A Game-Theoretic Approach for Hierarchical Policy-Making

Feiran Jia
Pennsylvania State University
fzj5059@psu.edu

Aditya Mate
Harvard University
aditya_mate@g.harvard.edu

Zun Li
University of Michigan
lizun@umich.edu

Shahin Jabbari
Harvard University
jabbari@seas.harvard.edu

Mithun Chakraborty
University of Michigan
dcsmc@umich.edu

Milind Tambe
Harvard University
milind_tambe@harvard.edu

Michael Wellman
University of Michigan
wellman@umich.edu

Yevgeniy Vorobeychik
Washington University in St. Louis
yvorobeychik@wustl.edu

ABSTRACT

We present the design and analysis of a multi-level game-theoretic model of hierarchical policy-making, inspired by policy responses to the COVID-19 pandemic. Our model captures the potentially mismatched priorities among a hierarchy of policy-makers (e.g., federal, state, and local governments) with respect to two main cost components that have opposite dependence on the policy strength, such as post-intervention infection rates and the cost of policy implementation. Our model further includes a crucial third factor in decisions: a cost of non-compliance with the policy-maker immediately above in the hierarchy, such as non-compliance of state with federal policies. Our first contribution is a closed-form approximation of a recently published agent-based model to compute the number of infections for any implemented policy. Second, we present a novel equilibrium selection criterion that addresses common issues with equilibrium multiplicity in our setting. Third, we propose a hierarchical algorithm based on best response dynamics for computing an approximate equilibrium of the hierarchical policy-making game consistent with our solution concept. Finally, we present an empirical investigation of equilibrium policy strategies in this game as a function of game parameters, such as the degree of centralization and disagreements about policy priorities among the agents, the extent of free riding as well as fairness in the distribution of costs.

KEYWORDS

Hierarchical Games, Empirical Game-Theoretic Analysis

ACM Reference Format:
Feiran Jia, Aditya Mate, Zun Li, Shahin Jabbari, Mithun Chakraborty, Milind Tambe, Michael Wellman, and Yevgeniy Vorobeychik. 2021. A Game-Theoretic Approach for Hierarchical Policy-Making. In *2nd International (Virtual) Workshop on Autonomous Agents for Social Good (AASG 2021), co-located with 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), London, UK, May 3–4, 2021*, IFAAMAS, 9 pages.

2nd International (Virtual) Workshop on Autonomous Agents for Social Good (AASG 2021), co-located with 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), A. Yadav, V. Kumar, S. Das, Y. Vorobeychik, May 3–4, 2021, London, UK.

1 INTRODUCTION

Democratic governments and institutions typically have a hierarchical structure. For example, policies in the U.S. emerge from complex interactions among the federal and state governments, as well as county boards and city councils and mayors. Similar structure exists in Canada and in European democracies. Such policy interactions are hierarchical, with higher levels in the hierarchy able to impose some constraints on the policies immediately below (e.g., the U.S. federal government can constrain what state policies can be). Violations of these constraints, in turn, entail a non-compliance cost to the violator, such as legal costs, penalties, or reputation loss. Many examples of such hierarchical policy structure commonly arise, such as in educational and vaccination decisions, as well as in devising policies for controlling a pandemic. Take COVID-19 social distancing policies as a concrete example. These policies commonly include recommendations at the national level, guidelines and restrictions at the state/province/district level, and policies for specific counties or cities. Moreover, a common feature of such hierarchical policy-making is that what ultimately matters are the policies actually deployed at the lowest level, since these are often most practical to enforce.

In general, policies are contentious. Agents at all levels of the policy-making hierarchy may disagree about the best policies, or more fundamentally, about the particular tradeoffs made in devising policies. For example, COVID-19 social distancing measures have considerable costs, both economic and socio-psychological, but lack thereof results in more people who become infected; different institutions disagree on how to trade off these concerns.

We propose a general model of hierarchical policy-making as a game among the policy-makers at all levels of the hierarchy. In this game, policies at the higher levels have an impact by imposing non-compliance costs on lower levels, but ultimate implementation of policies happens at the lowest level. Each agent in this game trades off two types of costs: policy *implementation cost* (e.g., socio-psychological or economical impacts of lockdowns) and *policy impact cost* (e.g., number of COVID-19 infections). Besides the impact on the structure of agent utilities, the hierarchy also impacts the sequence of moves: agents at higher levels precede lower levels (e.g. by announcing guidelines), with the latter observing and reacting to the policy recommendations by levels above them.

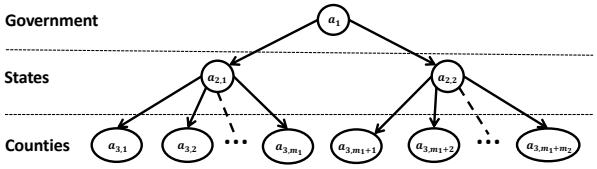


Figure 1: Hierarchy of policy-makers with 3 levels: Government, States and Counties.

In our game-theoretic model, each agent's action/pure strategy is a single, bounded scalar that represents the degree of social distancing in the pandemic-response context. Our first contribution is a novel solution concept which refines the subgame perfect Nash equilibrium, accounting for commonly occurring indifferences. Our second contribution is an analytic version of a recently proposed agent-based model (ABM) for COVID-19 pandemic spread estimation that accounts for social distancing [30]; we show that our analytic model closely mirrors short-term behavior of the ABM with a much shorter computational cost compared to the ABM.

We use our modeling framework to experimentally investigate possible phenomena arising from decentralized policy-making.

One of our questions relates to *policy free-riding*: Is it possible that (in equilibrium) a player lower in the hierarchy adopts a weak policy with a low implementation cost while imposing a negative externality on another player (perhaps on the same level) and hence also enjoying lower infection numbers owing to the latter player's stronger policy? Can a higher-level policy-maker mitigate such free-riding via non-compliance penalties? We show that the answer depends in a complex manner on different parameters such as initial infection rates, degree of contact among different parts of the population, weights on different types of cost, and the non-compliance cost structure. Our second set of experiments measures the *fairness* in the distribution of costs as a function of model parameters as well as degrees of centralization.

1.1 Related work

Our work is related to the line of research applying the social and behavioral sciences to the cost-benefit analysis of both centralized and decentralized decision-making under pandemic/epidemic conditions [26]. Some papers approach these trade-offs from an *optimal control* perspective [5, 19–21]. Others study the equilibria of various game-theoretic models of *individuals* deciding whether to follow guidelines for preventive measures (distancing, vaccination, etc.) and treatment, possibly against the (perceived) aggregate behavior of the population, under various models of disease propagation; e.g. the differential game model [17], the "wait and see" model of vaccinating behavior [1], evolutionary game-theoretic models [2, 11], and various others [3, 4, 8, 24] (see, e.g. [25] for a summary). We distinguish from these works by modeling the strategic interactions among *ideologically diverse, hierarchical policymakers* with explicit *non-compliance penalties*, and experimentally assess the impact of such interactions upon the actually implemented policies under various parameter settings.

