



# COVID-19 vaccine incentive scheduling using an optimally controlled reinforcement learning model

K. Stuckey<sup>a</sup>, P.K. Newton<sup>b,c,\*</sup>

<sup>a</sup> Department of Aerospace & Mechanical Engineering, University of Southern California, Los Angeles CA 90089-1191, United States of America

<sup>b</sup> Department of Aerospace & Mechanical Engineering, Mathematics, Quantitative and Computational Biology, University of Southern California, Los Angeles CA 90089-1191, United States of America

<sup>c</sup> The Ellison Institute for Transformative Medicine, University of Southern California, Los Angeles CA 90089-1191, United States of America



## ARTICLE INFO

### Article history:

Received 11 August 2022  
 Received in revised form 2 November 2022  
 Accepted 2 December 2022  
 Available online 16 December 2022  
 Communicated by V.M. Perez-Garcia

### Keywords:

Vaccine uptake dynamics  
 Evolutionary game theory  
 Hawk–Dove games  
 Optimal control  
 Reinforcement learning dynamics  
 Dynamic incentives

## ABSTRACT

We model Covid-19 vaccine uptake as a reinforcement learning dynamic between two populations: the vaccine adopters, and the vaccine hesitant. Using data available from the Center for Disease Control (CDC), we estimate the payoff matrix governing the interaction between these two groups over time and show they are playing a Hawk–Dove evolutionary game with an internal evolutionarily stable Nash equilibrium (the asymptotic percentage of vaccinated in the population). We then ask whether vaccine adoption can be improved by implementing dynamic incentive schedules that reward/punish the vaccine hesitant, and if so, what schedules are optimal and how effective are they likely to be? When is the optimal time to start an incentive program, how large should the incentives be, and is there a point of diminishing returns? By using a tailored replicator dynamic reinforcement learning model together with optimal control theory, we show that well designed and timed incentive programs can improve vaccine uptake by shifting the Nash equilibrium upward in large populations, but only so much, and incentive sizes above a certain threshold show diminishing returns.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

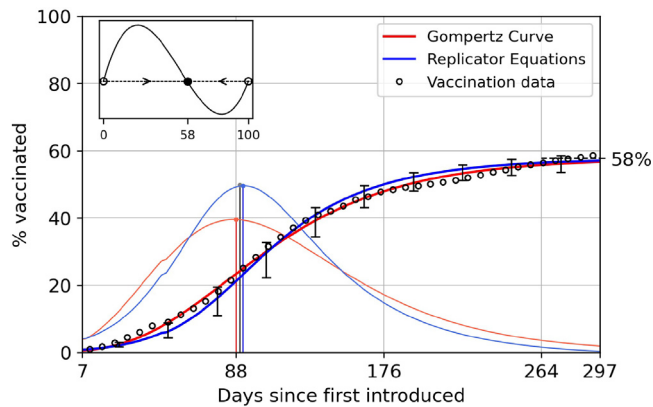
The voluntary uptake of vaccines for Covid-19 has proven to be a challenge around the world, but particularly in the United States where multiple vaccine options have been available since early in 2021 yet there remains a sizable unvaccinated group. After an initial population of early adopters were vaccinated, the surge began to slow, despite widespread availability, and has now reached what looks to be a fairly stable resting point (see Fig. 1). While 100% voluntary compliance is rarely if ever achievable, vaccine hesitancy [1] has proven to be more widespread for Covid-19 than for other vaccines, such as seasonal flu vaccines [2], the polio vaccine, smallpox, HPV and others [3]. With any widespread nationally coordinated vaccination effort, there will always be a population of people who we label vaccine adopters (e.g. elderly, immuno-compromised, healthcare workers) who get vaccinated as soon as they are eligible, or shortly thereafter, then others follow. The vaccine uptake curve for this group in these early stages is limited mostly by vaccine availability and logistics. There is also a vaccine hesitant population who will delay their initial chances to get vaccinated, then as they see others getting sick

and weigh evidence and public opinion, some might decide to vaccinate (adopters), while still others might further delay, or forgo their chance altogether for various reasons (hesitant) [4]. We view the full population as a collection of two types of players in a time-evolving game, who interact, learn, and receive payoffs (reward/punishment) according to the strategy (adopt/forgo) they choose, where the interactions determine the fitness of the players using one of two strategies, and the survival of that strategy is determined by the fitness function and the population frequencies of the players in each group. The two competing behaviors ultimately result in a growth curve describing vaccine uptake, shown in Fig. 1, that starts out rapidly (exponential), then slows down, passing through an inflection point to a fairly stable resting percentage of vaccinated individuals, which in the United States seems to have settled at just under 60% of the population (Fig. 1).

The way to think about the evolutionary dynamics unfolding in a vulnerable unvaccinated population of *players* is to imagine that each individual carries with him/her a complex and ever-changing set of beliefs about their faith in the efficacy of the vaccine that is being introduced. Individuals begin to learn about the vaccine well in advance of the rollout by talking with friends and relatives, and listening and reading news reports. It is widely appreciated that the social dynamics that takes place in this pre-rollout phase is crucial to the ultimate success of

\* Corresponding author.

E-mail addresses: [kstuckey@usc.edu](mailto:kstuckey@usc.edu) (K. Stuckey), [newton@usc.edu](mailto:newton@usc.edu) (P.K. Newton).



**Fig. 1.** Covid-19 vaccine uptake data in the United States (1 dose for Johnson & Johnson or 2 doses for Pfizer/Moderna) starting in January 2021 as is publicly available at <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>. Red curve shows a Gompertzian fit to the data (first derivative curve also in red), blue curve shows the results of the Hawk–Dove evolutionary game theory model (first derivative curve also in blue). Vertical lines mark the inflection point (maxima of the first derivative) where vaccine uptake begins to slow, resulting in an asymptote (Nash equilibrium) at roughly 58% of the population. Upper left inset shows the phase-plane diagram for the Hawk–Dove dynamical system with an internal evolutionary stable state (ESS) at 58%. How much can well designed incentive programs push this percentage up?

the program, and one can assume that the evolution of each individuals' perception regarding the vaccine is already forming before the actual rollout begins. Once vaccines become available, individual decisions are made over time, and from the initial uniformly unvaccinated population, a subpopulation of vaccine adopters begins to form and grow monotonically (a vaccinated person cannot become unvaccinated). The exact shape of this growth curve depends very much on the relative sizes of the two sub-populations that are forming between the unvaccinated and the vaccinated as the two groups continue to play this evolving game. The frequencies of each of the two groups within the overall population influence the decision process of each remaining unvaccinated person as it is known that unvaccinated people surrounded by a sea of vaccinated are much more likely to choose to be vaccinated than those who are surrounded by vaccine skeptics.

The question we address in this paper is whether or not a well designed (optimized) punishment/reward system can significantly alter this natural dynamic, and if so, how best to achieve the goal of obtaining a higher percentage of vaccinated people making up a population? Vaccine incentive programs have been utilized with varying degrees of success for other vaccines, but for the Covid-19 vaccine they have largely been local (county-wide and state-by-state) and somewhat haphazard, ranging from small cash rewards handed out at vaccination clinics, medium-sized vaccination add-ons, or larger lottery-style rewards [5,6]. Punishments for the unvaccinated have also been levied, ranging from the small extra hassle of requiring weekly Covid testing, more severe restrictions of not being allowed entry to restaurants or public events, and larger vaccine mandates that require vaccines as part of the employment requirement or school enrollment [7–9]. Table 1 shows a compilation of the mostly ad hoc strategies that have been implemented in states across the country but larger scale national programs have not been systematically designed or implemented.

Specific questions we address in this paper include whether there are inherent limitations to well designed (optimized) punishment/reward systems, if implemented on a wide-scale basis? Are there optimal schedules that can be designed that would work most effectively? What are the optimal starting and ending times for such dynamic incentive programs? Is there a point of diminishing returns where larger incentives are no longer as effective? By modeling the vaccine uptake problem as a reinforcement learning evolutionary game played between two sub-populations of players (the vaccine adopters and the vaccine hesitant), we address these questions within the context of a mathematical model calibrated with vaccine uptake data obtained from the Center for Disease Control both on a nationwide level, and a state-by-state level. With models tailored to individual states and for different age groups, we are able to test various types of incentive schedules to produce upper and lower bounds (using the Pontryagin maximum/minimum principle from optimal control theory) on the inherent limitations of dynamic incentive programs, and by producing incentive/response curves (analogous to chemotherapeutic dose/response curves [10]), we are able to hypothesize likely responses to different types and sizes of the incentive schedules.

