effects in exercise behavior[14], and social spillovers in online news consumption[15]. Most relevantly, Holtz et al[10] also leveraged weather instruments to estimate the impact of endogenous peer behavior.

Weather Instruments and First Stage We construct our instruments from the county-level weather dataset that we constructed described in Supplementary Note 1. In order to construct our alter county instruments, we first we first construct a sequence of county-level indicator variables that take a value of 1 if the amount of rainfall in county i on date t falls within or exceeds a specific precipitation decile, conditional on non-zero precipitation[7]. We generate a similar sequence of indicator variables for maximum temperature as well. To avoid perfect multicolinearity, we remove only first max temperature decile indicator. We need not remove the first precipitation decile as these deciles are computed only for non-zero precipiation, meaning that "no precipitation" functions as the base case. We then construct 19 alter-state weather measures again by taking the socially weighted averages of of each of 10 preciptation and 9 max temperature deciles[8] to form the alter state weather instruments ($W_{-it} = V_{-it}^{\text{prcp},1}, ..., Q_{-it}^{\text{prcp},10}, Q_{-it}^{\text{tmax},2}, ..., Q_{-it}^{\text{tmax},10}$). This leads to the following first-stage specification:

$$\log(Y_{-it}) = \delta_{(ip)}^{fs} D_{it}^{(ip)} + \delta_{(sh)}^{fs} D_{it}^{(sh)} + \delta_{(ro)}^{fs} D_{it}^{(ro)} + \gamma_{(ip)}^{fs} D_{-it}^{(ip)} + \gamma_{(sh)}^{fs} D_{-it}^{(sh)} + \gamma_{(ro)}^{fs} D_{-it}^{(ro)} +$$
$$\sum_{d=1}^{10} \left(\zeta_d^{\text{prcp}} Q_{-it}^{\text{prcp},d} + \zeta_d^{\text{tmax}} Q_{-it}^{\text{tmax},d} \right) + \alpha_{-i} + \tau_t + \nu_{-it} \tag{S5}$$

---

[7] For example $Q_{it}^{\text{prcp},1} = \mathbf{1}(\text{prcp}_{it} > q) : q = \arg_x Pr(\text{prcp}_{it} \geq x) = 0$, $Q_{it}^{\text{prcp},2} = \mathbf{1}(\text{prcp}_{it} \geq q) : q = \arg_x Pr(\text{prcp}_{it} \geq x) = 0.5$, etc. It is also worth noting that this construction means that if $V_{it}^{\text{prcp},k} = 1$, then $V_{it}^{\text{prcp},j} = 1 : j < k$.

[8] More formally: $Q_{-it}^{\text{prcp},k} = \sum_j \omega_{j \to i} * Q_{jt}^{\text{prcp},k}$ and $Q_{-it}^{\text{tmax},k} = \sum_j w_i j * Q_{jt}^{\text{tmax},k}$.

In theory, we should be able to use these alter-state weather variables to instrument for alter-state peer behavior, However, common major concern with weather instruments is that geographically proximate locations tend to have similar weather. Theoretically, this should not pose an issue: even if "alters' weather" is highly correlated with "own weather," it should be conditionally ignorable so long as the effects of "own weather" are controlled for. However, this can be quite challenging due to the potential nonlinearities and interactions in the impact of weather. For instance, the likelihood of going outside is going to change much more going from 0mm to 1mm of rain relative to going from 20mm to 21mm. In a similar vein, the impact of rain is likely to be very different if it is a comfortable day outside than if it is cold and dreary. Such complexities may therefore cause a "technical" violation of conditional ignorability since alters' weather may by providing additional information about own weather that cannot be captured linearly. As such, we adopt a flexible DML procedure to model the impact of weather that we explain in greater detail in Supplementary Note 6.

Results The results of estimating Equation 4 can be found in Table 3. Included in this table are the first-stage F-statistics testing the relevancy of the instruments. As can be seen in the table, the F-stats range from 60-70 indicating that we do not have a weak instruments problem.

Across our four main outcomes, several clear trends emerge. First, Ego State shelter-in-place still has a statistically significant negative effect on mobility. While these estimates are smaller in magnitude across the board, they are not quite statistically distinguishable from the estimates found in Table 3. Second, again consistent with Holtz et al[10], we find strong evidence of endogenous peer effects. Our estimates indicate that a 1% increase or drop in mobility by all peers in different states will cause mobility in an ego county to increase or drop by approximately 2% using the 2SLS estimate. As with Holtz et al[10], once

we introduce endogenous peer behavior into our model, the alter state policy coefficients move significantly closer to 0 and most are no longer statistically significant. Such results seem to confirm that the effects of alter state policies are being mediated by peer behavior in those states.

Supplementary Note 4: Dyadic Travel Difference-in-differences

In this section, we explore the potential impacts of both origin and destination policies on cross-state travel. As our policy variation is at the state level, we specifically focus on cross-state county-pairs. Similar to Supplementary Note 2, our analysis is based on a difference-in-differences approach, where we consider origin and destination policies as orthogonal treatments.

Model Specifications Here we use the following 2 model specifications:

$$\log(Y_{o \to d,t}) = \sum_m \lambda_m D_{ot}^m + \sum_n \psi_n D_{dt}^n + \alpha_{o \to d} + \tau_t + \epsilon_{o \to d,t} \tag{S6}$$

$$\log(Y_{o \to d,t}) = \sum_m \lambda_m D_{ot}^m + \sum_n \psi_n D_{dt}^n + \sum_m \sum_n \pi_{m,n}(D_{ot}^m * D_{dt}^n) + \alpha_{o \to d} + \tau_t + \epsilon_{o \to d,t} \tag{S7}$$

$\log(Y_{o \to d,t})$ refers to the one of our log-transformed cross-county mobility metrics (`log_ndotd`, `log_ptotd` described in Supplementary Note 1 above) based on the number of devices identified with a home in an origin county o and stopping for at least one minute in a destination county d. D_{ot}^m and D_{dt}^n denote origin and destination policies respectively, where $m, n \in \{(ip), (sh), (ro)\}$. As with Equation 2 above, these policy are binary indicators that flip to 1 once the corresponding policy period begins. This means that for Equation 6 the associated parameters $\lambda_{(ip)}$, $\lambda_{(sh)}$, $\lambda_{(ro)}$, $\gamma_{(ip)}$, $\gamma_{(sh)}$, and $\gamma_{(ro)}$ are interpreted as the marginal effect, but only within each parameter family[9].

In Equation 7, which models all potential interactions between origin and destination policies, the λs capture the marginal effects of origin policy if the destination is in the pre-

---

[9]That is to say each successive λ is marginal to only the previous λ's and each successive γ is marginal to the previous γs

policy period. Likewise, the γs capture the marginal effects of destination policy if the origin is in the pre-policy period. The interaction parameters $\pi_{m,n}$ are then the additional marginal effect above and beyond the sum of preceding parameters. That is, $\pi_{(sh),(rop)}$ is the additional marginal effect when compared to $\lambda_{(ip)} + \lambda_{(sh)} + \gamma_{(ip)} + \gamma_{(sh)} + \gamma_{(ro)} + \pi_{(ip),(ip)} + \pi_{(ip),(sh)} + \pi_{(ip),(ro)} + \pi_{(sh),(ip)}, \pi_{(sh),(sh)}$. Lastly, $\alpha_{o \to d}$ and τ_t denote directed dyad and time fixed effects, and $\epsilon_{o \to d,t}$ captures the error term.

