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

K. Stuckey, P.K. Newton

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.

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
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 vacation 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 × 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 [19–22]. 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**: 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.
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.

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

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,

\[
f(t) = a \exp[\exp(b - ct)]
\]

which has a long history of use in actuarial sciences (laws of human mortality), economics (growth laws of wealth), biology
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, $x(t) = (x_A, x_H)^T$; $x_A + x_H = 1$. 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:

$$x_A = x_0 x_A (f_A - f_A),$$
$$x_H = x_0 x_H (f_H - f_H).$$

Here, $f_A$ and $f_H$ denote the fitness of the vaccine adopters and the vaccine hesitant populations, while $f_A$ denotes the average fitness of the entire population under consideration. The system simply says that the growth rate of each sub-population $(x_A/x_A, x_H/x_H)$ 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.
case with vaccine uptake dynamics. The fitnesses are defined via a 2 × 2 payoff matrix A as:

\[ f_A = (Ax)_1 = a_{11}x_A + a_{12}x_H, \]

\[ f_H = (Ax)_2 = a_{21}x_A + a_{22}x_H, \]

\[ (f) = x^T(A)x = x_0f_A + x_0f_H, \]

with:

\[ A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \]

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 \((0, 1)\) 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:

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

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 \( \bar{u}(t) = (u_1(t), u_2(t)) \in \mathbb{R}^2 \) are bounded above and below (based on an incentive size parameter \( p \)):

\[ -\frac{ap}{a_{12} + a_{21}} \leq u_1(t), u_2(t) \leq \frac{ap}{a_{12} + a_{21}} \]  

(17)

\[ 0 \leq p \leq 1 \]  

and a global constraint on the incentive schedule, \( \bar{U}(t) = \int_0^t \bar{u}(\tau)d\tau = \text{const.} \) 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:

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

(18)
Since we are optimizing the endpoint cost, the running cost, while the second is called the endpoint cost. From the target is minimized). We briefly describe the basic context of missile guidance problems where final distance we take \( x \in \mathbb{R} \).

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

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

\[
J(\tilde{x}(\cdot), \tilde{u}(\cdot), t_0, t_f) = \int_{t_0}^{t_f} L(\tilde{x}(t), \tilde{u}(t), t) dt + \Phi(\tilde{x}(t_0), \tilde{u}(t_0), t_1)
\]

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, \( \phi(\tilde{x}(t_0), \tilde{u}(t_0), t_1) = x_t = \chi(t) \) (i.e., the asymptotic value of the 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:

\[
H(\tilde{x}(t), \tilde{U}(t), \tilde{\lambda}(t)) = \tilde{x}^T \tilde{F}(\tilde{x}) + L(\tilde{x}, \tilde{u}(t), t)
\]

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

\[
\dot{\lambda}_i(t) = -\frac{\partial H}{\partial \mu_i^*}
\]

\[
\dot{\mu}_i^*(t) = -\frac{\partial H}{\partial U_i^*}
\]

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

\[
\tilde{x}^*(t_0) = \tilde{x}_0 \quad \tilde{U}^*(t_0) = 0, \quad \tilde{U}^*(t_f) = \tilde{U}_f^*
\]

\[
\lambda_i(t_f) = \frac{\partial \phi(\tilde{x}(t_f))}{\partial x_i^* (t_f)}
\]

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

\[
\tilde{u}^*(t) = \arg \min_{\tilde{u}(t)} H(\tilde{x}^*(t), \tilde{U}^*(t), \tilde{\lambda}^*(t), \tilde{u}(t))
\]

The optimization problem becomes a two-point boundary value problem (using (26)-(28)) with unknowns \( (\lambda_i^2(t_0), x_i^*(t_f)) \) whose solution gives rise to the optimal trajectory \( \tilde{x}^*(t) \) (from (22)) and the corresponding control \( \tilde{u}^*(t) \) 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. 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 \( T \approx 88 \) days with \( \approx 18\% \) of the population vaccinated. The dependent variables in our system are optimized through one fundamental cycle time of length \( T/4 \approx 22 \) days using only an endpoint cost (Meyer problem). Dark blue/red color corresponds to the \( u_i \) schedule, light blue/red color corresponds to the \( u_0 \) schedule in Eq. (16).

**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).
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 \( u(t) = (u_1(t), u_2(t)) \) 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 Gompertian uptake curve, is able to push the vaccinated population up to a value of 38%, as compared to the unvaccinated 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 \( (\dot{x}_1, x_1) \) 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 \( x_1(t_0) \). 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...
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 (∼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
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 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 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 [35] 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 underutilized 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