Aspects of vaccine policy and individual decision making surrounding these policies have been studied, for example, by Korn et al. [11] who argue that vaccine uptake can be viewed as a social contract where individuals reward others who comply and punish those who do not. Bauch et al. [12] frame the uptake problem in terms of the complex trade-offs between group interests versus self-interest arguing that, in the case of extreme events (such as bio-terrorist attacks), it is unlikely that voluntary vaccination levels alone would reach the group optimal level necessary for obtaining herd immunity. Bauch et al. [13,14] have also used game-theoretical models to help explain human decision-making surrounding vaccine uptake studying how vaccine scares unfold [15]. In [16], they invoke imitation dynamics models to understand the complex interplay between vaccine coverage, disease prevalence, and individual decision making. In a very comprehensive recent book, Tanimoto [17] describes many popular models associated with the spread of epidemics. Chapter 9 addresses the topic of pre-emptive vs. late vaccination strategies, Chapter 10 discusses the flu vaccine uptake problem, while Chapter 11 discusses the optimal design of vaccination subsidy policies. More general modeling frameworks have used tools borrowed from statistical physics in interesting ways [18] to model vaccine dynamics.

Our approach makes use of the vaccine uptake data (country-wide as well as state by state data) available at <https://covid.cdc.gov/covid-data-tracker/#datatracker-home> to fit three Gompertzian parameters ( $a, b, c$ ) and then use these to estimate the entries of the  $2 \times 2$  payoff matrix that describes the evolutionary game played between the vaccine adopters and the vaccine hesitant populations. The data shows that the population is effectively playing a Hawk–Dove game with an evolutionary stable internal fixed point (ESS) representing the percentage of vaccine adopters (Doves) in the population. We then use optimal control theory on this dynamical system to design time-dependent incentive schedules that alter the baseline payoff matrix entries (altering the reward/punishment balance) in order to obtain upper (and lower) bounds on how different incentive strategies can shift the asymptotic percentage of vaccine adopters in the population. This control technique was originally developed for the design of adaptive/optimal chemotherapy schedules for controlling resistance in tumors [19–26]. Here, we exploit the observation that optimizing vaccine incentive schedules is analogous to optimizing chemotherapy schedules to produce dose–response curves [10] for specific goals, such as, for example, avoiding chemotherapeutic resistance [19–22]. The adoption of these techniques to

**Table 1**

Compiled information on different forms of vaccine incentives states have used. States are ranked in order of highest adoption percentage to lowest. Not listed are states where we were not able to obtain information on any incentive programs.

Rank	% Vaccinated	State	One Dose	Fully Vaccinated
2	70.68	Connecticut	Free event admission Concert tickets Free food	Free drinks
4	70.56	Maine	License or event pass	\$1 per person vaccinated
5	69.69	Massachusetts		5 \$1 million prize 5 \$300,000 scholarship
6	66.72	New York	Baseball tickets	State Park pass
7	66.4	New Jersey	Free Beer	State Park Vax Pass Dinner with governor
8	66.18	Maryland	\$100 for state employees	\$2 million lottery
9	63.48	Washington	Lottery tickets	
11	62.88	Oregon	\$1 million prize 36 \$10,000 prizes 5 \$100,000 scholarships	
14	62.14	New Mexico	5 wkl \$250,000 prize \$5 million prize \$100 10 prize wheels Travel prize	
15	61.64	Colorado	Weekly lottery \$50,000 tuition	\$500 for CDOC
16	61.31	California	10 \$1.5 million prize 30 \$50,000 prizes \$50 gift card 6 vacations	
18	60.42	Illinois		50,000 six flags tickets 3 \$1 million prizes 40 \$100,000 prizes
19	59.9	Minnesota	Free/discounted drinks	
20	59.81	Hawaii	Travel perks	
22	59.73	Delaware	Inmate incentives Scholarship raffle Vacation passes	\$302,000 prize Free drinks
27	53.54	Michigan	\$5 million in cash prizes	
31	52.98	Nevada	"Vax Nevada Days"	
32	52.67	North Carolina	4 \$1 million prizes \$25 cash cards	
35	51.88	Ohio	\$1 million drawings	
36	50.72	Kentucky	Lottery Tickets 3 \$1 million prizes 15 scholarships	
40	49.87	Indiana	Girl scout cookies	
43	48.04	Arkansas	\$20 lottery tickets	\$100 for state employees
44	47.71	Louisiana	State park access \$100,000 prizes	Free drinks
45	47.66	Tennessee	Car sweepstakes	
47	44.78	Alabama	Talladega Sweepstakes \$250 gift cards	
48	43.89	Wyoming	"Shots for swag"	
49	43.92	Idaho	4 hr paid leave	
50	41.04	West Virginia	\$100 gift cards \$1.58 million prize	

vaccine incentive scheduling presents a different set of questions and challenges but can be addressed within a similar modeling framework. Other recent work that makes use of feedback control ideas to develop COVID-19 policies includes [27]. While the merging of reinforcement learning models with optimal/adaptive control theory is a new and promising field with many potential applications, a nice introduction to the field, described mostly in the robotics framework, can be found in a recent monograph [28].

## 2. The vaccine uptake model

### Calibrating the Gompertzian curves

The vaccine uptake curve shown in Fig. 1 is a three parameter ( $a, b, c$ ) Gompertzian curve,

$$f(t) = a \exp[-\exp(b - ct)] \tag{1}$$

which has a long history of use in actuarial sciences (laws of human mortality), economics (growth laws of wealth), biology

**Table 2**  
Model parameters for different population groups.

Population	$T$ (days)	$f'(0)$	$f'(T)$	$a$	$b$	$c$	$a_{21}$	$a_{11}$	$a_{12}$
US population	88	0.028256504	0.405499161	57.71	1.675	0.0191	0.034	0.02	0.0191
18 – 39 year olds	106	0.008343649	0.382741771	57.8	1.907	0.018	0.032	0.019	0.018
40 – 64 year olds	90	0.004432806	0.597362637	70.6	2.077	0.023	0.033	0.023	0.023
65+ year olds	58	0.015256548	1.024397092	81.9	1.971	0.034	0.043	0.035	0.034
Connecticut	86	0.012707894	0.563297	69.6	1.901	0.022	0.032	0.022	0.022
Vermont	89	0.002882054	0.612739997	69.4	2.14	0.024	0.035	0.024	0.024
Idaho	77	0.030170215	0.327559854	42.4	1.608	0.021	0.051	0.022	0.021
West Virginia	63	0.065387774	0.331017921	40.9	1.389	0.022	0.055	0.0231	0.0221

(population growth and saturation), cancer (tumor growth) [29,30]. Key parameters for fitting such a curve to this data are: (1)  $T$ : the location of the inflection point (shown in Fig. 1); (2)  $f'(T)$ : the slope of the tangent line at the inflection point (the growth rate when the growth curve changes from concave up to concave down); (3)  $f'(0)$ : the slope of the tangent line at the origin (initial growth rate). In Eq. (1),  $a$  is the asymptote ( $\lim_{t \rightarrow \infty} f(t)$ ) (also known as carrying-capacity in other contexts [31,32]),  $b$  is the displacement along the  $t$ -axis (time-shift parameter), and  $c$  is a time-scaling factor. In terms of those parameters, the inflection point is located at  $T$ , where:

$$T = b/c, \quad (2)$$

is the ratio of time-shift to time scaling parameters, while the slope of  $f(t)$  at the origin and inflection points are given by:

$$f'(0) = ac \exp(b - \exp(b)), \quad (3)$$

$$f'(T) = ac / \exp(1). \quad (4)$$

Inverting (2)–(4) for  $(a, b, c)$  gives:

$$a = \exp(1)f'(T)T/b, \quad (5)$$

$$c = b/T, \quad (6)$$

$$b - \exp(b) = \ln(f'(0)/\exp(1)f'(T)). \quad (7)$$

The transcendental equation (7) can be solved numerically to give the value of  $b$ , and the results are shown in Table 2, and corresponding curve in Fig. 1 (in red) for the US population. For this, the asymptote is roughly 58% vaccinated, and the uptake inflection point is  $T \approx 88$  days from when vaccines first became available. Error bars in Fig. 1 are produced using a stochastic process that governs the evolutionary game dynamics in finite populations [33–35], which in the limit of large populations converges to the deterministic problem. The stochastic simulation is run and at each time step, the relative fitness of each population is used to calculate the probability of birth and the probability of death. These probabilities are sampled and the proportion of individuals in each population is updated accordingly. This process is repeated 10,000 times. The error bars show one standard deviation around the mean.