Pre Trends As with Supplementary Note 2, our analysis here is based on a difference-in-differences approach. Naturally, this means that it is important to verify that there aren't any systematic differences in pre-trends. In Supplementary Figure 4, we plot the average residuals of our 2 cross-state mobility variables after partialing out dyad and date fixed effects for the period between Jan. 1, 2020 and Feb. 29, 2020. as with above, it is difficult to find any systematic trend amongst the different groups organized around destination county's statewide reopening start. Again, each series looks a like a mean 0 random walk suggesting that parallel trends does indeed hold.

Results The results from estimating Equations 6 and 7 are displayed in Supplementary Tables 4 and 5. While `log_ndotd` is estimated using OLS, `log_pdotd` is estimated using WLS where weights are proportional to origin county population.

Looking at Table 4, we generally see that the impacts of origin policy are generally statistically significant. However both destination closures and reopenings have major effects, with closures decreasing cross-state travel by 13-18% from distant counties and reopenings increasing travel by 12-13% from both nearby and distant counties. Evaluating all the interactions effects is rather difficult, but after appropriately summing the coefficients we find that destination policies did not have a meaningful effect on travel while origin counties were in their initial policy period. However, once an origin starts sheltering-in-place, travel

to distant counties decreased by 10% while travel to nearby counties increased by 52-65%, if those counties had not yet implemented a shelter-in-place order. Once destinations implemented shelter-in-place orders, distant cross-state travel decreased by 14-16%, with no detectable effect on nearby cross-state travel, conditional on the origin also sheltering-in-place. Once the destinations was reopened (but the origin was still locked down), travel to destinations increased by 11-12% from nearby counties and 24% from distant counties. In contrast, if the origin had reopened, travel to both nearby and distant locked down destination counties was decreased by 9-17% and 21-27% respectively. Once the destination reopened, conditional on the origin already being reopened, there was a travel from distant origins increased by 14-19%.

Supplementary Note 5: Robustness

To further support the results presented in the main text, we run a series of robustness checks. To address concerns about the representativeness of our mobility data, we first begin by replicating our results using mobility data provided by Facebook. We also explore the potential anticipation or lagging effects of policy with regards to mobility behavior. Lastly, we provide additional validation of our statistical inference using Fischerian Randomization Inference (FRI).

Replication with Facebook Data   Here we make use of Facebook's Data for Good Initiative publicly available movement range maps[10] which provide differentially private, daily, county-level measures of mobility from Mar 1, 2020 onwards. Specifically, we make use of: `bing_tiles_visited_relative_change` (btvrc), which captures the relative change[11] in the number of "bing tiles" (0.6km $\times$ 0.6km blocks) visited by Facebook users in a given county on a given day, and `ratio_single_tile_users`, which is simply the fraction of Facebook users who are recorded staying within a single bing tile for that entire day. To make this measure jibe with our other metrics, we subtract it from one to construct `ratio_not_single_tile_users` (rnstu) and log transform it. Here, we restrict our data the 2368 counties and 122 dates (March 1, 2020 to June 30, 2020) found in both the Facebook and Safegraph datasets. Results from re-estimating Equations 2 and 3 are in Supplementary Figure 5. The estimates produced using Facebook mobility measures are both qualitatively and quantitatively similar to those produced using the Safegraph data, providing support for the generalizability of our results.

---

[10]https://dataforgood.fb.com/tools/movement-range-maps/
[11]The baseline for this metric is constructed by averaging across the number of bing tiles visited in February across each day of the week, excluding President's Day.

Lagging and Leading Policy Effects As our policy measures are relatively blunt, we explore whether there may be any leading or lagging effects of transitioning to different periods. Here we re-estimate Equations 3 and 6 with the addition of 5 lagging and leading terms for each policy parameters, plotted in Supplementary Figures 6 and 7 respectively. Note that the leads and and lags are coded marginally so meaning that estimates are additive in nature.

Fisherian Randomization Inference As a robustness check, we use a Fisherian randomization inference[16–19] (FRI) procedure to estimate the null distributions of different policy coefficients for both Equations 3 and 6. We form these distributions by re-sampling our data and shuffling the state policy assignments with each draw. Specifically, each state is randomly assigned the policy vectors of a different state, without replacement. Overall, we repeat this procedure 500 times and we plot the null distributions of the policy parameters estimated using Equations 3 and 6 in Supplementary Figures 8 and 9 respectively. The estimates produced by our real data are represented as black lines. For coefficients originally found to be statistically significant in our main analysis, we see that these generally take extreme values in the null distributions.

Supplementary Note 6: Double Machine Learning Weather Controls

To flexibly control for the impact of weather, we employ a "double machine learning" (DML) procedure[9]. This approach is designed to estimate and draw inferences on a low-dimensional parameter in the presence of high-dimensional nuisance parameters. Consider the following "canonical example" from Chernozhukov et al[9] which we reproduce here:

$$Y = D\theta_0 + g_0(Z) + U, \quad \mathbb{E}[U|D, Z] = 0$$

$$D = m_0(Z) + V, \quad \mathbb{E}[V|Z] = 0$$

Y denotes the outcome, D is a policy or treatment variable, θ_0 is the low-dimensional parameter of interest, Z is a high-dimensional vector of covariates ($g_0(Z)$ can be considered to be the high-dimensional nuisance parameter), and U and V are the errors. The basic intuition behind DML is that $g_0(.)$ and $m_0(.)$ can be estimated using non-parametric statistical methods (aka machine learning) and then "partialed out[20]" from both Y and D. Then one simply regresses the residuals of the dependent variable on the residuals of the treatment variable in order to estimate θ_0. In order to provide guarantees that key moment conditions are satisfied, the machine learning predictions needs to be orthogonalized which can be achieved via sample splitting. As such, the general double ML algorithm is as follows:

1. Split the dataset into K equal size partitions or "folds." Let $F_k, F_k^c : k \in 1, ..., K$ denote each fold and its complement.[12]

2. Estimate g_0 and m_0 with some non-parametric statistical model of choice using only the observations in B_1^c

---

[12]Suppose a dataset has 100 observations and is split into 5 block. B_1 consists of observations 1-20 and $F_1^c := F_2, F_3, F_4, F_5$ consists of the remaining observations 21-80.

3. Form residuals $\tilde{Y} := Y - \hat{g}_0(Z)$ and $\tilde{D} := D - \hat{m}_0(Z)$ only on observations in F_1.

4. Regress $\tilde{Y}$ on $\tilde{D}$ to obtain an estimate of θ_0. Overall, this estimate can be thought of as function of F_1 and F_1^c: $\hat{\theta}_0(F_1, F_1^c)$.