Also related is the literature on ABM for pandemic spread and response policies that account for preferences/incentives of individuals [6, 10, 30]. [30] is of particular importance since our policy impact cost is computed by a closed-form approximation to their model. Other recent work includes the assessment of the impact of prevention and containment policies on the spread COVID-19 via causal analysis [13], Gaussian processes [16], and state-of-the-art data-driven non-pharmaceutical intervention models [22].

Instead of an analytic treatment, we empirically compute the (approximate) equilibrium of our complex, multi-level, continuous-action game using algorithmic approaches that exploit the structure of the problem. Thus, our methods belong to the category of *empirical game-theoretic analysis* (see, e.g. [7, 27–29]).

2 A MODEL OF HIERARCHICAL POLICY-MAKING

2.1 The Game Model

Consider as a running example COVID-19 hierarchical policy-making with three levels of decision-makers: Government (a single agent), States, and Counties. Each agent is a player in a game and chooses a social distancing policy (recommended or enforced). Next, we present a formal game theoretic model of this kind of hierarchical policy-making, focusing on strategic interactions among players both within and across the levels in this hierarchy.

Let $[m]$ denote the set $\{1, 2, \ldots, m\}$ for any $m \in \mathbb{Z}^+$. We represent the *players* in the *hierarchical policy-making game (HPMG)* by *nodes* in a directed rooted tree (we will use the terms players and nodes interchangeably), as illustrated in Figure 1 for our running example. A general *HPMG* has $L > 1$ levels or layers. Each level $l \in [L]$ is associated with a set of nodes/players, denoted by $\mathcal{L}_l$, with $n_l = |\mathcal{L}_l|$ the number of players in level l. Without loss of generality, let $n_1 = 1$ (we can always add a dummy layer with a single player who has a single strategy); we call the player in level 1, denoted by a_1 the *root-node*. The i^{th} node/player in an arbitrary level l is denoted by $a_{l,i}$. For each player in each level $l < L$, $a \in \mathcal{L}_l$, let $\chi(a)$ be the set of its children in the tree; clearly, $|\chi(a)| \geq 0$ for every a, and $\sum_{a \in \mathcal{L}_l} |\chi(a)| = n_{l+1}$ for each l. Likewise, for every player $a \in \mathcal{L}_l$ and for each $l > 1$, let $\pi(a)$ be its unique parent in the tree, where a directed edge goes from a parent to its child. In our running example in Figure 1, a_1 is the Government in level 1, level 2 consists of two States, both children of a_1, $\chi(a_1) = \{a_{2,1}, a_{2,2}\}$, and level 3 consists of Counties.

Each player a can take a scalar *action* $\alpha_a \in [0, 1]$. $\overline{\alpha}$ denotes the profile of actions of all players, and $\overline{\alpha}_l$ the restriction of this profile to a particular level l. In our pandemic-response policy-making example, α_a is an abstraction of the policy adopted by a, capturing the *the extent of overall activity* (conversely, $1 - \alpha_a$ represents the extent of social distancing implemented/recommended by a). Thus, small α_a corresponds to the greatest reduction in infection spread (due to stricter social distancing). On the other hand, a large α_a will entail a higher policy implementation cost, such as socio-economic and psychological costs of social distancing. At the extremes of our illustration, $\alpha_a = 1$ signifies no intervention, while $\alpha_a = 0$ corresponds to a complete lockdown.

HPMG is a sequential game in which players make strategic decisions following the sequence of layers. Specifically, the player

in level 1 moves (i.e., chooses a strategy) first, followed by all players in level 2, who first observe the strategy of a_1 and simultaneously choose a joint strategy profile in response. This is then followed by all players in level 3, and so on. Thus, all players in the same level l make strategic choices simultaneously.

Because all the utilities in our main application of this model (COVID-19 social distancing policies) are negative (i.e., costs), we next define the general model in terms of costs (negative utilities). The cost function of each player a has three components: *policy impact cost*, $\mathtt{C}_a^{\text{inc}}(\overline{\alpha})$, *policy implementation cost*, $\mathtt{C}_a^{\text{dec}}(\overline{\alpha})$, and, for each player in levels $l > 1$, *non-compliance cost*, $\mathtt{C}_a^{\text{NC}}(\alpha_a, \alpha_{\pi(a)})$. In the COVID-19 example, policy impact cost is a measure of infection spread (number of people infected in the player's geographic area, say), while implementation cost can be a psychological and economic costs of a lockdown. The non-compliance cost, in turn, is a penalty imposed by a policy-maker upon an agent within its jurisdiction for deviating from its recommendation (e.g., a fine, litigation costs, or reputational harm). An important piece of structure to the policy implementation and impact costs is that they directly depend for a player a not on the full profile of strategies by all players, but only on the layer l of the player a if $l = L$, and only the layer immediately below otherwise. To formalize, we introduce for each player a the notion of its *share* $\mu_a \in [0, 1]$. In our running example, a node's share can be interpreted as the proportion of the total population of the country that is under the jurisdiction of the corresponding player (e.g., the share of a state is the proportion of the total population that resides in this state). Thus, $\mu(a_1) = 1$, while the shares of the nodes in the lowest level L are arbitrary, except for the constraint $\Sigma_{a\in\mathcal{L}_L}\mu_a = 1$. For a level $1 < l < L$, we have $\mu_a = \Sigma_{a'\in\chi(a)}\mu_{a'}$ for every $a \in \mathcal{L}_l$. We now use the notion of shares to formally define the impact and implementation costs of policies.

- For each lowest-level player $a \in \mathcal{L}_L$, $\mathtt{C}_a^{\text{inc}}(\overline{\alpha})$ depends only on $\overline{\alpha}_L$, lies in $[0, 1]$, and is non-decreasing in each $\alpha_a \in \overline{\alpha}_L$; we provide further specifics of this function for our pandemic-response example in Section 2.3. For a higher-level player $a \in \mathcal{L}_l$, $l < L$, this cost is the share-weighted aggregate of those of its child-nodes:
$$\mathtt{C}_a^{\text{inc}}(\overline{\alpha}) = \tfrac{1}{\mu_a}\Sigma_{a'\in\chi(a)}\mu_{a'}\mathtt{C}_{a'}^{\text{inc}}(\overline{\alpha}).$$
- For each $a \in \mathcal{L}_L$, $\mathtt{C}_a^{\text{dec}}(\overline{\alpha}) \in [0, 1]$ depends only on, and is non-increasing, in α_a; in particular, in our pandemic-response example, we simply focus on the function $\mathtt{C}_a^{\text{dec}}(\overline{\alpha}) = 1 - \alpha_a$. Also, for each $a \in \mathcal{L}_l$, $l < L$,
$$\mathtt{C}_a^{\text{dec}}(\overline{\alpha}) = \tfrac{1}{\mu_a}\Sigma_{a'\in\chi(a)}\mu_{a'}\mathtt{C}_{a'}^{\text{dec}}(\overline{\alpha}).$$