We want to emphasize the importance of the inflection point time  $T$  in our approach which we use to set the basic timescale over which we optimize. An optimization cycle of  $T/4$  was chosen to be short enough to allow for relatively frequent changes in the incentive schedules if necessary, but long enough for an optimized schedule to have a reasonable impact. Each  $T/4$  cycle is optimized individually, then longer periods  $nT/4$  (where  $n$  is an integer) are optimized sequentially using the final value of the previous cycle as the initial value for the next. Table 2 provides a summary of all of the model parameters we use for the different sub-populations. With these parameters, we develop the reinforcement learning dynamical system.

### The reinforcement learning/replicator model

We use the replicator dynamics equations from evolutionary game theory as our reinforcement learning model for vaccine uptake dynamics between the two populations  $x_A$  (vaccine adopters) and  $x_H$  (vaccine hesitant), where each represents a proportion of the entire population,  $\vec{x}(t) = (x_1, x_2)^T \equiv (x_A, x_H)^T$ ;  $x_A + x_H = 1$ . The essential feature of replicator dynamics is that people (reinforcement learners) copy others, and successful strategies get replicated more frequently than unsuccessful strategies [36] thereby spreading throughout the population. As discussed earlier, initially, none of the players are committed to just one way of behaving, but retain several potential ways of behaving simultaneously. Which behavior predominates depends on the experiences at the individual level. At the population level, the process operates analogously to biological evolution governed by the replicator dynamical system. This makes the model useful not only in contexts where biological evolution by natural selection (due to competition) is prominent [37], as cells and organisms with higher fitness (measured by their ability to replicate) more often pass along their genetic characteristics in the population, but also in any reinforcement learning setting where learners copy successful strategies [38] more often than unsuccessful ones (success begets success and failure spirals downward), with success determined by fitness level. The attractiveness of this framework in the present context is that it has been widely documented that vaccine uptake is more common in a positive uptake environment, and less common in settings where fewer people choose to get vaccinated. This dynamic is also the hallmark of a reinforcement learning process where people interact, learn strategies from others, receive payoffs (in the form of advantage or disadvantage) based on strategies they adopt, the payoffs determine the fitness (ability to survive) of those strategies in the overall population, the fitness controls the survival probability of the strategy.

To formulate the dynamical system, we use:

$$\dot{x}_A = x_A(f_A - \langle f \rangle), \quad (8)$$

$$\dot{x}_H = x_H(f_H - \langle f \rangle). \quad (9)$$

Here,  $f_A$  and  $f_H$  denote the fitness of the vaccine adopters and the vaccine hesitant populations, while  $\langle f \rangle$  denotes the average fitness of the entire population under consideration. The system simply says that the growth rate of each sub-population ( $\dot{x}_A/x_A$ ;  $\dot{x}_H/x_H$ ) is governed by the difference between the fitness of that population and the overall average fitness of both populations. The more each sub-population fitness deviates from the average (either above or below), the larger/smaller the instantaneous growth rate is of that strategy in the population. This models gradual evolution (as contrasted, for example, with catastrophism), i.e. behavior changes occur gradually as is the

case with vaccine uptake dynamics. The fitnesses are defined via a  $2 \times 2$  payoff matrix  $A$  as:

$$f_A = (A\vec{x})_1 = a_{11}x_A + a_{12}x_H, \quad (10)$$

$$f_H = (A\vec{x})_2 = a_{21}x_A + a_{22}x_H, \quad (11)$$

$$\langle f \rangle = \vec{x}^T(A\vec{x}) = x_A f_A + x_H f_H, \quad (12)$$

with:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (13)$$

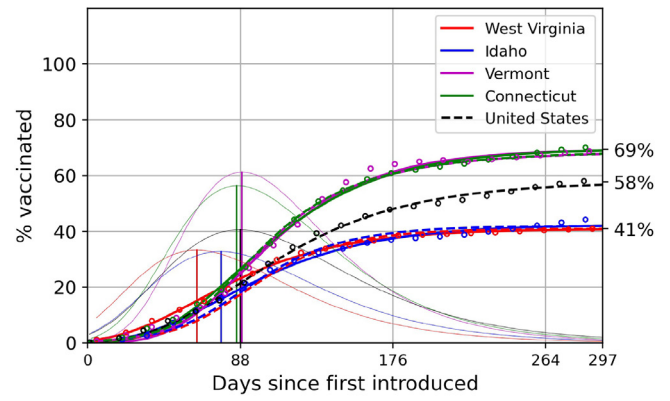
which defines the evolutionary game being played as determined by the CDC data. The four entries of this matrix encode the punishment-reward balance (i.e. payoffs) associated with competition between the two groups and is the heart of the model. As described in [39], the payoffs are decided by many complex factors, including each person's perceived risk of infection (which can vary in time), the severity of the disease (which certainly varies in time), perhaps measured in hospitalization rates, financial costs of vaccinations, and also the perceived uptake of vaccinations by others. Increasing/decreasing either of the entries of the top row of  $A$  increases/decreases the fitness of the vaccine adopter population, whereas increasing/decreasing either of the entries of the bottom row of  $A$  increases/decreases the fitness of the hesitant population. Without loss, we can choose  $a_{22} = 0$ , while the remaining three entries (shown in Table 2) can be obtained as functions of  $(a, b, c)$  (which were optimally fit to the data) via Eqs. (5)–(7). Thus, all of the complexities associated with the many decision processes which result in the Gompertzian uptake curves are neatly packaged into three of the four entries of the payoff matrix determining the evolutionary game which unfolds between the vaccine adopters and the vaccine hesitant.

It is a non-trivial result of this fitting process that the payoff matrix corresponds to a Hawk–Dove evolutionary game, where the vaccine adopters are the Doves, and the vaccine hesitant are the Hawks. This is based on the inequalities:  $a_{21} > a_{11} > a_{12} > a_{22} = 0$ . A key feature of a Hawk–Dove evolutionary game is the existence of an internal  $\in (0, 1)$  ESS (Nash equilibrium), which we denote by  $a$  (the asymptote of  $f(t)$ ). As shown in Fig. 1, for the US population as a whole,  $a \approx 0.58$ . Fig. 2 shows the data, curve fit, and replicator dynamic model for the two states with the highest vaccine uptake percentages (Vermont and Connecticut  $\approx 69\%$ ), and the two lowest (West Virginia and Idaho  $\approx 41\%$ ). Fig. 2(a) shows the results of the vaccine uptake data along with both the Gompertzian curve fit and the replicator dynamics model for the four states, along with the entire US population. In Fig. 2(b) we break the US data into three different age groups (18–39; 40–64; 65+) in a similar way that the vaccine rollout prioritized these groups. This is reflected in the leftward shift of the curve corresponding to the older compared to younger groups, with the oldest population showing the steepest uptake curve consistent with the notion that this group was among the most eager to be vaccinated. The inflection point for the US population as a whole is roughly at  $T \approx 88$  days which we take as a benchmark for scaling time when we implement our control strategy on this group. Similarly, for all other subgroups, we use the corresponding inflection point location associated with that subgroup (see Table 2).