5. Repeat steps 2-4 for the the remaining $K - 1$ folds

6. Form the final estimate of θ_0 by averaging across all estimates: $\hat{\theta}_0^* = \frac{1}{K} \sum_k \hat{\theta}_0(F_k, F_k^c)$

In our case, we consider a county's own weather to be the high-dimensional nuisance parameter, as we are not principally interested in identifying the effect of own weather on social distancing behavior. We use gradient boosted decision trees via XGBoost[21], a state-of-art machine learning algorithm, to estimate $f(.)$ in Equations 2, 3, and 4 as well as the effect of weather on any of the other variables included in our models. XGBoost is an ensemble method that works by fitting a series of forward stage-wise decision trees aimed to minimizing a specified loss function. To give a general idea of the basic procedure:

1. Fit an initial decision tree T_1 that minimizes $E\left[(Y - T_1(X))^2\right]$, where Y is the outcome and X are the covariates or features.

2. Each successive tree is then fitted on the residuals of the previous state[13]:
$T_n = \underset{T}{\arg\min}\ \mathbb{E}[(Y - \sum_{n-1}^{i=1} T_i(X) - T(X))^2]$

In order to prevent overfitting, this iterative process is stopped once out-of-sample predictive performance starts to decline.

As with many other machine learning algorithms, there are a number of hyperparameters that control this estimation procedure of XGBoost. In particular, we adjust:

---

[13]To be more precise, the degree to which each successive tree contributes to the ensemble can be controlled via tuning hyperparameter called a learning rate. We provide a little bit more detail on this below.

- **`tree_depth`**: Controls the depth that each tree-based model is allowed to grow to. The deeper the tree, the more complex the model.

- **`eta`**: Controls the "learning rate" or step size of each model. One way to think of this parameter is as a form of regularization on each model step in order to prevent over-fitting.

- **`nrounds`**: The maximum number of stages the fitting process is allowed to continue on for.

We fix **`tree_depth`** to 2 and **`eta`** = 0.5, but allow **`nrounds`** to run up to a maximum of 100. Then, for each individual variable, the optimal number of rounds (given our choice of **`tree_depth`** and **`eta`**) is determined via a cross-validation procedure for each variable individually[14]. Once the optimal **`nrounds`** is determined, we form the residuals for all our dependent variables and covariates by first partialing out the set of fixed effects and then following the DML approach described above.

---

[14]We note that it would be more optimal to do an exhaustive grid search across the entire hyperparameter space for each individual variable that needs to have the effect of weather partialed out. However, such a grid search would be extremely computationally expensive and would only yield very minor improvements in predictive accuracy.

Supplementary Note 7: Software

Data processing, analysis, and plotting was conducted in R[22] and Python[23]. `pandas`[24], `jsonlite`[25], and various `tidyverse` libraries[26]—`dplyr`, `lubridate`, `readr`, `stringr`, `tidyr`, etc.—were used to process and prepare the data for analysis. Regression analysis was performed using `lfe`[27] and our DML approach relied on `xgboost`[28]. `doMC`[29] was used to parallelize computation. Tables were created using the `stargazer` package[30]. Plots were generated using `ggplot2`[31], `viridis`, `ggsci`, and `urbnmapr`.
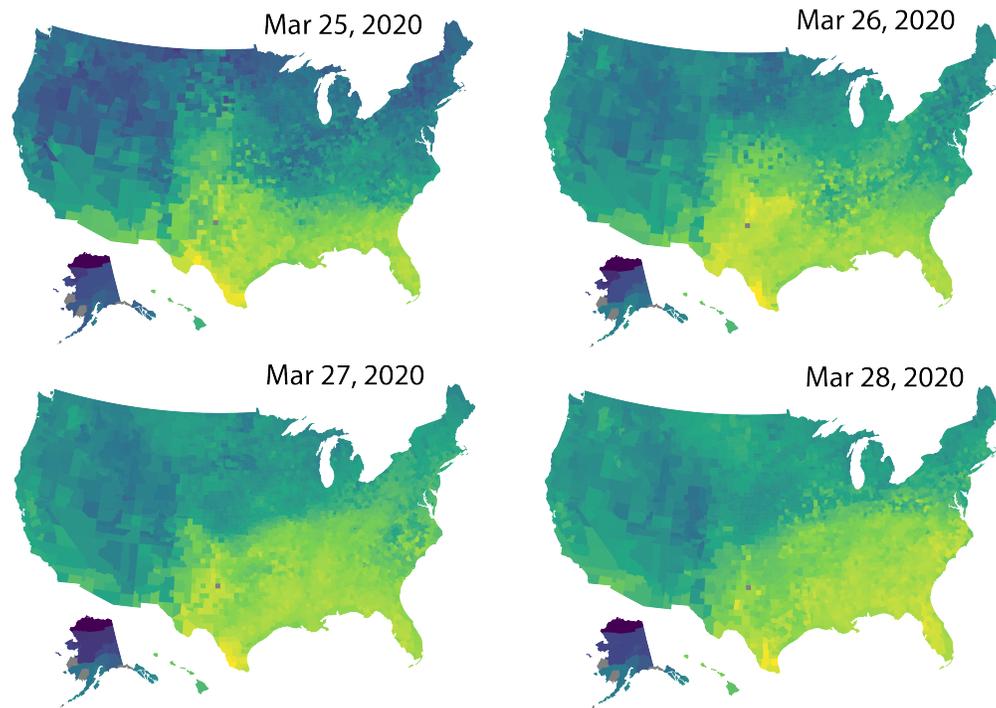
Supplementary Figure 1: The maximum daily temperature (in degrees Celsius) at the county level over four consecutive days. The brighter color indicates higher maximum temperature.
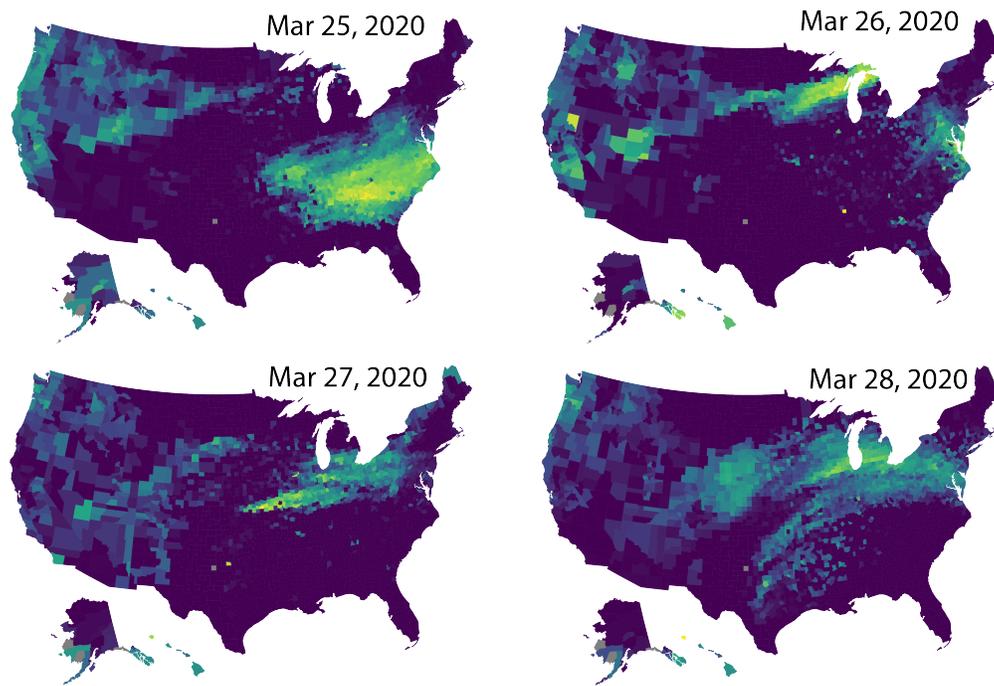
Supplementary Figure 2: Daily precipitation (in millimeters) at the county level over four consecutive days. The brighter color indicates higher precipitation.