Finally, we consider two variants of the non-compliance cost: *one-sided* under which there is no penalty for an α lower than that of the parent (capturing scenarios such as a policy-maker only punishing policy responses weaker than its recommendation), and *two-sided* under which any deviation is penalized regardless of direction [18], with the discrepancy being measured by the Euclidean distance for either variant:

$$\mathtt{C}_a^{\text{NC}}(\alpha, \alpha') = \begin{cases} (\max\{0, \alpha - \alpha'\})^2, & \text{if one-sided;} \\ (\alpha - \alpha')^2, & \text{if two-sided.} \end{cases}$$

Finally, each player $a \in \mathcal{L}_l$ for $l > 1$ has an idiosyncratic set of *weights* $\kappa_a \geq 0$ and $\eta_a \geq 0$ that trade its three cost components off against each other via a convex combination, and account for differences in ideology. Thus, the overall cost of such a player a is given by

$$\mathtt{C}_a(\overline{\alpha}) := \kappa_a\mathtt{C}_a^{\text{inc}}(\overline{\alpha}) + \eta_a\mathtt{C}_a^{\text{dec}}(\overline{\alpha}) + \gamma_a\mathtt{C}^{\text{NC}}(\alpha_a, \alpha_{\pi(a)}),$$

where $\gamma_a = 1 - \kappa_a - \eta_a$. The player a_1 obviously has no non-compliance issues, hence it has only one weight $\kappa_{a_1} > 0$, its overall cost being

$$\mathtt{C}_{a_1}(\overline{\alpha}) := \kappa_{a_1}\mathtt{C}_{a_1}^{\text{inc}}(\overline{\alpha}) + (1 - \kappa_{a_1})\mathtt{C}_{a_1}^{\text{dec}}(\overline{\alpha}).$$

2.2 Solution Concept

The solution concept we are primarily interested in is a *pure-strategy subgame perfect Nash equilibrium* (PSPNE) [23] of our continuous-action game which is sequential-move between levels and simultaneous-move within a level. However, the game may have multiple such equilibria, leading to an *equilibrium selection problem*.

An extreme but simple motivating scenario which gives rise to a multiplicity of equilibria, many of which are unreasonable, is when a lowest-level player a has non-compliance weight $\gamma_a = 1$ under a one-sided cost structure: player a would be indifferent among all values $\alpha_a \in [0, \alpha_{\pi(a)}]$ since any such value induces an overall cost of 0. Such indifference could also characterize the best response of a higher-level player. Consider a two-level variant of the game in Figure 1 (e.g., when counties are constrained to be compliant with the respective states); for each state $a \in \{a_{2,1}, a_{2,2}\}$, let $\kappa_a = 0$ and $\eta_a = 0.6$, hence $\gamma_a = 0.4$. Straightforward calculations show that the local minimum α_a^* of the overall cost of any such any state a over $[0, \alpha_{a_1}]$ is $\alpha_a^* = \alpha_{a_1}$ with a cost of $0.6(1 - \alpha_{a_1})$ and that over $(\alpha_{a_1}, 1]$ is

$$\alpha_{a'}^* = \begin{cases} 1, & \text{cost} = 0.4(1 - \alpha_{a_1})^2, \quad \alpha_a \geq 0.25; \\ \alpha_{a_1} + 0.75, & \text{cost} = 0.375 - 0.6\alpha_{a_1}, \quad \text{otherwise.} \end{cases}$$

Thus, the unique best response of either state (whose costs are independent of each other) to *any* government policy $\alpha_{a_1} \geq 0.25$ is 1, i.e., there are infinitely many equilibria with the government recommending any $\alpha_{a_1} \geq 0.25$ but each state choosing 1 regardless. The fact that the government would recommend a policy intervention (which could be as strong as $\alpha_{a_1} = 0.25$) knowing fully well that both states would choose no intervention even under the threat of a non-compliance penalty seems absurd, but this absurdity cannot be eliminated by the above solution concept.

With this in mind, we propose and use the following equilibrium selection criterion. For any player $a \in \mathcal{L}_l$, $l < L$, define its *social cost* $\mathtt{SC}_a(\overline{\alpha}_{\chi(a)})$ for any action profile $\overline{\alpha}_{\chi(a)}$ of its children as the share-weighted aggregate of the overall costs of its children, that is: $\mathtt{SC}_a(\overline{\alpha}_{\chi(a)}) := \frac{1}{\mu_a}\sum_{a'\in\chi(a)}\mu_{a'}\mathtt{C}_a(\overline{\alpha})$. Evidently, this quantity is, in general, distinct from $\mathtt{C}_a(\overline{\alpha})$.

If multiple values of α_a induce equilibria for a particular $\overline{\alpha}_{\chi(a)}$, then we will pick the α_a which minimizes $\mathtt{SC}_a(\overline{\alpha}_{\chi(a)})$, breaking further ties in favor of a higher α_a (i.e., smaller policy impact). We refer to this solution concept, which is a refinement of PSPNE, as *minimal-impact pure-strategy subgame perfect Nash equilibrium (MI-PSPNE)*.

In general, a MI-PSPNE will not exist. Consequently, we will seek to compute an ϵ-MI-PSPNE, where ϵ is the highest benefit from deviation by any player a. Below (Section 3) we present a general approach for finding such approximate equilibria in our setting.

2.3 Infection Dynamics and Cost

We now come to the particular instantiation of $\mathtt{C}_a^{\text{inc}}(\cdot)$ for each of the lowest-level players $a \in \mathcal{L}_L$ (Counties in Figure 1). Recently, Wilder et al. [30] developed and analyzed an agent-based model (ABM) for COVID-19 spread that accounts for the *degree of contact* (both within and between households) among individuals from different parts of a population.[1] However, this ABM is computationally expensive, making its use for equilibrium computation impractical at scale. In this section, we will derive a closed-form model of infection spread that (as we show below) relatively closely mirrors the expected number of infections of the ABM over a short horizon.

Let N_a and I_a^0 denote the fixed population of County a and the number of infections in a before policy intervention respectively. An individual who is not currently infected but can develop an infection on contact with someone infected is *susceptible*. We call an individual from County a' *active* in County a if that individual is capable of making contact (through travel etc.) with a susceptible individual in County a; if $a' = a$, we say that the individual is active within County a. A major parameter of the ABM is the *transport matrix* $R = \{r_{aa'}\}_{a,a' \in \mathcal{L}_L}$, where $r_{aa'} \geq 0$ is the proportion of the population of County a' that is active in County a in the absence of an intervention. Thus, in the absence of policy intervention, the total number of individuals from County a' active in County $a \neq a'$ is $N_{a'} r_{aa'}$ and the total number of infected individuals from County a' active in County $a \neq a'$ is $I_{a'}^0 r_{aa'}$.