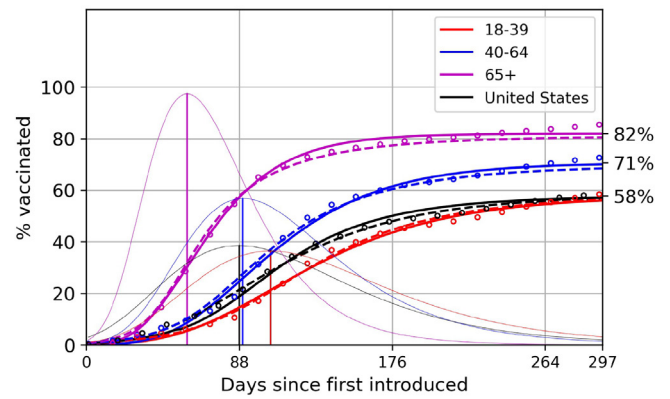
#### Shifting vaccine uptake curves with time-dependent payoffs

To implement an optimal vaccine incentive strategy, we now consider the time-dependent payoff matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A_0 + A_1(t) \quad (14)$$



(a)



(b)

**Fig. 2.** Vaccine uptake curves for different subgroups. Dots (data), Dashed (Gompertzian fit), Solid (replicator model). Vertical lines mark inflection points at maximizers of the derivative curves. (a) United States (dashed black); Connecticut (green); Vermont (purple); West Virginia (red); Idaho (blue). (b) Three different age groups: 18–39 (red), 40–64 (blue) and 65+ (purple).

$$= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} 0 & u_1(t) \\ u_2(t) & 0 \end{bmatrix} \quad (15)$$

$$= \begin{bmatrix} a_{11} & a_{12} + u_1(t) \\ a_{21} + u_2(t) & a_{22} \end{bmatrix}, \quad (16)$$

where  $A_1(t)$  represents our control with entries in the off-diagonal terms (without loss of generality), and  $A_0$  is the baseline Hawk–Dove payoff matrix as obtained from the vaccine uptake data. The time-dependent controllers  $\vec{u}(t) = (u_1(t), u_2(t)) \in \mathbb{R}^2$  are bounded above and below (based on an incentive size parameter  $p$ ):

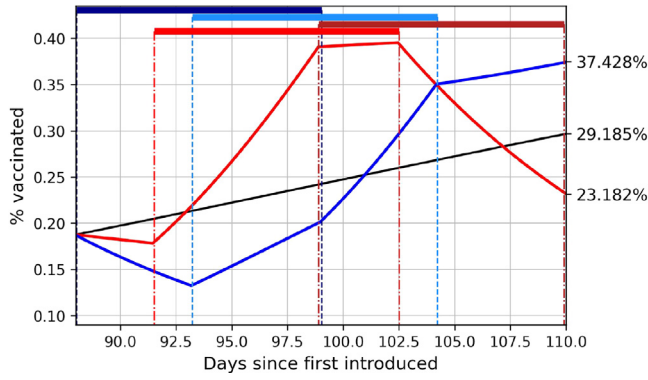
$$-\frac{ap}{a_{12} + a_{21}} \leq u_1(t), u_2(t) \leq \frac{ap}{a_{12} + a_{21}} \quad (17)$$

( $0 \leq p \leq 1$ ) and a global constraint on the incentive schedule,  $\bar{U}(t) = \int_0^t \vec{u}(\tau) d\tau = \text{const.}$  is enforced, all of which play a role in determining the detailed outcome of the optimization procedure.

### 3. Methods

To implement the Pontryagin maximum (minimum) principle with boundary value constraints in order to compute upper (maximum principle) and lower (minimum principle) bounds, we follow standard methods [40] and denote:

$$\vec{X} = [\vec{x}(t), \vec{U}(t)]^T, \quad \vec{X} \in \mathbb{R}^4 \quad (18)$$



**Fig. 3.** Optimal schedules (top horizontal bars indicating incentives on) and response curves (underneath) associated with the US population. Black curve is the uncontrolled subpopulation, blue curve is the maximized population, red curve is the minimized population. In this case, the incentive schedules start at the inflection point of the uptake curve, which is roughly at  $T \sim 88$  days, with  $\sim 18\%$  of the population vaccinated. The dependent variables in our system are optimized through one fundamental cycle time of length  $T/4 \sim 22$  days using only an endpoint cost (Meyer problem). Dark blue/red color corresponds to the  $u_1$  schedule, light blue/red color corresponds to the  $u_2$  schedule in Eq. (16).

$$\dot{\vec{X}} = \vec{F}(\vec{X}) = [\dot{\vec{x}}, \dot{\vec{U}}(t)]^T, \quad \vec{F} : \mathbb{R}^4 \rightarrow \mathbb{R}^4 \quad (19)$$

where we would like to minimize or maximize a mathematical cost function:

$$J[\vec{x}(\cdot), \vec{u}(\cdot), t_0, t_f] = \int_{t_0}^{t_f} L(\vec{x}(t), \vec{u}(t), t) dt + \varphi[\vec{x}(t_0), t_0, \vec{x}(t_f), t_f] \quad (20)$$

over times from  $t_0$  to  $t_f$ . The first term on the right is called the running cost, while the second is called the endpoint cost. Since we are optimizing the endpoint cost,  $\varphi[\vec{x}(t_0), t_0, \vec{x}(t_f), t_f] = x_1 = x_A(t_f)$  only (i.e. the asymptotic vaccine acceptance value), we take  $L = 0$  (called a Meyer problem [40] developed in the context of missile guidance problems where final distance from the target is minimized). We briefly describe the basic

framework and refer readers to [40] for more details on how to implement the approach. In particular, we construct the control theory Hamiltonian:

$$H(\vec{x}(t), \vec{U}(t), \vec{\lambda}, \vec{u}(t)) = \vec{\lambda}^T \vec{F}(\vec{x}) + L(\vec{x}, \vec{u}(t), t) \quad (21)$$

where  $\vec{\lambda} = [\lambda_1, \lambda_2, \mu_1, \mu_2]^T$  are the co-state functions (i.e. momenta) associated with  $\vec{x}$  and  $\vec{U}$  respectively. Assuming that  $\vec{u}^*(t)$  is the optimal control for this problem, with corresponding trajectory  $\vec{x}^*(t), \vec{U}^*(t)$ , the canonical equations satisfy:

$$\dot{x}_i^*(t) = \frac{\partial H}{\partial \lambda_i^*} \quad (22)$$

$$\dot{U}_i^*(t) = \frac{\partial H}{\partial \mu_i^*} \quad (23)$$

$$\dot{\lambda}_i^*(t) = -\frac{\partial H}{\partial x_i^*} \quad (24)$$

$$\dot{\mu}_i^*(t) = -\frac{\partial H}{\partial U_i^*} \quad (25)$$

where  $i = (1, 2)$ . The corresponding boundary conditions are:

$$\vec{x}^*(t_0) = \vec{x}_0 \quad (26)$$

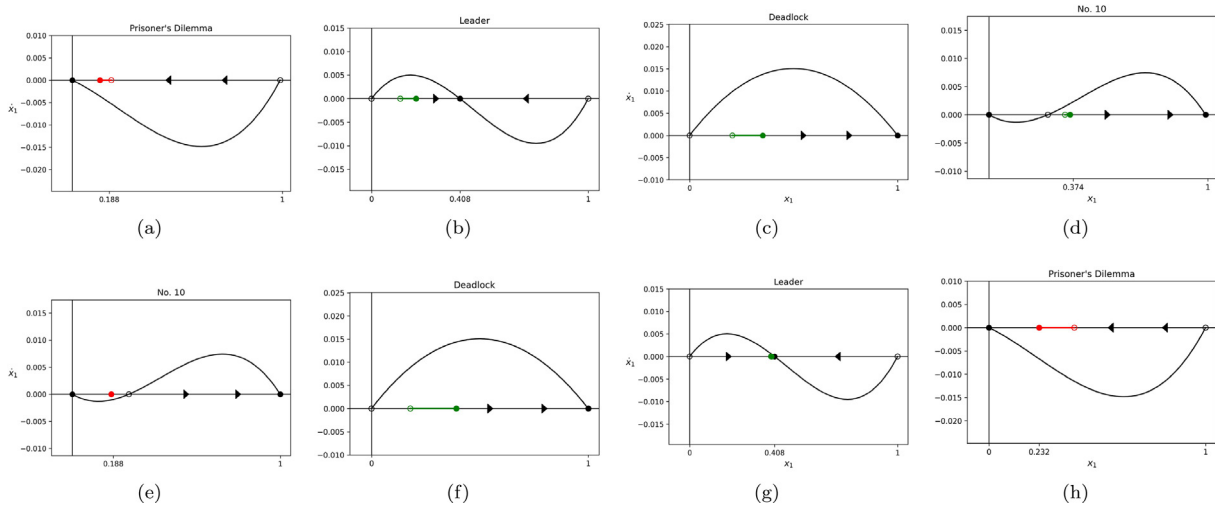
$$\vec{U}^*(t_0) = 0, \vec{U}^*(t_f) = \vec{U}_f^* \quad (27)$$

$$\lambda_i^*(t_f) = \frac{\partial \varphi(\vec{x}(t_f))}{\partial x_i^*(t_f)} \quad (28)$$