Supplementary Figure 3: The average residuals across counties grouped by state-level re-opening date quintiles after partialing out county and date fixed effects from Jan 1, 2020. to Feb. 29, 2020.
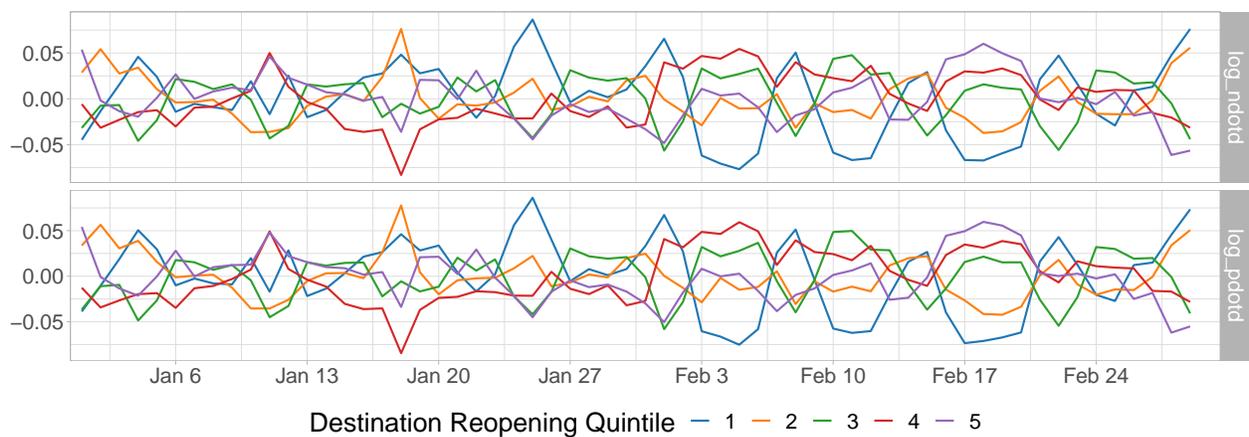
Supplementary Figure 4: The average residuals across dyads grouped by destination state-level reopening date quintiles after partialing out dyad and date fixed effects from Jan 1, 2020. to Feb. 29, 2020.
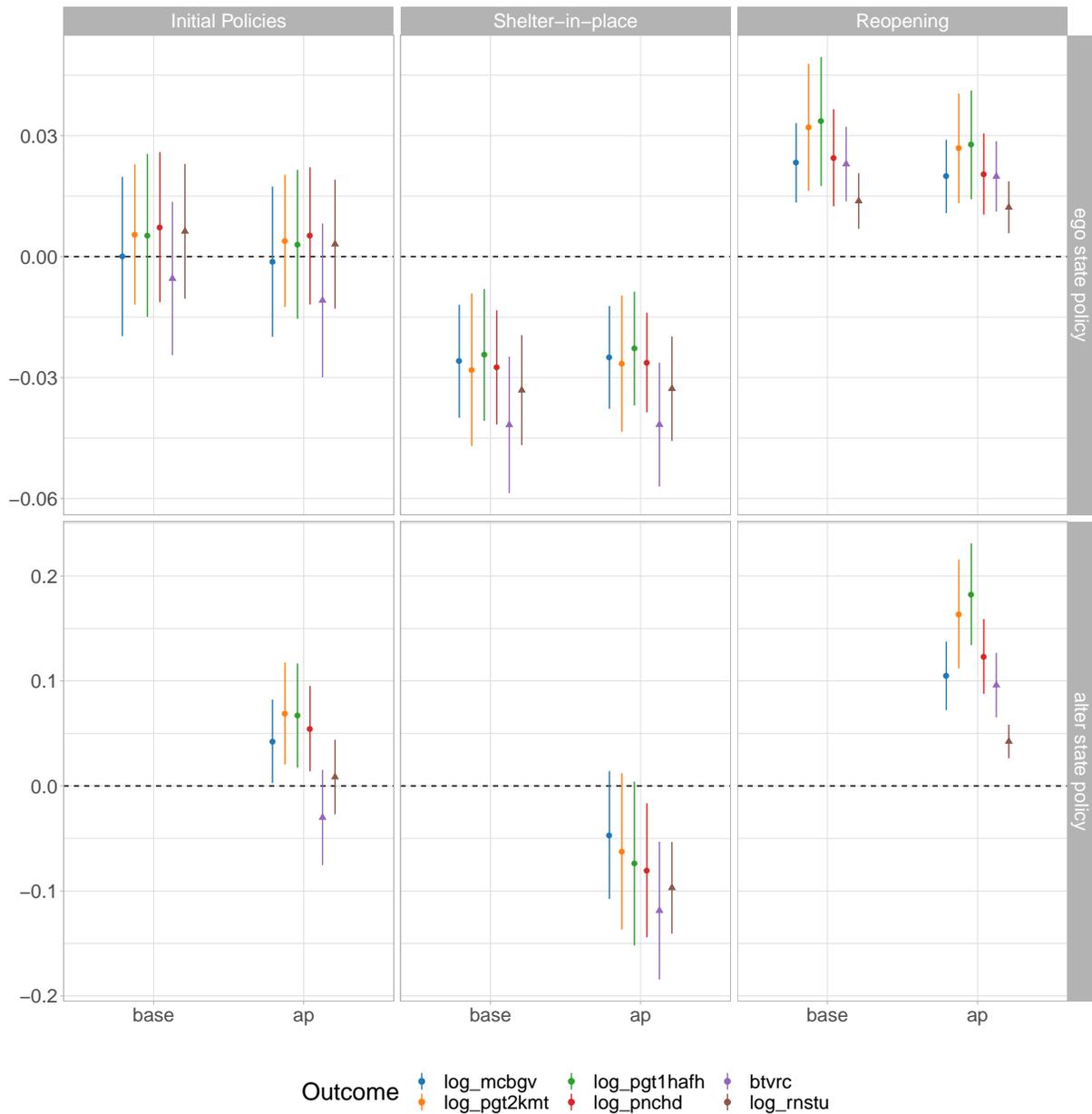
Supplementary Figure 5: Policy coefficients produced using Equations 2 and 3 estimated with WLS using county population as weights across 7 different mobility measures: log_mcgbv (blue), log_pgt2kmt (orange), log_pgt1hafh (green), log_pnchd (red), btvrc (purple), and log_rnstu (brown). Circular points correspond to Safegraph mobility measures and triangular points correspond to Facebook mobility measures.