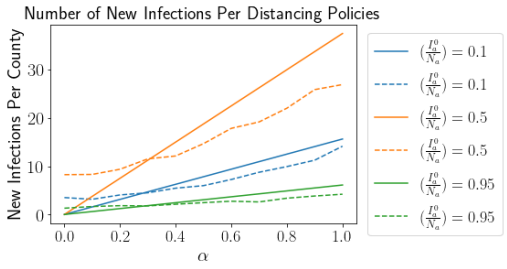
The policy α_a affects the population in two ways: it scales down both the susceptible and active sub-populations. In other words, under the policy intervention, County a has $(N_a - I_a^0)\alpha_a$ susceptible individuals, and there are $N_{a'}\alpha_{a'} r_{aa'}$ active individuals in County a from County a', out of whom $I_{a'}^0 \alpha_{a'} r_{aa'}$ are (initially) infected. Hence, the proportion of infected active individuals in County a is given by

$$\rho_a(\overline{\alpha}_L) := \frac{\sum_{a' \in \mathcal{L}_L} I_{a'}^0 \alpha_{a'} r_{aa'}}{\sum_{a' \in \mathcal{L}_L} N_{a'} \alpha_{a'} r_{aa'}}.$$

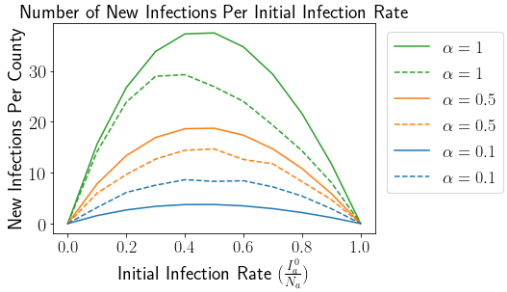
We will now focus on an arbitrary susceptible individual in County a and lay down our assumptions on the process why which she may contract an infection: This individual makes actual contact with a random sample of X active individuals drawn from a Poisson distribution with mean C, which is a parameter in our model [30]; this distribution is fixed across all individuals in all Counties, and all these contacts are mutually independent. The next assumption is that, in this sample of X contacts for a susceptible individual in County a, the proportion of infected individuals is $\rho_a(\overline{\alpha}_L)$.

Let $p \in (0, 1)$ denote the probability that a susceptible individual becomes infected upon contact with an infected individual, i.e. the probability that contact with an infected individual does not infect a susceptible individual is $(1 - p)$. Since all $X\rho_a(\overline{\alpha})$ infected contacts

[1]The model in Wilder et al. [30] is an individual-level variant of the well-known susceptible-exposed-infectious-recovered or SEIR model but this paper assumes that every exposed person eventually becomes infected after an incubation period.



(a) New infections vs. policy (shared by Counties).



(b) New infections vs. initial infection rate (shared by Counties).

Figure 2: Comparison of ABM output (solid lines) with closed-form approximation (dashed lines).

of an arbitrary susceptible individual are mutually independent, the probability that the susceptible individual develops an infection is $1 - (1 - p)^{X\rho_a(\overline{\alpha})}$. We also interpret this as the proportion of the $(N_a - I_a^0)\alpha_a$ susceptible individuals in County a who end up getting infected. Let $\mathtt{Infect}_a(\overline{\alpha})$ denote the *expected* number of *additional, post-intervention infections* in County a. Thus,

$$\begin{aligned}\mathtt{Infect}_a(\overline{\alpha}) &= \mathbb{E}_X[(N_a - I_a^0)\alpha_a(1 - (1-p)^{X\rho_a(\overline{\alpha})})] \\ &= (N_a - I_a^0)\alpha_a(1 - \mathbb{E}_X[((1-p)^{\rho_a(\overline{\alpha})})^X]).\end{aligned}$$

Define $y_a(\overline{\alpha}_L) := (1-p)^{\rho_a(\overline{\alpha})}$. Since $X \sim \textsc{Poisson}(C)$, we can show from the properties of the Poisson distribution that

$$\mathtt{Infect}_a(\overline{\alpha}) = (N_a - I_a^0)\alpha_a(1 - e^{-C(1-y_a(\overline{\alpha}_L))}). \tag{1}$$

Finally, we define the infection cost to be $\mathtt{C}_a^{\text{inc}}(\overline{\alpha}) = \mathtt{Infect}_a(\overline{\alpha})/N_a$.

We ran some preliminary experiments comparing Equation (1) with the actual output of the ABM [30]; partial results are shown in Figure 2. Note that Equation (1) is a one-shot formula, whereas the ABM computes contacts and infections recursively over several time-periods with an initial *incubation period* so that the effect of the first-period contacts are manifested only after a delay. Hence, we contrast the ABM output after 8 periods (to account for the average incubation period of 7 days [12]) with the above closed-form estimation. In the experiments we report, we have 2 States under the Government, each State having 2 Counties (4 Counties in total); each County a has a population of $N_a = 250$; the transport matrix is symmetric, given by $r_{aa'} = 0.25$ for every pair of Counties a, a'. We set $p = 0.047$ [30] and $C = 15$ (calculated based on Prem et al. [14]). For each set of experiments (represented by a separate color in

Figure 2), each County has the same *initial infection rate* I_a^0/N_a and applies the same policy α. In Figure 2a, we vary α on the x-axis, for different (fixed) values of I_a^0/N_a which is the same for all Counties; similarly, in Figure 2b, for different policies, we vary the initial infection rates. The plots indicate qualitative similarity between the ABM and our approximation; a salient point of similarity is that the additional number of infections decreases as the initial infection rate gets higher or lower than a middling point, everything else remaining the same. This is because a higher infection rate implies less "room for growth" due to a fixed population, whereas a lower value of the same rate causes fewer further infections over the same horizon.

3 SOLUTION APPROACH

Our HPMG model is essentially an extensive-form game model endowed with one-dimensional action space for each agent resulting in a non-convex strategic landscape. To seek for PSPNE in the hierarchical game(HG), we propose a backward induction algorithm incorporated with a payoff point query interface and a best response computation component solving for a joint-policy profile in equilibrium. The algorithm exploits the hierarchical structure by propagating strategic information between consecutive levels, detailed as follows. Given a joint action profile at levels $1, \ldots, l-1$, the players at level l compose a simultaneous-move game whose payoffs emerge from the strategic interactions from levels below them. To obtain payoffs for a certain action profile at level l, we recursively call to the next level $l+1$ till we reach the bottom level L. Then at level l, we use these payoffs to solve for an approximate Nash equilibrium. Since every such simultaneous-move game lacks the tractable analytic payoff structure for gradient-based optimization, in our current implementation we discretize the infinite strategy space and adopt *best response dynamics* (BRD) for equilibrium computation.