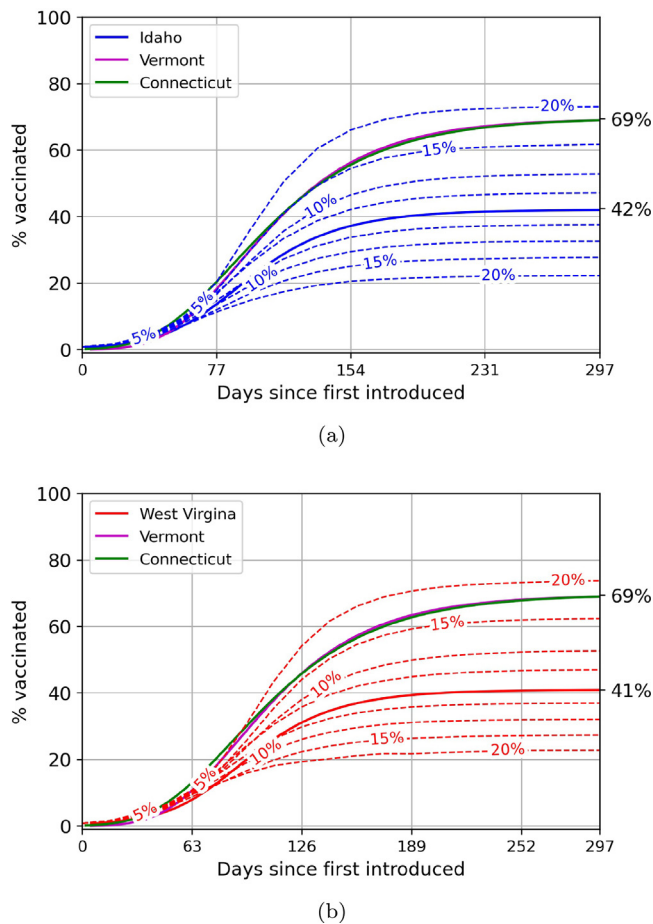
Then, at any point in time, the optimal control  $\vec{u}^*(t)$  will minimize the control theory Hamiltonian:

$$\vec{u}^*(t) = \arg \min_{\vec{u}(t)} H(\vec{x}^*(t), \vec{U}^*(t), \vec{\lambda}^*(t), \vec{u}(t)) \quad (29)$$

The optimization problem becomes a two-point boundary value problem (using (26)–(28)) with unknowns  $(\lambda_2^*(t_0), x_2^*(t_f))$  whose solution gives rise to the optimal trajectory  $\vec{x}^*(t)$  (from (22)) and the corresponding control  $\vec{u}^*(t)$  that produces it, as shown, for example, in Fig. 3. In practice, we solve the optimization problem numerically using the dynamic optimization software GEKKO. Documentation for the software can be found at <https://gekko.readthedocs.io/>.



**Fig. 4.** Phase portraits associated with US population with incentive schedules starting at the inflection point of the uptake curve as shown in Fig. 3. In each of the four legs of the cycle (see Fig. 3), a different evolutionary game is being played, which drives the system along an optimal path. The dynamics along the  $x_1$  horizontal axis proceeds from the initial open circle to the closed one. (a) Phase diagram of vaccinated population through first leg of the maximizing schedule (Prisoner's dilemma game); (b) Phase diagram of vaccinated population through second leg of the maximizing schedule (Leader game); (c) Phase diagram of vaccinated population through third leg of the maximizing schedule (Deadlock game); (d) Phase diagram of vaccinated population through fourth leg of the maximizing schedule (Game No. 10); (e) Phase diagram of vaccinated population through first leg of the minimizing schedule (Game No. 10); (f) Phase diagram of vaccinated population through second leg of the minimizing schedule (Deadlock game); (g) Phase diagram of vaccinated population through third leg of the minimizing schedule (Leader game); (h) Phase diagram of vaccinated population through fourth leg of the minimizing schedule (Prisoner's dilemma).



**Fig. 5.** Maximizing and minimizing vaccination percentages for state with lowest vaccination rates, West Virginia and Idaho, in comparison with states with the highest vaccination rates, Vermont (purple) and Connecticut (green). (a) Solid blue curve depicts Idaho's natural vaccine uptake curve. Dashed curves show the Idaho optimized model with upper bounds using 5%, 10%, 15%, 20% incentive sizes (normalized using baseline value) and lower bounds using 5%, 10%, 15%, 20% incentive sizes (normalized using baseline value); (b) Solid red curve depicts West Virginia's natural vaccine uptake curve. Red dashed curves show the West Virginia optimized model with upper bounds using 5%, 10%, 15%, 20% incentive sizes (normalized using baseline value) and lower bounds using 5%, 10%, 15%, 20% incentive sizes (normalized using baseline value).

#### 4. Results

As a first important example, we show in Fig. 3 an optimized result using the US population curve as a baseline, with optimized schedules for the controllers  $\vec{u}(t) = (u_1(t), u_2(t))$  shown as horizontal bars at the top, and the maximized (blue) and minimized (red) values of the % vaccinated at the end of one cycle  $T/4$ . The optimized incentive schedule, which begins at the inflection point of the Gompertzian uptake curve, is able to push the vaccinated population up to a value of 38%, as compared to the unincentivized value of 29% that would have naturally occurred in the absence of any incentives. The control schedule is bang–bang (off/on), with the bars showing the time each of the controllers is on.

Associated with the schedules and responses shown in Fig. 3, we show a complementary and useful interpretation of our method in Fig. 4 in terms of the phase portraits ( $\dot{x}_1$  vs.  $x_1$ ) that correspond to the optimized trajectories through the four separate time-regions defined from the off/on schedules in Fig. 3. Figs. 4(a)–(d) show the maximized dynamics in the four regions along with the name of the evolutionary game defined by the controlled

payoff matrix in each regime. On the first leg (Fig. 4(a)), the subpopulation  $x_1$  actually decreases while a Prisoner's dilemma game is the governing matrix. But then in the next three legs (Fig. 4(b), (c), (d)), the games are changed (Leader-Deadlock-Game #10), and the endpoint (shown by the blue curve in figure 3 is pushed up to its maximally achievable value of  $\sim 37\%$ . In Figs. 4(e)–(h) we show the corresponding four legs associated with the minimization procedure from Fig. 3. Here, the sequence of evolutionary games cycle through the reverse order as in the maximization procedure (Game #10-Deadlock-Leader-Prisoner's Dilemma) to push the vaccinated population down to its lowest value of  $\sim 23\%$  at the end of the optimization cycle. Note that in all cases, the exact switching times from one game to another are determined from the outcome of the optimization procedure, and depends on the starting value  $\vec{x}_1(t_0)$ . We mention a recent paper by Tanimoto [41] which also discusses the connection between scaling the social dilemma strengths and its affect on the resulting phase plane dynamics which is related to this interpretation.

We now use our optimization method to answer several specific questions that give insight into how well an optimized incentive vaccine rollout program can perform.