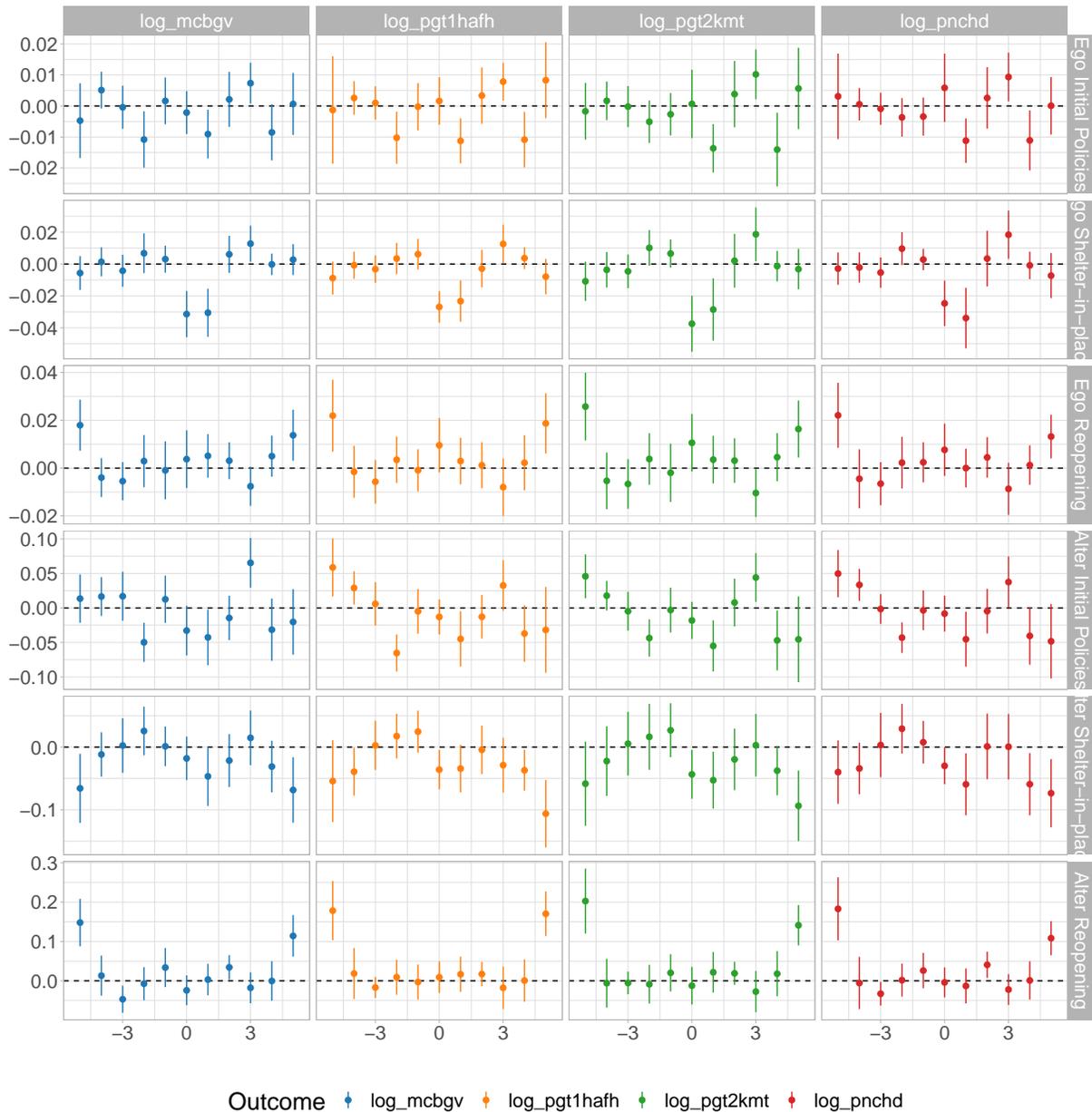
Supplementary Figure 6: Leading and Lagging Effects of different policy periods estimated using Eq. 3 with 5 leading and lagging terms for each both ego and alter state policy variables. Along the X-axis, negative values correspond to leading effects while positive values correspond to lagging effects (leading effects come before the policy, while lagging effects come after). Different rows of plots refer to different policy coefficients, while different columns refer to different mobility outcomes.
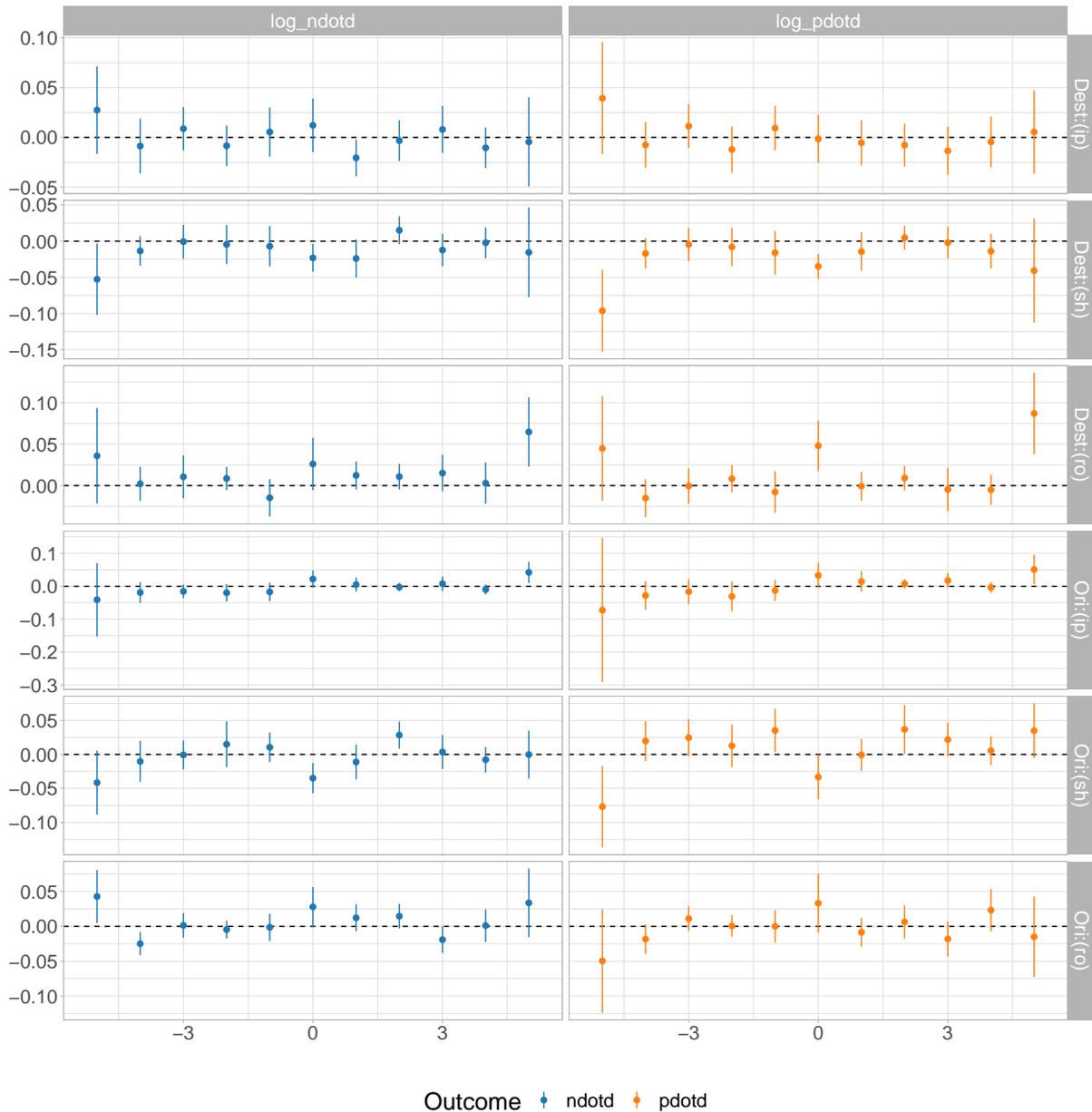
Supplementary Figure 7: Leading and Lagging Effects of different policy periods estimated using Eq. 6 with 5 leading and lagging terms for each both ego and alter