Algorithm 1 computes the $\epsilon-$ equilibrium among players at a single level l given tunable parameters. An $\epsilon-$ equilibrium at level l is an action profile $\overline{\alpha}_l$ where no agent $\alpha_{a_{l,i}}$ can decrease their cost by more than ϵ by a unilateral deviation (Lines 3–6). Let $\boldsymbol{\alpha}_{l_1:l_2}$ denote the sequence of actions $\overline{\alpha}_{l_1}, ..., \overline{\alpha}_{l_2}$; $\boldsymbol{T}_{l:L} = T_l, ..., T_L$ the maximum numbers of steps of BRD at each level $l > 1$; $\boldsymbol{e}_{l:L} = e_l, ..., e_L$ the limits of ϵ for each level. At each round $t < T$, we randomly select a subset of k_l out of n_l agents to best respond simultaneously to the existing profile; we call the variant with $k_l = n_l$ *synchronous BRD*. To analyze the dependence of the number of BRD steps to reach equilibrium on the sample size k_l, we performed a small set of experiments on a 2-level game with 1 Government and a variable number n_2 of States. We focused on symmetric settings, i.e. all states have equal population, equal initial infections, and equal weight vectors. The transport matrix is also symmetric (see Section 4 for details). Figure 3 shows that setting $k_2 = n_2$ in Algorithm 1 (synchronous BRD) results in fastest convergence to equilibrium for several choices of n_2. To increase efficiency, we should pick a subset of agents that can improve their payoffs after best-responding (Line 7). The synchronous BRD might get trapped in a cycle of moves. The way we solve this issue is to keep a memory of the moves, then check whether the new profile already exists in the memory and, if yes (i.e. a cycle is detected), we jump to a new profile and resume

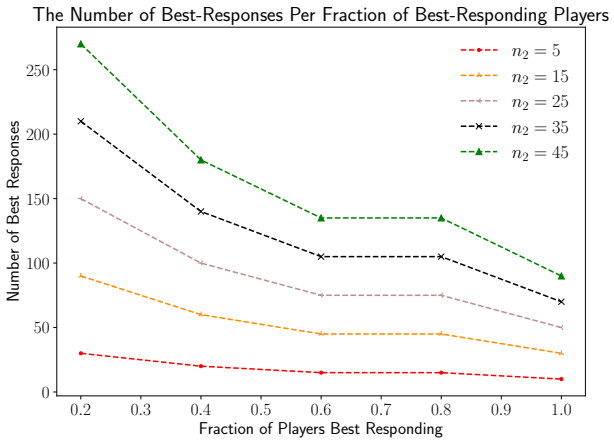


Figure 3: The number of best response steps (y-axis) as a function of the fraction of players that are best responding (x-axis). Each curve corresponds to a game with n_2 players (ranging from 5 to 45) in the second level which is also the lowest level of game.

ALGORITHM 1: HG-PSPNE

Input: $\boldsymbol{\alpha}_{1:l-1}$.

Parameter:param$_l = \{\boldsymbol{T}_{l:L}, \boldsymbol{k}_{l:L}, \boldsymbol{e}_{l:L}\}$.

1: Let $t \leftarrow 1, \epsilon_l \leftarrow \infty$. Initialize $\overline{\alpha}_l$ randomly.
2: **while** $t \leq T_l$ or $\epsilon_l \leq e_l$ **do**
3: **for** $a_{l,i}$ in $\mathcal{L}_l$ **do**
4: **if** l is the lowest level L **then**
5: $\alpha'_{a_{l,i}} \leftarrow \arg\min_{\alpha_{l,i}} \mathrm{C}_{a_{l,i}}(\overline{\alpha})$
6: **else**
7: $\alpha'_{a_{l,i}} \leftarrow \arg\min_{\alpha_{l,i}} \mathrm{C}_{a_{l,i}}(\text{HG-PSPNE}(\boldsymbol{\alpha}_{1:l}))$
8: **end if**
9: **end for**
10: Calculate ϵ_l for profile $\overline{\alpha}_l$ and update ϵ_l if lower than the current value.
11: Pick k_l agents to best respond to $\alpha'_{l,i}$.
12: $t \leftarrow t+1$.
13: **end while**
14: **return** $\overline{\alpha}^*$ where $\overline{\alpha}^*_l$ has the lowest ϵ_l.

the BRD. Finally, the algorithm returns the profile with the lowest ϵ_l when the termination condition is met.

To search for the best strategy, we discretize the continuous strategy space and use grid search with tie-breaking (smaller policy impact) to recover the optimum value. However, in the experiments shown Figure 2a, we observe that our approximation of the infection cost is nearly linear. Although we have no guarantees, it is reasonable to ask whether the overall cost of a lowest-level player is almost convex in its policy (given the particular closed forms we use for the implementation and non-compliance costs) and hence whether we could use *binary search* (i.e. the bisection methods) to speed up our BRD. Figure 4 shows the run-time of a two-level game for $n_2 = 10$ to 100 players in the second level when we replace the grid search in the lowest level with the binary search under a

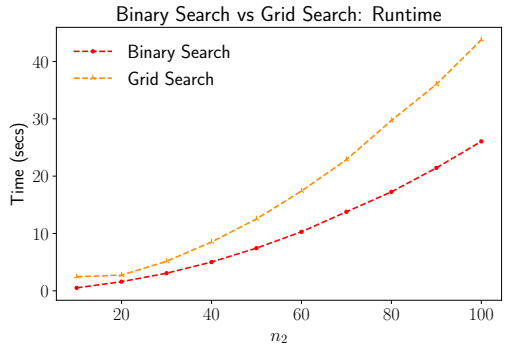


Figure 4: Run-time in secs (y-axis) of binary search and grid search as a function of n_2 (x-axis) in a symmetric setting.

symmetric setting (i.e. equal populations and symmetric transport matrix in level 2). In those experiments, we all find the PSPNEs with $\epsilon_2 = 0$. Binary search yields same results as grid search, but is more efficient.