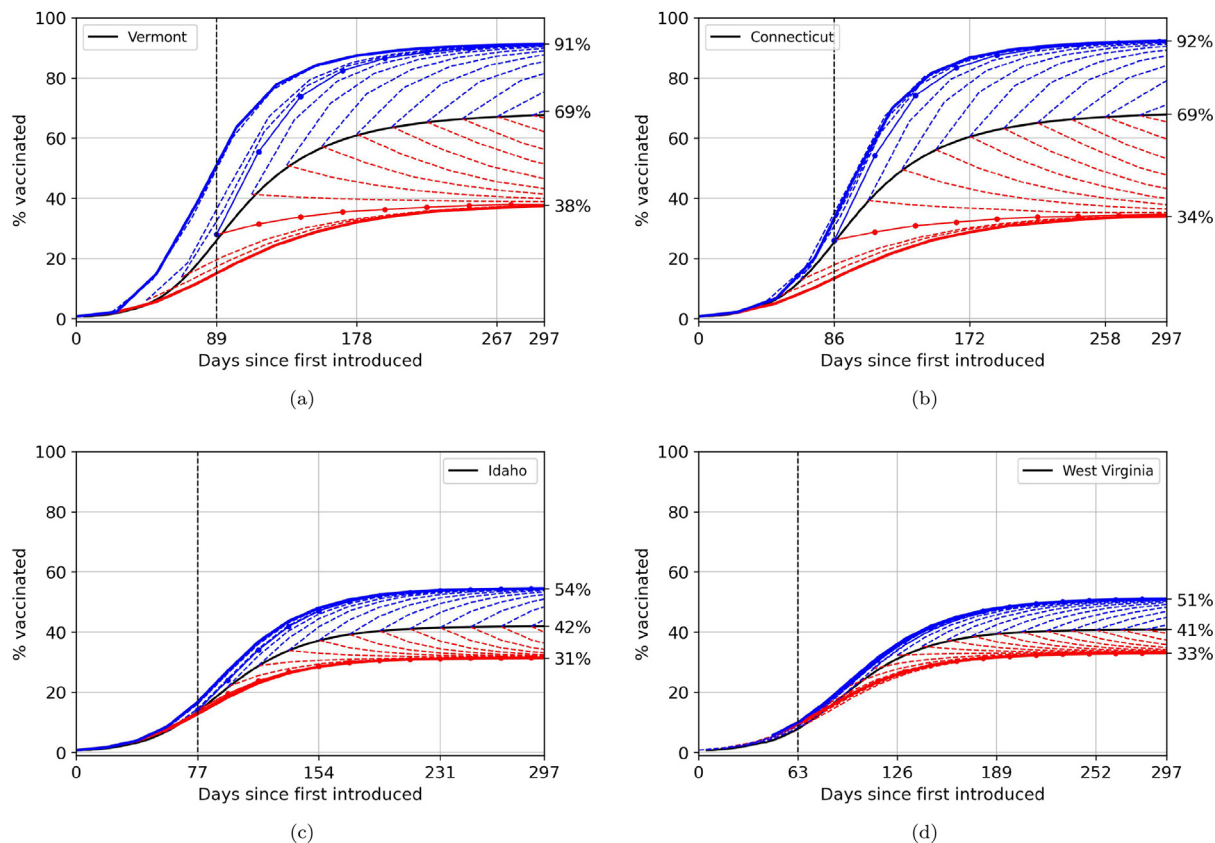
#### State-by-state results

The first question we address using our optimized incentive model, is whether it is possible to incentivize the states with low vaccine uptake curves (West Virginia and Idaho) to bring them up to the level of states with high uptake curves (Vermont and Connecticut). Fig. 5(a) shows the result of our simulations for Idaho. With relatively large incentive sizes roughly between 15%–20% (measure normalized by the baseline value), we show this is possible. But we consider this range of incentive sizes to be so large that the price of implementing them might be prohibitive. Fig. 5(b) shows our simulations for West Virginia, with the same general conclusions as Idaho. Incentive sizes this large can have a considerable effect, but the price would be high to implement them. At this point we have not included any running penalty in our cost function which could account for incentive sizes.

#### Optimal timing

We next address whether or not the initial start-time of our optimal incentive schedule has much impact on the end result. The short answer to this is no, it does not, as shown in Fig. 6 panel for the (a) Vermont population, (b) Connecticut population, (c) Idaho population, and (d) West Virginia population. In all cases, the incentivized curves (dashed) asymptote to the absolute max/min curves no matter when the schedules begin. This indicates that we could begin the schedules at the inflection point of the uptake curves, allowing us to collect and develop the model in real time as the uptake dynamics unfolds, designing the optimal incentive schedules to use going forward. The one caveat with this approach is that although the curves all reach the same asymptote, if time is of the essence (say because of high death rates in the unvaccinated population), there could well be advantages to starting the incentive schedules as early as possible. To design optimal schedules in real time before reaching the inflection point of the uptake curve would require a separate careful forecasting model based only on earlier data.

In Fig. 7 we highlight both the results of the incentive schedules as well as the optimized results starting at three different times. The time-windows are shown in Fig. 7(a) in black rectangles, before the inflection point, at the inflection point, and after the inflection point. Corresponding optimized results with schedules are shown in Figs. 7(b), (c), (d). The plots show that



**Fig. 6.** Maximizing and minimizing vaccination uptake with controllers turning on at different times. All plots show incentive schedules beginning at  $t_0 = \frac{T}{4}n$  for increasing  $n$ . In all cases, the incentivized schedules reach upper (blue) and lower (red) asymptotes, indicating relative insensitivity of the optimal outcome to start times. (a) Vermont unincentivized uptake curve (black),  $T = 88$  days. Blue shows optimized upper bounds using incentive schedules, red shows lower bounds using incentive schedules; (b) Connecticut unincentivized uptake curve (black),  $T = 86$  days. Blue shows optimized upper bounds using incentive schedules, red shows lower bounds using incentive schedules; (c) Idaho unincentivized uptake curve (black),  $T = 77$  days. Blue shows optimized upper bounds using incentive schedules, red shows lower bounds using incentive schedules; (d) West Virginia unincentivized uptake curve (black),  $T = 63$  days. Blue shows optimized upper bounds using incentive schedules, red shows lower bounds using incentive schedules.

incentive schedules are generally more effective at pushing the final value up for the later time-windows, which is somewhat counter-intuitive. We interpret this result to mean that incentives are most effective when rolled out into a population of unvaccinated people who make up a smaller part of a larger population with a relatively high percentage of vaccinated individuals. The unvaccinated are more receptive to the incentives when they are surrounded by people who have already been vaccinated. This is what the model predicts, but whether or not this pans out in an actual vaccine rollout would need to be tested, as there might well be other factors at play not considered in our model.

### Incentive-response curves

We now address the question of what incentive size leads to the best response? Fig. 8 shows the percent shift in the asymptote (US population) for different size incentives, both upper (blue) and lower (red) bounds. With no incentive (0%), the asymptote remains at 58% of the population, as expected. In general, the larger the incentive, the larger the response.

Notice that for higher percentage incentives, the curves rise much faster than for low, which means the optimized results will be achieved more quickly. But it is also important to point out that even for lower incentives, the upper asymptote will eventually be reached. The bottom line here is that, as would be expected, it seems that the smaller the incentive, the longer it takes to achieve the desired result.