state policy variables. Along the X-axis, negative values correspond to leading effects while positive values correspond to lagging effects (leading effects come before the policy, while lagging effects come after). Different rows of plots refer to different policy coefficients, while different columns refer to different mobility outcomes.
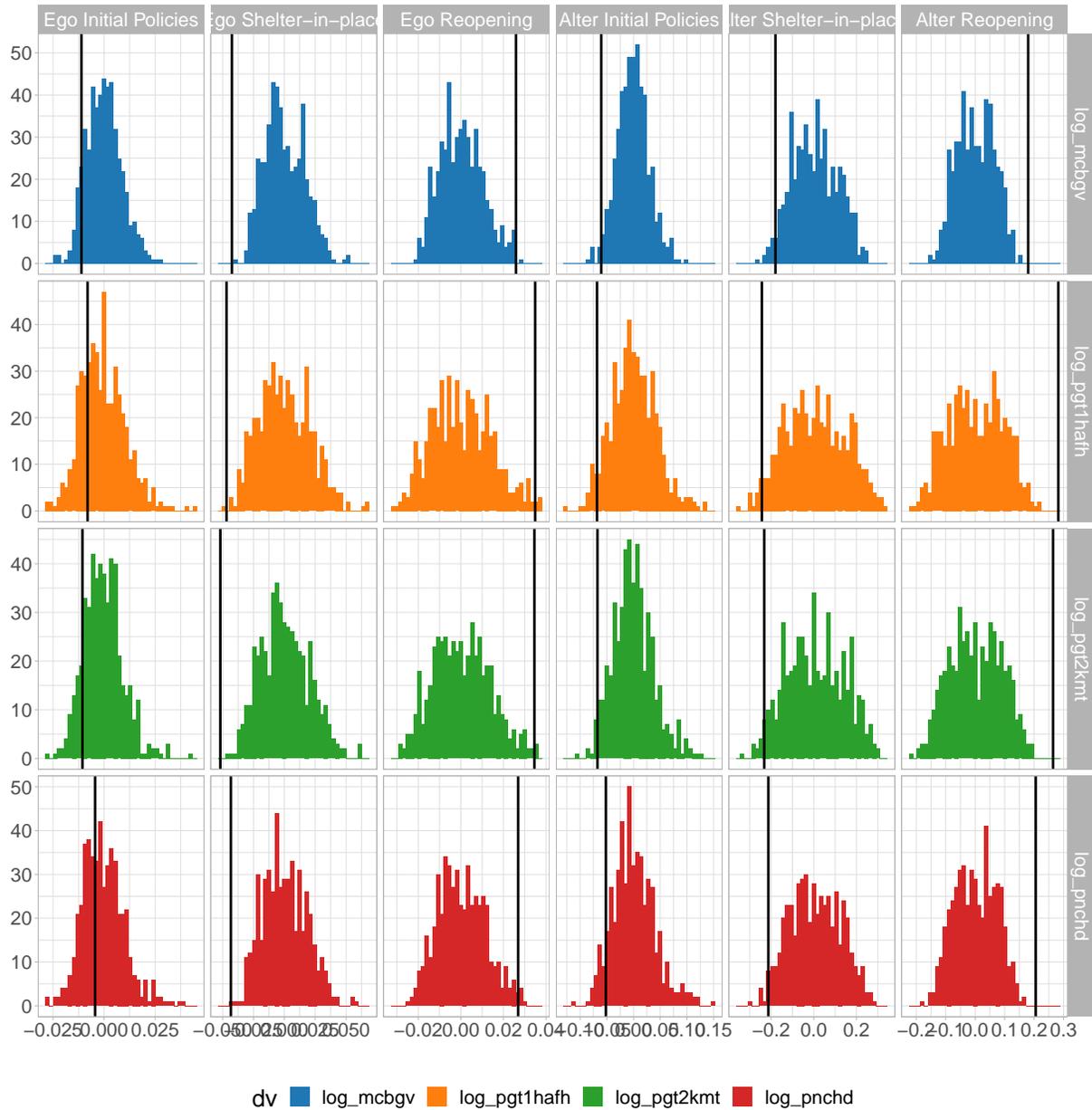
Supplementary Figure 8: Null distributions of the policy parameters (columns) in Eq. 3 across different mobility outcomes (rows). Original estimates represented by black vertical lines.
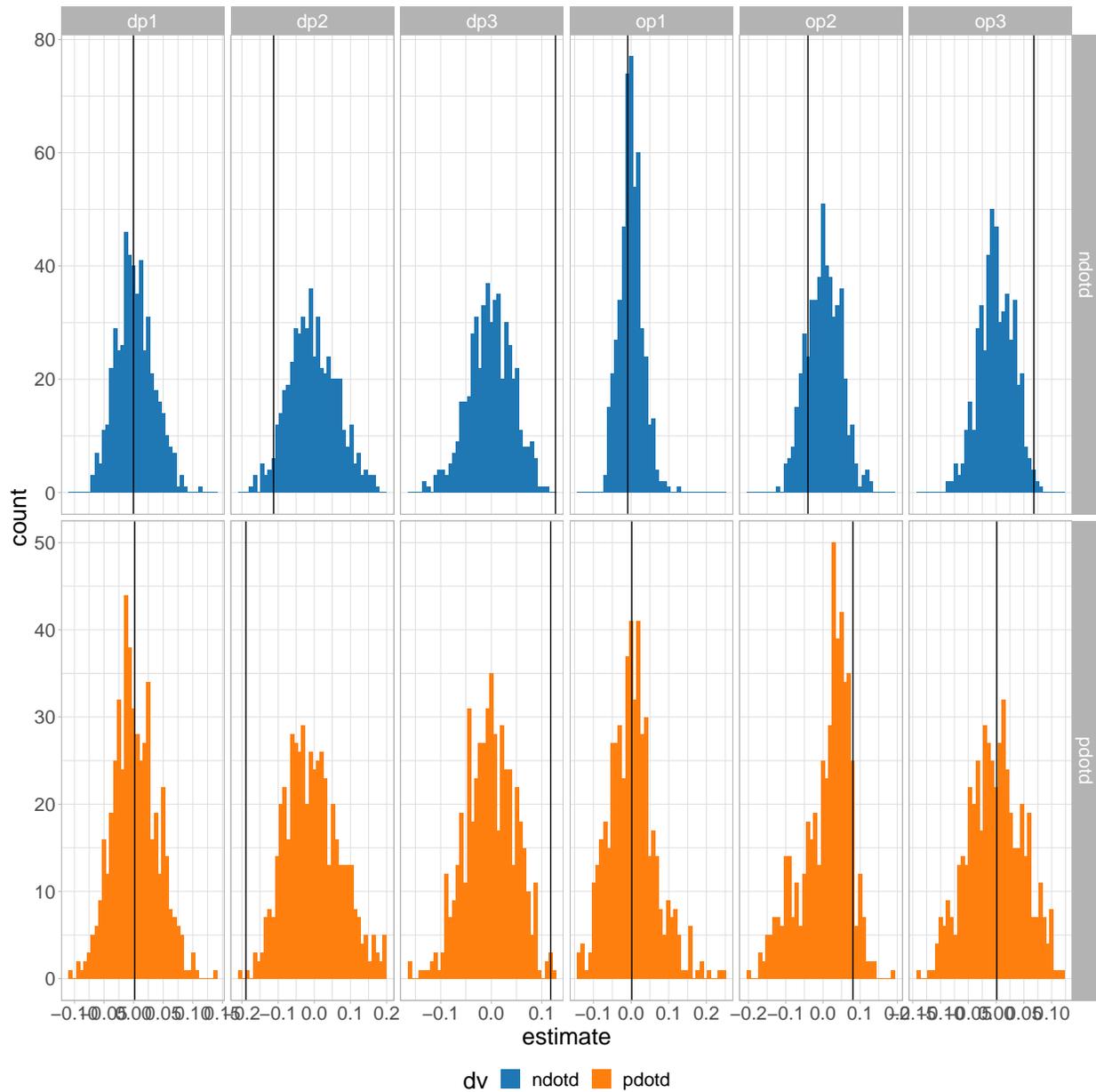
Supplementary Figure 9: Null distributions of both origin and destination policy parameters (columns) in Eq. 6 across different mobility outcomes (rows). Original estimates represented by black vertical lines.

