Heterogeneous Incentives and Their Impact on Prosocial Behavior: Evidence from COVID-19 Vaccinations

Daniel Gulti Kebede [*] James C. Reeder, III[†]

December 2, 2024

Abstract

This paper examines the impact of heterogeneous public policy incentives on the adoption of prosocial behaviors. To test the effect of incentives, we use COVID-19 vaccination incentives in the United States as our empirical setting. We analyze four types of incentives, categorized across three dimensions: is it monetary or nonmonetary, is the reward direct (immediate and guaranteed) or probabilistic (future and uncertain), and the incentive's value. Adopting the utility-based, Bass Diffusion model (Cosguner and Seetharaman 2022), we estimate the heterogeneous effects of these incentives at the county level. Our findings show that, on average, public policy incentives reduced vaccination adoption, with a 0.5 decrease in utility and a 3% reduction in vaccination rates across treated counties. A more granular analysis reveals that probabilistic incentives, regardless of type, are the primary driver of this negative effect. Additionally, local-level characteristics (e.g., income), moderate the impact of incentives, with higher-income counties exhibiting positive change in utility when incentives are offered. Counterfactual policy analyses of different incentive strategies show that direct interventions—whether cash payments or direct nonmonetary rewards—lead to higher vaccination rates and faster adoption. Our study provides valuable insights for policymakers that employ incentive-based strategies to encourage prosocial behaviors.

Keywords: Prosocial adoption, political polarization, incentives, public policy, machine learning, diffusion models, COVID-19 vaccination.

[*]Department of Economics, Colgate University email: dkebede@colgate.edu
[†]KU School of Business, University of Kansas email: jcreederiii@ku.edu

INTRODUCTION

Research indicates that engaging approximately 25% of a target population in prosocial activities is often sufficient to drive significant prosocial behavioral change within the broader community (Centola 2019).[1] Public policymakers often employ incentives aimed at motivating the target population to reach the desired threshold. This paper examines the effectiveness of heterogeneous incentive programs offered by public sector organizations to promote prosocial behavior. To illustrate how these incentives function, we provide two examples. The Bolsa Família program in Brazil offers direct cash payments of approximately $35 USD to 50 million low-income families, contingent upon children attending school and maintaining regular medical visits (World Bank 2010). With an estimated cost of $8 billion USD, the program has achieved notable success in improving educational and health outcomes. Similarly, Australia implemented a rebate program ranging from $300 to $1,000 USD to encourage the adoption of energy-efficient appliances, with a total investment of about $44 million USD (Queensland Government 2024). This initiative led to increased rates of energy-efficient appliance adoption, contributing to environmental and economic benefits. The success of both programs exemplify the potential of incentive-driven public policies.

While there are notable examples of successful incentive programs, others have disappointing results contingent upon the amount of money spent. For instance, the state of Ohio's VAX-a-Million program, created to encourage COVID-19 vaccination, allocated millions of dollars in incentives but led to no net increase in vaccination rates (Walkey et al. 2021). Similarly, the Cash for Clunkers program, which aimed to incentivize the trade-in of older vehicles for more fuel-efficient ones, cost of $2.85 billion (Gayer and Parker 2013), but only resulted in a modest increase in sales of new cars, primarily among individuals who were already considering purchasing a new vehicle (Mian and Sufi 2012). These mixed outcomes suggest that further research is needed to better understand how policymakers can

[1] Prosocial behavior refers to voluntary actions taken by individuals that benefit both themselves and others, or society at large, for example charitable acts or adherence to health guidelines.

effectively design and implement incentive programs to promote prosocial behavior.

This paper examines the effectiveness of various incentive plans in promoting prosocial behavior. Specifically, we focus on the use of incentives to encourage COVID-19 vaccinations within the United States, where vaccinations can be considered a prosocial activity (Böhm and Betsch 2022). To motivate our study, we develop a conceptual model that builds on the work of Bénabou and Tirole (2003, 2006), which suggest that actions are taken based upon a comparison of intrinsic motivation versus extrinsic cues. Further, Bénabou and Tirole (2006) show that incentives, as an extrinsic factor, may reduce the desire to engage in prosocial behavior. To understand how these extrinsic cues may be moderated, we leverage Self-Determination Theory (Ryan and Deci 2000). By focusing on the concept of autonomy, we examine moderators that may explain the perception of the rewards being considered as controlling, which would cause individuals to be less likely to engage in prosocial behavior. Thus, we identify three contextual moderators related to the incentives themselves, as well as three additional moderators related to individual characteristics. Based upon this model, and our empirical setting, we classify observed incentives into four categories: direct monetary, lottery-based, direct nonmonetary, and scholarship.[2]

To quantify the effects of these incentives on vaccinations, we study the vaccinations of 2,269 counties in the United States, 1,232 of which received an incentive, while 1,037 did not. To estimate the local-level reaction of these heterogeneous incentives, we augment the utility-based, Bass Diffusion model of Cosguner and Seetharaman (2022). This approach allows us to study heterogeneity in incentive type, heterogeneity in local-level characteristics, and the interaction of both. As a result of this estimation, we conduct a series of counterfactual policy exercises with the goal of reconciling the mixed findings in the literature related to incentives and prosocial behavior.