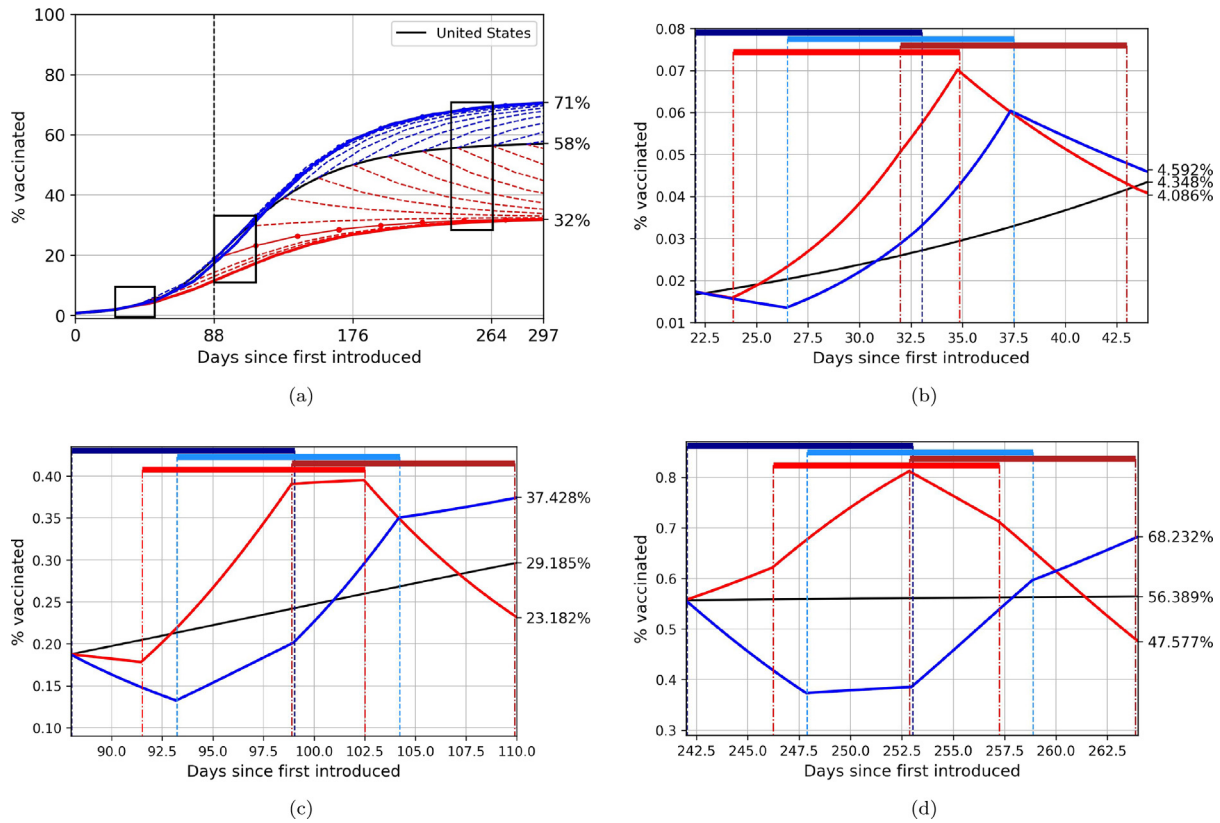
### Diminishing returns

Is there a point of diminishing returns on implementing larger incentives, after which the response diminishes? Fig. 9 shows an incentive-response curve for the US population. Our model produces a curve (data points fit to three-parameter Gompertzian curve) depicting the incentive strength (abscissa) versus the change in asymptote (ordinate). For incentive strengths below 11%, the curve is concave up, indicating a better response with higher incentives. Above 11%, however, the curve is concave down, indicating a weaker response to higher incentives. We can think of this threshold value ( $\sim 11\%$ ) as a point of diminishing returns. This is in many ways analogous to dose-response curves in chemotherapy settings [10] where past a threshold, increasing the dose further shows a diminished response. This leads to a threshold value of optimal incentive size, which our model predicts is roughly 11%. We place more value in showing that such a threshold exists in our model, than the actual threshold value, which can be tricky to pin down accurately without more detailed analysis.

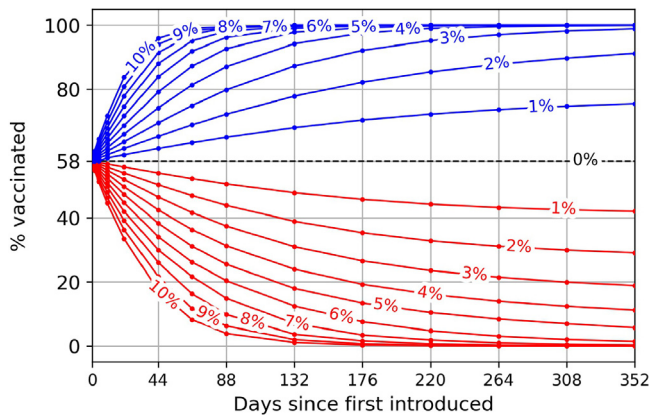
## 5. Discussion

Although it is presumably unrealistic to assume that optimality will actually be achievable in practice, optimal control nonetheless gives clear upper and lower bounds on what is theoretically possible in an ideal setting. But there are several tangible ways the model could be improved. First, we make the



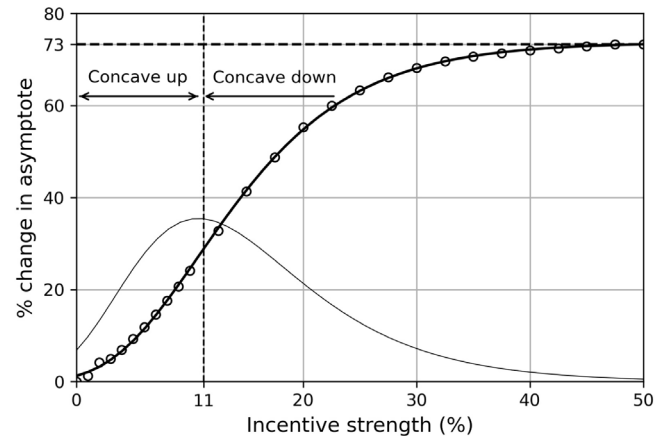


**Fig. 7.** Optimal schedules and optimal responses for early, medium, and late times. Blue schedules and curves maximize the % vaccinated at the end of one cycle time. Red schedules and curves minimize the % vaccinated at the end of one cycle time. Dark blue/red color corresponds to the  $u_1$  schedule, light blue/red color corresponds to the  $u_2$  schedule. Black curve is the uncontrolled response curve. (a) Three black rectangles show early, medium, and late time regions over which we introduce the first optimization cycle; (b) Max and min schedules and responses for early time window; (c) Max and min schedules and responses for medium time window at the Gompertzian inflection point; (d) Max and min schedules and responses for late time window.



**Fig. 8.** Absolute maximizers (blue) and minimizers (red) for a range of incentive intensities using US population model. Cycle times  $44 \text{ days} \leq T \leq 352 \text{ days}$ .

simplifying assumption that response times to incentives are instantaneous. Building in finite-time responses (i.e. time delays) would make the model more realistic. Second, the hesitant population could be further sub-divided into groups, such as hesitant but willing, hesitant and unwilling, with incentives influencing each of those groups differently. This would lead to a higher dimensional model with more complexity but perhaps higher fidelity. Third, the model assumes what is called a well-mixed population (i.e. no spatial structure). A spatially dependent model would be significantly more complex but has the potential to be more targeted geographically. Finally, the psychological aspects of



**Fig. 9.** Incentive-response curve fit to data. Abscissa indicates the incentive size (measured as % of baseline). Ordinate shows the shift in the asymptote (measured as % of baseline). Also shown is the first derivative of the response curve. Response curve is concave up for incentive strengths below 11%, and concave down for larger incentives indicating diminishing returns in terms of response.

how people, states, groups, respond to different incentives are not considered in our model. Matching the size of our controllers with actual incentives/punishments would best be handled by experts in human psychology and is not addressed in our approach. Additionally, the size of the incentives (in terms of financial cost) are not part of our cost function in the optimization procedure which may well be desirable in future model improvements by including

a running penalty term to our mathematical cost function (20) which takes into account incentive size.

Two strengths of our model we would like to emphasize are that only data up until the inflection point needs to be used, and starting the incentives after that point will ultimately lead to the same shift in the asymptotic percentage of vaccinated people as would have happened if the incentives started earlier. It is not a priori clear whether or not nationwide, state-wide, or even more localized data is most useful, but models that use more localized information (at least state-wide) would probably be more useful as it seems probable that different regions of the country would respond differently to different kinds of incentives.

## 6. Conclusion

Every vaccine rollout associated with each new epidemic will have its own natural uptake curve, depending essentially on the complex nature of the interactions between the vaccine adopters and the vaccine hesitant populations, and also the interactions within each group, all of which are nicely encoded as elements of the payoff matrix as determined by the data. But it is not unreasonable to speculate that they should all commonly follow the general form of a three-parameter Gompertzian, with different parameters in each case, and geographic location (targeted population), but of the same universal form. This general form is an outcome of the fact that there are, generally speaking, early adopters, followed by a population of players who decide to adopt as time proceeds, leaving only the most hesitant who remain towards the later stages of a rollout. As a vaccine rollout unfolds, the key parameters to obtain from the vaccine uptake curve are: (i) the initial rate of uptake (which we write as % of the relevant population per day), (ii) the inflection point location on the uptake curve (i.e. when uptake begins to slow down), and (iii) the slope of the tangent line at the inflection point (rate of uptake). As long as reliable data is available up until the inflection point, the reinforcement learning model described can then be developed and calibrated in real-time. When the uptake rate begins to slow (i.e. at or near the inflection point), using the controlled replicator dynamical system model, vaccine incentive schedules can be optimized going forward and likely responses can be predicted from the dose–response curves produced by the model. A recent review paper [39] has highlighted the importance and need for using game theory and mathematical models in designing vaccine policy, which we enthusiastically endorse and feel is an under-utilized tool in the arsenal of developing science-based decision making during an infectious outbreak. The framework developed here, which allows the mathematical models to be tailored to specific settings, offers the possibility of testing different strategies in real time for many different scenarios and is flexible, generalizable, relatively simple, and potentially actionable.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data is publicly available and links are provided.