Supplementary Tables

Supplementary Table 1: Base Model Results

	Dependent variable:			
	log_mcbgv	log_pgt2kmt	log_pgt1hafh	log_pnchd
	(1)	(2)	(3)	(4)
Ego State Initial Policies	−0.005	−0.002	0.001	0.003
	(0.006)	(0.007)	(0.006)	(0.006)
Ego State Shelter-in-place	−0.048***	−0.059***	−0.055***	−0.050***
	(0.008)	(0.011)	(0.011)	(0.009)
Ego State Reopening	0.034***	0.046***	0.047***	0.036***
	(0.007)	(0.011)	(0.011)	(0.009)
Observations	470,106	470,106	470,106	470,106
R^2	0.049	0.043	0.043	0.048
Adjusted R^2	0.043	0.037	0.037	0.042

Note: State-Clustered Standard Errors are reported. *$p<0.1$; **$p<0.05$; ***$p<0.01$

Supplementary Table 2: Alters' Policies Model Results

	Dependent variable:			
	log_mcbgv	log_pgt2kmt	log_pgt1hafh	log_pnchd
	(1)	(2)	(3)	(4)
Ego State Initial Policies	−0.011**	−0.011	−0.008	−0.004
	(0.005)	(0.006)	(0.006)	(0.006)
Ego State Shelter-in-place	−0.043***	−0.052***	−0.047***	−0.043***
	(0.007)	(0.009)	(0.009)	(0.007)
Ego State Reopenings	0.026***	0.035***	0.035***	0.027***
	(0.007)	(0.010)	(0.010)	(0.008)
Alter States Initial Policies	−0.061***	−0.068**	−0.068**	−0.052**
	(0.018)	(0.028)	(0.028)	(0.023)
Alter States Shelter-in-place	−0.179***	−0.232***	−0.243***	−0.213***
	(0.046)	(0.057)	(0.056)	(0.048)
Alter States Reopenings	0.180***	0.264***	0.282***	0.206***
	(0.028)	(0.042)	(0.040)	(0.032)
Observations	470,106	470,106	470,106	470,106
R^2	0.087	0.085	0.090	0.095
Adjusted R^2	0.082	0.080	0.085	0.090

Note: State-Clustered Standard Errors are reported. *p<0.1; **p<0.05; ***p<0.01

Supplementary Table 3: IV 2SLS Results

	Dependent variable:			
	log_mcbgv	log_pgt2kmt	log_pgt1hafh	log_pnchd
	(1)	(2)	(3)	(4)
Ego State Initial Policies	−0.005	−0.002	0.0003	0.002
	(0.006)	(0.008)	(0.007)	(0.006)
Ego State Shelter-in-Place	−0.032***	−0.036***	−0.030***	−0.028***
	(0.005)	(0.007)	(0.006)	(0.005)
Ego State Reopening	0.014**	0.014	0.013	0.012*
	(0.007)	(0.010)	(0.010)	(0.007)
Alter States Initial Policies	−0.037**	−0.032	−0.041*	−0.037*
	(0.018)	(0.025)	(0.023)	(0.020)
Alter States Shelter-in-Place	0.018	0.045	0.048	0.029
	(0.025)	(0.035)	(0.036)	(0.027)
Alter States Reopening	0.039**	0.031	0.031	0.029
	(0.018)	(0.033)	(0.030)	(0.023)
Endogenous Alter States Behavior	1.835***	2.228***	2.193***	2.073***
	(0.190)	(0.235)	(0.217)	(0.193)
First-Stage F	71.879	60.895	62.580	72.430
Observations	470,106	470,106	470,106	470,106
R^2	0.476	0.402	0.467	0.492
Adjusted R^2	0.473	0.398	0.464	0.489

Note: State-Clustered Standard Errors are reported. *$p<0.1$; **$p<0.05$; ***$p<0.01$

Supplementary Table 4: Dyadic Travel Results

	Dependent variable:					
	log_ndotd			log_pdotd		
	(1) All	(2) Nearby	(3) Distant	(4) All	(5) Nearby	(6) Distant
Origin Initial Policies	−0.009	0.012	−0.016	0.002	0.085	0.0005
	(0.027)	(0.025)	(0.035)	(0.034)	(0.051)	(0.035)
Origin Shelter-in-Place	−0.040	−0.117***	−0.012	0.081***	−0.137***	0.090***
	(0.026)	(0.034)	(0.026)	(0.029)	(0.045)	(0.029)
Origin Reopening	0.068*	0.106***	0.065	0.001	0.090**	0.002
	(0.040)	(0.036)	(0.044)	(0.055)	(0.044)	(0.061)
Destination Initial Policies	−0.0003	0.010	−0.004	0.002	0.036	0.002
	(0.028)	(0.020)	(0.029)	(0.030)	(0.030)	(0.031)
Destination Shelter-in-Place	−0.112**	−0.095***	−0.110*	−0.189***	−0.061	−0.193***
	(0.051)	(0.032)	(0.057)	(0.045)	(0.046)	(0.045)
Destination Reopening	0.127***	0.165***	0.130***	0.118***	0.185***	0.124***
	(0.033)	(0.025)	(0.032)	(0.040)	(0.029)	(0.042)
Observations	3,107,468	708,708	2,398,760	3,107,468	708,708	2,398,760
R^2	0.764	0.905	0.664	0.853	0.928	0.819
Adjusted R^2	0.762	0.905	0.662	0.853	0.927	0.818