4 EXPERIMENTS

In this section, we describe two sets of experiments on our HPMG framework using the methodology discussed in Section 3. Given the large and complex set of game parameters, we report the most insightful experimental results we obtained, deferring additional results to the full version. In Section 4.1, we quantify the notion of *free-riding* and explore conditions under which free-riding appears in equilibrium and can be circumvented by non-compliance penalties. In Section 4.2, we study how different degrees of centralization and mismatched priorities of players in HPMG can impact fairness in the distribution of costs. In all our experiments, we use the 3-level HPMG of Figure 1 with 2 States, denoted simply by 1 and 2 (also in subscripts), and a number of Counties (to be specified) with equal population. Moreover, we say that a setting has *transport symmetry* if the transport matrix R is proportional to an identity matrix i.e. $r_{aa'} = 1/n_L$ for any two lowest-level players a, a'. There are infinite ways in which R could be asymmetric; we focus on a particular type of asymmetry where one subset of Counties (or more generally lowest-level players) $\mathcal{F}$ are *globally favorite destinations* (and equally popular) and all others are equally (un)popular, i.e. for each County j, $r_{ij} = r^H > r^L = r_{kj}$ for each $i \in \mathcal{F}$ and each $k \in \mathcal{L}_3 \setminus \mathcal{F}$ for some $0 < r^L < r^H < 1$, and $\sum_{i \in \mathcal{L}_3} r_{ij} = 1$.

4.1 Free-riding

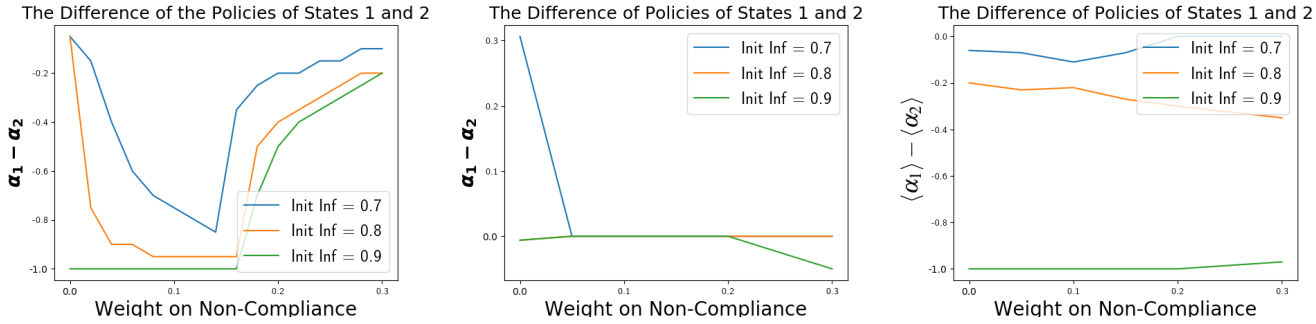
We begin with the rationale for our measurement and visualization of free-riding. Suppose State 2 has a higher initial infection (than State 1); then, intuitively, it may prefer a weak distancing policy ($\alpha_2 \gg 0$) to reduce its implementation cost as most of its population is already infected; at the same time, since infection can spread from State 2, State 1 could still suffer a large infection cost unless it employs a strong distancing policy ($\alpha_2 \ll 1$). This creates the possibility for State 2 to free-ride off State 1; but, whether this actually happens depends on the combination of parameters (including non-compliance issues with the Government above and Counties below) and the same possibility may be created by other conditions. We use this difference between policy strengths of the States as an indicator of the degree of free-riding. Intuitively as $\alpha_1 - \alpha_2$ approaches -1 (1), the degree of free-riding of State 2 (1) off State 1 (2) increases.

In our reported experiments, we assume Government indifference between infection and implementation (i.e. $\kappa_{a_1} = 0.5$) and an even split of the population between States (i.e. $N_1 = N_2 = 500$, hence $\mu_1 = \mu_2 = 0.5$); both States use the same weight vector which we vary. The crucial difference between States is in the initial infection rate I_a^0/N_a (as we discuss shortly).[2] In all experiments, each State consists of 5 Counties.

Figure 5 depicts our results under transport symmetry and other conditions which we will now detail. First, we assume that all Counties are constrained to comply with their respective States so that the policies set by States 1 and 2 actually get implemented in their respective jurisdictions (i.e. we essentially have a 2-level HPMG, hence the number and weights of Counties are immaterial). In Figure 5a, we plot the variation in this policy-difference against the States' shared non-compliance weight γ_a under the following conditions: State 1's initial infection rate is fixed at 0.1 (low) while that of State 2 varies over $\{0.7, 0.8, 0.9\}$; for either State $a \in \{1, 2\}$, we have $\kappa_a = 0.9(1 - \gamma_a)$ and $\eta_a = 0.1(1 - \gamma_a)$ for each value of the non-compliance weight. We observe that free-riding is exacerbated as State 2's initial infection rate becomes larger, although a high enough non-compliance weight will mitigate the problem. However, interestingly, lower values of the non-compliance weight also exhibit a lower degree of free-riding. Increasing the non-compliance weight forces both States to monotonically weaken their policies (towards 1) but State 1 is more conservative, maintaining a strict policy (at 0) up to a non-compliance weight of (at least) 0.15 and only then weakening its policy to the level of State 2, as shown in Figures 6 and 7. This accounts for the non-monotonic dependence of free-riding on non-compliance cost weight as observed in Figure 5a.

What happens when we allow Counties to not comply with the respective States? We report results for a setting where each County's initial infection rate and weight vector is identical to that of its corresponding State. Recall that, with Counties no longer constrained to comply, State policies are *recommendations* policies whereas those that are implemented are County actions. With this mind, we report in Figures 5b and 5c the difference in State policies $\alpha_1 - \alpha_2$ as well as the difference $\langle \alpha_1 \rangle - \langle \alpha_2 \rangle$, where $\langle \alpha_a \rangle$ is the average of the equilibrium policies set by all Counties ins State $a \in \{1, 2\}$, over the same combinations of weights and initial infection rates as Figure 5a. We find virtually no evidence of free-riding from either measure (in the extreme case represented by the lowest curve Figure 5c is perhaps better interpreted as State 2 giving up on policy intervention rather than free-riding off State 1). This indicates that distributing autonomous policy-making among several smaller-scale actors may also have a mitigating effect on free-riding, making

[2] We only report results for two-sided compliance costs at all levels; we did not observe any evidence of free-riding mitigation using the one-sided variant in our experiments.



(a) Counties constrained to comply. **(b) Counties free to not comply.** **(c) Counties free to not comply.**

Figure 5: Free-riding (y-axis) as a function of non-compliance cost weight (x-axis). Each curve corresponds to a different initial infection rate of State 2 (Init Inf) as indicated in the legend.

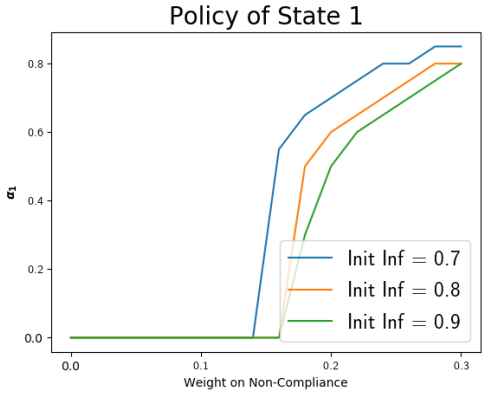


Figure 6: The Policy of State 1 (y-axis) as a function of non-compliance weight (x-axis) under transport symmetry. Each curve corresponds to a different initial infection rate for State 2 as specified in the legend.