## Acknowledgments

We gratefully acknowledge support from the Army Research Office, United States of America MURI Award #W911NF1910269 (2019–2024). All authors approved the version of the manuscript to be published.

## References

- [1] N. MacDonald, Vaccine hesitancy: Definition, scope and determinants, *Vaccine* 33 (2015) 4161–4164.
- [2] S. Quinn, A. Jamison, J. An, G. Hancock, V. Freimuth, Measuring vaccine hesitancy, confidence, trust and flu vaccine uptake: Results of a national survey of white and African American adults, *Vaccine* 37 (2019) 1168–1173.
- [3] J. Smith, M. Lipsitch, J. Almond, Vaccine production, distribution, access, and uptake, *Lancet* 378 (2011) 428–438.
- [4] J. Eskola, P. Duclos, M. Schuster, N. MacDonald, How to deal with vaccine hesitancy? *Vaccine* 33 (2015) 4215–4217.
- [5] S. Higgins, E. Klemperer, S. Coleman, Looking to the empirical literature on the potential for financial incentives to enhance adherence with COVID-19 vaccination, *Preventive Med.* 145 (2021) 106421.
- [6] B. Acharya, C. Dhakal, Implementation of state vaccine incentive lottery programs and uptake of COVID-19 vaccinations in the United States, *JAMA Network OPEN* 4 (12) (2021) e2138238.
- [7] M. McArdle, Harsher and harsher punishments are not the way to convince the unvaccinated, *Washington Post* Dec. 15 (2021).
- [8] A. Jena, C. Worsham, Facts alone aren't going to win over the unvaccinated. This might, *New York Times* Dec. 21 (2021).
- [9] H. Thorp, Colleges need vaccine mandates, *Science* 373 (6553) (2021) 369.
- [10] J. Martin, S. Dimmitt, The rationale of dose-response curves in selecting cancer drug dosing, *Br. J. Clin. Pharmacol.* 85 (2019) 2198–2204.
- [11] L. Korn, R. Böhm, N. Meier, C. Betsch, Vaccination as a social contract, *Prod. Nat'L Acad. Sci.* 117 (26) (2020) 14890–14899.
- [12] C. Bauch, A. Galvani, D. Earn, Group interest versus self-interest in smallpox vaccination policy, *Prod. Nat'L Acad. Sci.* 100 (18) (2003) 10564–10567.
- [13] C. Bauch, D. Earn, Vaccination and the theory of games, *Prod. Nat'L Acad. Sci.* 101 (36) (2004) 13391–13394.
- [14] S. Bhattacharyya, C. Bauch, Wait and see vaccinating behavior during a pandemic: A game theoretic analysis, *Vaccine* 29 (2011) 5519–5525.
- [15] C. Bauch, S. Bhattacharyya, Evolutionary game theory and social learning can determine how vaccine scares unfold, *PLOS Comp. Bio.* 8 (4) (2012) e1002452.
- [16] C. Bauch, Imitation dynamics predict vaccinating behavior, *Proc. Roy. Soc. B* 272 (2005) 1669–1675.
- [17] J. Tanimoto, *Sociophysics Approach to Epidemics*, Springer, 2021.
- [18] Z. Wang, C. Bauch, S. Bhattacharyya, D. Zhao, Statistical physics of vaccination, *Phys. Rep.* 664 (2016) 1–113.
- [19] Y. Ma, P. Newton, Role of synergy and antagonism in designing multi drug adaptive chemotherapy schedules, *Phys. Rev. E* 103 (2021) 032408.
- [20] J. West, L. You, J. Zhang, R. Gatenby, J. Borwn, P. Newton, A. Anderson, Towards multi-drug adaptive therapy, *Cancer Res.* 80 (7) (2020) 1578–1589.
- [21] P. Newton, Y. Ma, Nonlinear adaptive control of competitive release and chemotherapeutic resistance, *Phys. Rev. E* 99 (2019) 022404.
- [22] J. West, Y. Ma, P. Newton, Capitalizing on competition: An evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment, *J. Theor. Bio.* 455 (2018) 249–260.
- [23] P. Newton, Y. Ma, Maximizing cooperation in the prisoner's dilemma evolutionary game via optimal control, *Phys. Rev. E* 103 (1) (2021) 012304.
- [24] K. Stuckey, R. Dua, Y. Ma, J. Parker, P. Newton, Optimal dynamic incentive scheduling for Hawk-Dove evolutionary games, *Phys. Rev. E* 105 (014412) (2022).
- [25] J. West, Z. Hasnain, J. Mason, P. Newton, The prisoner's dilemma as a cancer model, *Converg. Sci. Phys. Oncol.* 2 (3) (2016) 035002.
- [26] J. West, P. Newton, Chemotherapeutic dose scheduling based on tumor growth rates provides a case for low-dose metronomic high-entropy therapies, *Cancer Res.* (2017) 10.1158/0008-5472.CAN-17-1120.
- [27] G. Li, S. Shivam, M. Hochberg, Y. Wardi, J. Weitz, Disease-dependent interaction policies to support health and economic outcomes during the COVID-19 epidemic, *IScience* 24 (7) (2021) 102710.
- [28] D. Bertsekas, *Reinforcement Learning and Optimal Control*, Athena Scientific, Belmont MA, 2019.
- [29] C. Winsor, The Gompertz curve as a growth curve, *Proc. Nat'L Acad. Sci.* 18 (1) (1932) 1–8.
- [30] A. Laird, Dynamics of tumor growth: Comparison of growth rates and extrapolation of growth curve to one cell, *Br. J. Cancer* 19 (2) (1965) 278–291.
- [31] P. Gerlee, The model muddle: In search of tumor growth laws, *Cancer Res.* 73 (8) (2013) 2407–2411.
- [32] J. West, Z. Hasnain, P. Macklin, P. Newton, An evolutionary model of tumor cell kinetics and the emergence of molecular heterogeneity and Gompertzian growth, *SIAM Rev.* 58 (4) (2016) 716–736.
- [33] A. Traulsen, J.C. Claussen, C. Hauert, Coevolutionary dynamics: From finite to infinite populations, *Phys. Rev. Lett.* 95 (23) (2005) 238701.

- [34] R. Dua, Y. Ma, P. Newton, Are adaptive chemotherapy schedules robust? A three-strategy stochastic evolutionary game theory model, *Cancers* 13 (2880) (2021).
- [35] J. Park, P. Newton, Stochastic competitive release and adaptive chemotherapy, *BioRxiv* (2022) <http://dx.doi.org/10.1101/2022.06.17.496594>.
- [36] M.A. Nowak, *Evolutionary Dynamics*, Harvard University Press, 2006.
- [37] P. Schuster, Replicator dynamics, *J. Theor. Bio.* 100 (1983) 533–538.
- [38] T. Börgers, R. Sarin, Learning through reinforcement and replicator dynamics, *J. of Econ. Theory* 77 (1997) 1–14.
- [39] M. Piraveenan, S. Sawleshwarkar, M. Walsh, I. Zablotska, S. Bhattacharyya, H. Farooqui, T. Bhatnagar, A. Karan, M. Murhekar, S. Zodpey, K. Rao, P. Pattison, A. Zomaya, M. Perc, Optimal governance and implementation of vaccination programmes to contain the COVID-19 pandemic, *Roy. Soc. Open Sci.* 8 (210429) (2021).
- [40] J. Hsu, A. Meyer, *Modern Control Principles and Applications*, McGraw Hill Book Company, 1968.
- [41] H. Ito, J. Tanimoto, Scaling the phase-planes of social dilemma strengths show game-class changes in the five rules governing the evolution of cooperation, *Royal Soc. Open Sci.* 5 (181085) (2018).