Note: 2-way Origin State and Destination State Clustered Standard Errors are reported. *p<0.1; **p<0.05; ***p<0.01

Supplementary Table 5: Dyadic Travel With Interactions

	Dependent variable:					
	log_ndotd			log_pdotd		
	(1) All	(2) Nearby	(3) Distant	(4) All	(5) Nearby	(6) Distant
Origin Pre-Policies × Destination Initial Policies	0.038 (0.034)	0.034 (0.026)	0.032 (0.034)	0.055* (0.030)	0.053 (0.034)	0.055* (0.030)
Origin Pre-Policies × Destination Shelter-in-Place	−0.135* (0.070)	−0.059 (0.042)	−0.140 (0.085)	−0.201*** (0.030)	−0.132** (0.050)	−0.196*** (0.035)
Origin Pre-Policies × Destination Reopening	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Origin Initial Policies × Destination Pre-Policies	0.024 (0.029)	0.041 (0.027)	0.013 (0.039)	0.058 (0.042)	0.100* (0.053)	0.058 (0.044)
Origin Initial Policies × Destination Initial Policies	−0.072** (0.027)	−0.054** (0.026)	−0.067** (0.030)	−0.106*** (0.026)	−0.048 (0.031)	−0.108*** (0.028)
Origin Initial Policies × Destination Shelter-in-Place	0.119** (0.054)	0.024 (0.058)	0.141** (0.070)	0.148*** (0.041)	0.121** (0.049)	0.143*** (0.043)
Origin Initial Policies × Destination Reopening	0.068 (0.052)	0.154*** (0.046)	0.062 (0.050)	0.141** (0.068)	0.267*** (0.084)	0.147** (0.071)
Origin Shelter-in-Place × Destination Pre-Policies	−0.006 (0.035)	−0.364*** (0.115)	0.058 (0.043)	0.192*** (0.067)	−0.399*** (0.075)	0.193*** (0.071)
Origin Shelter-in-Place × Destination Initial Policies	0.004 (0.041)	0.290*** (0.102)	−0.029 (0.043)	−0.057 (0.058)	0.293*** (0.100)	−0.048 (0.059)
Origin Shelter-in-Place × Destination Shelter-in-Place	−0.094** (0.038)	−0.110** (0.045)	−0.098** (0.037)	−0.111*** (0.036)	−0.060 (0.050)	−0.111*** (0.039)
Origin Shelter-in-Place × Destination Reopening	0.064 (0.041)	0.070** (0.034)	0.065 (0.042)	−0.032 (0.053)	0.003 (0.061)	−0.032 (0.054)
Origin Reopening × Destination Pre-Policies	0.175** (0.069)	0.143** (0.058)	0.198** (0.079)	0.124 (0.101)	0.256** (0.122)	0.129 (0.109)
Origin Reopening × Destination Initial Policies	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Origin Reopening × Destination Shelter-in-Place	−0.110** (0.043)	0.019 (0.039)	−0.147*** (0.048)	−0.142** (0.067)	−0.075 (0.103)	−0.147** (0.069)
Origin Reopening × Destination Reopening	−0.010 (0.045)	−0.097** (0.038)	0.003 (0.052)	0.008 (0.042)	−0.164** (0.062)	0.009 (0.045)
Observations	3,107,468	708,708	2,398,760	3,107,468	708,708	2,398,760
R²	0.764	0.905	0.665	0.854	0.928	0.819
Adjusted R²	0.763	0.905	0.663	0.853	0.928	0.818

Note: 2-way Origin State and Destination State Clustered Standard Errors are reported. *p<0.1; **p<0.05; ***p<0.01

References

1. Dwork, C., Roth, A. *et al.* The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* **9**, 211–407 (2014).

2. Raifman, J. *et al.* Covid-19 us state policy database (2020).

3. Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives* **32**, 259–80 (2018).

4. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *Journal of atmospheric and oceanic technology* **29**, 897–910 (2012).

5. Allcott, H. *et al.* Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. Tech. Rep. 26946, National Bureau of Economic Research (2020).

6. Painter, M. & Qiu, T. Political beliefs affect compliance with COVID-19 social distancing orders. Tech. Rep. (2020). Available at SSRN 3569098.

7. Chiou, L. & Tucker, C. Social distancing, internet access and inequality. Tech. Rep. 26982, National Bureau of Economic Research (2020).

8. Goodman-Bacon, A. Difference-in-differences with variation in treatment timing. Tech. Rep., National Bureau of Economic Research (2018).

9. Chernozhukov, V. *et al.* Double/debiased machine learning for treatment and structural parameters (2018).

10. Holtz, D. *et al.* Interdependence and the cost of uncoordinated responses to covid-19. *Proceedings of the National Academy of Sciences* (2020). URL https://www.pnas.org/content/early/2020/07/29/2009522117. https://www.pnas.org/content/early/2020/07/29/2009522117.full.pdf.

11. Manski, C. F. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* **60**, 531–542 (1993).

12. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: homophily in social networks. *Annual review of sociology* **27**, 415–444 (2001).

13. Coviello, L. *et al.* Detecting emotional contagion in massive social networks. *PloS one* **9** (2014).

14. Aral, S. & Nicolaides, C. Exercise contagion in a global social network. *Nature communications* **8**, 1–8 (2017).

15. Aral, S. & Zhao, M. Social spillovers in online news consumption. *Available at SSRN 3328864* (2019).

16. Fisher, R. A. *The Design of Experiments* (Oliver and Boyd, Edinburgh, 1935).

17. Imbens, G. W. & Rubin, D. B. *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge University Press, 2015).

18. MacKinnon, J. G. & Webb, M. D. Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics* (2020).

19. Bowers, J., Fredrickson, M. M. & Panagopoulos, C. Reasoning about interference between units: A general framework. *Political Analysis* **21**, 97–124 (2013).

20. Frisch, R. & Waugh, F. V. Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society* 387–401 (1933).

21. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794 (2016).

22. Team, R. C. *et al.* R: A language and environment for statistical computing (2013).

23. Rossum, G. Python reference manual (1995).

24. McKinney, W. *et al.* Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, vol. 445, 51–56 (Austin, TX, 2010).

25. Ooms, J. The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv preprint arXiv:1403.2805* (2014).

26. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).

27. Gaure, S. lfe: Linear group fixed effects. *The R Journal* **5**, 104–117 (2013).

28. Chen, T. *et al.* *xgboost: Extreme Gradient Boosting* (2019). URL https://CRAN.R-project.org/package=xgboost. R package version 0.90.0.1.

29. Analytics, R. domc: Foreach parallel adaptor for the multicore package. *R package version* **1** (2014).

30. Hlavac, M. Stargazer: Well-formatted regression and summary statistics tables. *R package version* **5** (2015).

31. Wickham, H. *ggplot2: elegant graphics for data analysis* (Springer, 2016).