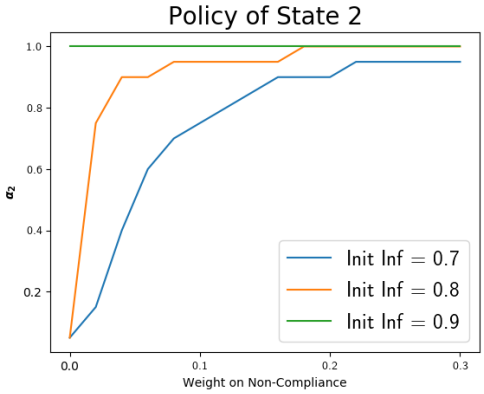


Figure 7: The Policy of State 2 (y-axis) as a function of non-compliance weight (x-axis) under transport symmetry. Each curve corresponds to a different initial infection rate for State 2 as specified in the legend.

the impact on free-riding of non-compliance penalties from the highest level weaker.

We now repeat these experiments but in a specific setting violating transport symmetry with Counties constrained to comply: we make State 1 the favorite destination with $r^H = 0.8$ (equivalently, all Counties in State 1 are equally favored as destinations). Figure 8 shows results when the States' infection weights are $\kappa_1 = 0.8(1-\gamma_1)$ and $\kappa_2 = 0.9(1-\gamma_2)$ respectively (still with $\gamma_1 = \gamma_2$), State 1's initial infection rate is fixed at 0.5 (moderate) while that of State 2 varies in $\{0.1, 0.5, 0.6, 0.7, 0.8, 0.9\}$, and Counties are constrained to comply. While it is true that the States' aversion to non-compliance is able to lessen free-riding monotonically and more readily as State 2's initial infection rate grows, the most salient feature is the sudden reversal in the status of the apparent free-rider as State 2's initial infection rate crosses a (high) threshold. Further inspection reveals that, although State 1 has a higher proportion of active individuals even from State 2 and cares about infections only slightly less than State 2 (but still with κ as high as 0.8), the initial infection rate of 0.5 is high enough for it to respond weakly ($\alpha_1 \approx 1$) while

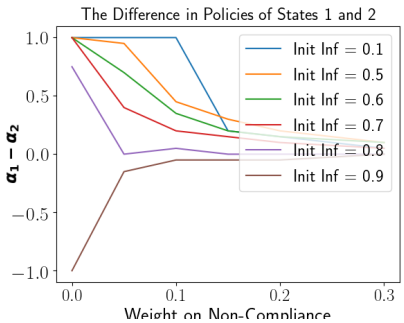
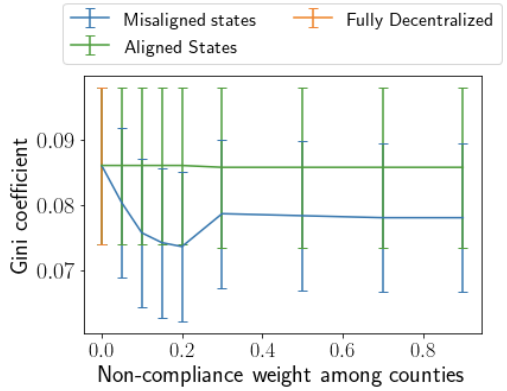
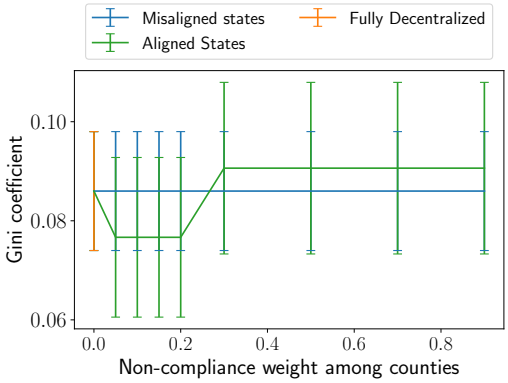


Figure 8: Free-riding (y-axis) as a function of non-compliance cost weight (x-axis). Each curve corresponds to a different initial infection rate of State 2 (Init Inf) as indicated in the legend.



(a) Transport symmetry



(b) Transport asymmetry (1 favorite County per State)

Figure 9: Gini coefficient of costs averaged over (a) 50 trials for each scenario; (b) 30 trials for *Aligned States*, 50 for each other scenario. Error bars show one standard error.

State 2 weakens its policy more gradually with increase in the non-compliance weight γ_a, enabling State 1 to free-ride. However, once State 2's initial infection rate becomes sufficiently high, it gives up on distancing policy even for low γ_a, forcing State 1 to strengthen its policy — this is reflected in the sign reversal and increased magnitude of the policy difference (figures deferred to the full version).

4.2 Fairness

Another property of the equilibria of an HPMG worth studying is how *fair* the distribution of costs is among the Counties for different degrees of centralization and different priorities of the States. Of the many fairness concepts that exist in the literature, we apply the popular measure, the *Gini coefficient* [9], to Counties' overall costs at profile $\overline{\alpha}$ returned by Algorithm 1:

$$\texttt{Gini}(\overline{\alpha}) = \frac{\sum_{a\in\mathcal{L}_3}\sum_{a'\in\mathcal{L}_3}|\mathsf{C}_a(\overline{\alpha}) - \mathsf{C}_{a'}(\overline{\alpha})|}{2n_L\sum_{a\in\mathcal{L}_3}\mathsf{C}_a(\overline{\alpha})}.$$

We report experiments with 5 Counties under each of 2 States, all non-compliance costs being one-sided. For the Government, $\kappa_{a_1} = \eta_{a_1} = 0.5$; for each State $b \in \{1,2\}$, $\gamma_b = 0.5$, and there are two different scenarios of the *full game* based on the ratios κ_b/η_b: (1) *Misaligned States* if this ratio is 20/80 for State 1 and 80/20 for State 2, (2) *Aligned States* if it is 50/50 for either state. A third scenario we study is *full decentralization* where we set each County's non-compliance weight to 0 so that HPMG degenerates into a simultaneous-move game among Counties. For each scenario, we apply two treatments with respect to the transport matrix: transport symmetry and a specific asymmetry where each State has 1 County that is (universally) favorite with $r^H = 0.35$. In each situation, we vary the shared non-compliance weight γ_a of every County a as an independent variable (unless it is fixed at 0), draw a uniform random sample $\kappa'_a \sim \mathbb{U}[0,1]$, and set the infection weight at $\kappa_a = \kappa'_a(1-\gamma_a)$. Each set of draws for all Counties constitutes one trial. Figure 9 provide Gini coefficient scatter plots for transport symmetry and our specific asymmetry respectively. The distribution of overall costs seems reasonably and comparably fair (lower is better) across scenarios.

5 DISCUSSION AND FUTURE WORK

We have initiated the study of a new game-theoretic model motivated by decentralized policy-making under pandemic conditions, and experimentally uncovered interesting aspects of its equilibria. Directions for future work are: more extensive experimentation over parameter configurations (with the formulation and testing of causal hypotheses); using the actual ABM [30] instead of our closed-form infection estimation and handling the resulting computational efficiency issues; considering more complex policies (e.g. adaptive strategies) and invoking more sophisticated EGTA approaches. It would also be interesting to apply HPMG or its natural variants to other problems of hierarchical decision-making such as within a corporation or political organization, or in the interplay of domestic and international politics Putnam [15], where "superiors" can impose a non-compliance penalty or offer a compliance bonus.

ACKNOWLEDGMENTS

This work is supported by the ARO under contract W911NF1810208.

REFERENCES

[1] Samit Bhattacharyya and Chris Bauch. 2011. "Wait and see" vaccinating behaviour during a pandemic: A game theoretic analysis. *Vaccine* 29, 33 (2011), 5519–5525.

[2] Martin Brüne and Daniel Wilson. 2020. Evolutionary perspectives on human behavior during the Coronavirus pandemic: insights from game theory. *Evolution, medicine, and public health* 2020, 1 (2020), 181–186.

[3] Frederick Chen. 2012. A mathematical analysis of public avoidance behavior during epidemics using game theory. *Journal of Theoretical Biology* 302 (2012), 18–28.

[4] Frederick Chen, Miaohua Jiang, Scott Rabidoux, and Stephen Robinson. 2011. Public avoidance and epidemics: insights from an economic model. *Journal of Theoretical Biology* 278, 1 (2011), 107–119.

[5] Eli Fenichel. 2013. Economic considerations for social distancing and behavioral based policies during an epidemic. *Journal of Health Economics* 32, 2 (2013), 440–451.

[6] Eli Fenichel, Carlos Castillo-Chavez, Graziano Ceddia, Gerardo Chowell, et al. 2011. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences* 108, 15 (2011), 6306–6311.

[7] Nicola Gatti and Marcello Restelli. 2011. Equilibrium approximation in simulation-based extensive-form games. In *International Conference on Autonomous Agents and Multiagent Systems*. 199–206.

[8] Mark Gersovitz. 2010. Disinhibition and immiserization in a model of SIS diseases. *Unpublished Papers, Johns Hopkins University* (2010).

[9] Corrado Gini. 1912. Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Eds. E. Pizetti, T. Salvemini)* (1912).

[10] Nicolas Hoertel, Martin Blachier, Carlos Blanco, Mark Olfson, et al. 2020. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature medicine* 26, 9 (2020), 1417–1421.

[11] Ariful Kabir and Jun Tanimoto. 2020. Evolutionary game theory modelling to represent the behavioural dynamics of economic shutdowns and shield immunity in the COVID-19 pandemic. *Royal Society Open Science* 7, 9 (2020), 201095.

[12] Stephen Lauer, Kyra Grantz, Qifang Bi, Forrest Jones, Qulu Zheng, et al. 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine* 172, 9 (2020), 577–582.

[13] Atalanti Mastakouri and Bernhard Schölkopf. 2020. Causal analysis of COVID-19 spread in Germany. *Advances in Neural Information Processing Systems* 33 (2020).

[14] Kiesha Prem, Alex Cook, and Mark Jit. 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Computational Biology* 13, 9 (2017), e1005697.

[15] Robert D Putnam. 1988. Diplomacy and domestic politics: the logic of two-level games. *International organization* (1988), 427–460.

[16] Zhaozhi Qian, Ahmed Alaa, and Mihaela van der Schaar. 2020. When and How to Lift the Lockdown? Global COVID-19 Scenario Analysis and Policy Assessment using Compartmental Gaussian Processes. *Advances in Neural Information Processing Systems* (2020).

[17] Timothy Reluga. 2010. Game theory of social distancing in response to an epidemic. *PLoS Computational Biology* 6, 5 (2010), e1000793.

[18] Rick Rojas. 2020. Trump Criticizes Georgia Governor for Decision to Reopen State. https://www.nytimes.com/2020/04/22/us/trump-georgia-governor-kemp-coronavirus.html. *New York Times* (2020). Accessed: 2020-01-19.

[19] Bob Rowthorn and Flavio Toxvaerd. 2012. The optimal control of infectious diseases via prevention and treatment. *CEPR Discussion Paper No. DP8925* (2012).

[20] Robert Rowthorn and Jan Maciejowski. 2020. A cost–benefit analysis of the COVID-19 disease. *Oxford Review of Economic Policy* 36, Supplement (2020), S38–S55.

[21] Suresh Sethi. 1978. Optimal quarantine programmes for controlling an epidemic spread. *Journal of the Operational Research Society* (1978), 265–268.

[22] Mrinank Sharma, Sören Mindermann, Jan M Brauner, Gavin Leech, Anna B Stephenson, Tomáš Gavenčiak, Jan Kulveit, Yee Whye Teh, Leonid Chindelevitch, and Yarin Gal. 2020. How Robust are the Estimated Effects of Nonpharmaceutical Interventions against COVID-19? *Advances in Neural Information Processing Systems* (2020).

[23] Yoav Shoham and Kevin Leyton-Brown. 2008. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

[24] Flavio Toxvaerd. 2019. Rational disinhibition and externalities in prevention. *International Economic Review* 60, 4 (2019), 1737–1755.

[25] Flavio Toxvaerd. 2020. Equilibrium Social Distancing. Cambridge–INET Working Paper Series No. 2020/08.

[26] Jay Van Bavel, Katherine Baicker, Paulo Boggio, Valerio Capraro, et al. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* (2020), 1–12.

[27] Yevgeniy Vorobeychik, Daniel Reeves, and Michael Wellman. 2007. Constrained Automated Mechanism Design for Infinite Games of Incomplete Information. In *Conference on Uncertainty in Artificial Intelligence*. 400–407.

[28] Yevgeniy Vorobeychik and Michael Wellman. 2008. Stochastic search methods for Nash equilibrium approximation in simulation-based games. In *International Conference on Autonomous Agents and Multiagent Systems*. 1055–1062.

[29] Michael Wellman. 2006. Methods for empirical game-theoretic analysis. In *AAAI Conference on Artificial Intelligence*. 1552–1556.

[30] Bryan Wilder, Marie Charpignon, Jackson Killian, Han-Ching Ou, Aditya Mate, Shahin Jabbari, et al. 2020. Modeling between-population variation in COVID-19 dynamics in Hubei, Lombardy, and New York City. *Proceedings of the National Academy of Sciences* 117, 41 (2020), 25904–25910.