Using our analysis of the heterogeneous responses to each incentive type, we address four key questions that have either been overlooked or yielded inconsistent findings in the liter-

[2] Figure 1 and Web Appendix Section A provide additional information about incentive types.

ature. First, does the incentive type, defined as monetary (cash payments) or nonmonetary (non-cash payments), result in more or fewer vaccinations? Second, does the psychological distance, the gap between the action and when the reward is obtained, defined as direct (immediate) or probabilistic (future and uncertain), affect vaccination rates? Third, does the stated incentive amount influence vaccination rates, conditional on the prior two factors? Fourth, how do county-level characteristics, such as income, prosocial beneficiaries, or political polarization, moderate the effectiveness of incentives? In answering these questions, our study provides actionable guidance for public policymakers on how to target incentives by type, value, and local conditions.

Broadly speaking, our study makes three contributions to the literature. First, we provide novel insight into the effectiveness of incentives in driving prosocial behavior. Unlike prior studies, (Thirumurthy et al. 2022; Walkey et al. 2021; Milkman et al. 2021), which focus on a single type of incentive, we consider all incentives offered during the observation period. This approach allows us to better estimate the differential effects and compare their effectiveness to one another. While Kuznetsova et al. (2022) examines several incentive policies at the country level, we focus on the county level to capture more localized reactions. To our knowledge, we are the only study examining both the heterogeneity in incentives (across three dimensions) and county-level characteristics, providing more in-depth understanding into the mechanisms that moderate the effectiveness of prosocial incentives. By adopting this granular approach, we identify a backfire effect, which has drawn limited attention in other empirical studies.

Second, we contribute to the literature by examining the role of political polarization as a moderator. Public policy initiatives are increasingly being viewed as partisan, even if they are trying to benefit the social good(Pasquini and Saks 2022). Thus, understanding the intersection of political polarization and how it moderates public policy incentives is important. However, prior work has primarily focused on polarization as a function of party affiliation,(Thirumurthy et al. 2022; Liaukonytė et al. 2023; Schoenmueller et al. 2022), while the political science literature suggests other dimensions of polarization exist related to ideological extremism (Schedler 2023; Patkos 2020), which is a collection of beliefs or actions that are far from what is considered moderate within a society. Our study considers both party affiliation and ideological extremism simultaneously and finds that these dimensions have differential effects. We highlight the need that future studies should carefully select the approach dimension of political polarization when investigating its role as a moderator.

Third, our study provides a framework for obtaining the causal estimates of interventions by modifying the utility-based, Bass Diffusion model of Cosguner and Seetharaman (2022). Our approach allows for the local-level estimation of treatment effects, using a counterfactual estimation technique. By focusing on local-level heterogeneous effects, we provide novel estimands for exploring heterogeneity and offer a flexible framework for policy evaluation. Additionally, the Bass Diffusion model allows us to comment on both the speed at which people adopt vaccinations and the overall vaccination rate within a region. These two constructs are critical for understanding how prosocial behavior spreads in incentivized areas.

Our paper is organized as follows. Our second section, *Related Literature and Conceptualization*, provides a comprehensive literature review and develops a conceptual model to test our hypotheses. Our third section, *Data Sources*, describes our empirical setting and the data sets used in our study. Our fourth section, *Estimation*, details our estimation framework that leverages the Bass Diffusion model of Cosguner and Seetharaman (2022). Our fifth section, *Estimation Results*, explores the average and heterogeneous effects estimated by our model. Our sixth section, *Policy Counterfactual*, shows how different incentive policies would impact vaccination adoption. Our seventh and final section, *Discussion*, highlights how our work contributes to both the academic literature and public policymaker practice.

RELATED LITERATURE AND CONCEPTUALIZATION

In this section, we first highlight the contribution of our study to three streams of literature. We then detail our conceptual model, grounded in the theoretical work of Bénabou and Tirole (2003), which serves as a basis for our hypothesis development.

Related Literature

Our work contributes to three streams of literature: the effectiveness of incentives in driving prosocial behavior, the potential backfiring effects of these incentives, and the role of political polarization in moderating behavioral responses. We review the literature in each of these areas and explain how our study advances existing research.

First, we contribute to the literature on incentive design to encourage prosocial behavior. Pioneering empirical work by Hendren and Sprung-Keyser (2020); Hoekstra et al. (2017); Mian and Sufi (2012) show that large-scale, financial incentives often fail to recoup the costs of their investment. For example, Mian and Sufi (2012) find that the Cash for Clunkers program primarily attracted individuals already predisposed to trade in their cars, rather than encouraging other people to participate in the prosocial behavior. Our study shows that many of the policies chosen to encourage COVID-19 vaccinations, a type of prosocial behavior, deterred vaccinations, rather than encouraging it.

Our study builds on the recent research related to COVID-19 vaccination incentives.[3] Prior empirical studies find limited evidence that incentives encouraged COVID-19 vaccinations. Studies, such as Milkman et al. (2021); Thirumurthy et al. (2022); Walkey et al. (2021), find no significant effects of lottery incentives on vaccination rates. Similarly, Sprengholz et al. (2023*b*) find no increase in vaccinations in Germany due to financial incentives. In contrast, Schneider et al. (2023) examines of financial incentives in Sweden and shows small positive effects in vaccinations. Experimental studies show greater effects of financial incentives, contingent upon the incentive's value. Serra-Garcia and Szech (2023) designed and conducted an online experiment and show that when participants were given a low amount of money, they were less likely to state positive intention to get vaccinated. Zhang and Lane (2023), using a hypothetical online study, argue that direct incentives between $8 and

[3] For a comprehensive overview of studies related to vaccinations and incentives, see Schwalbe et al. (2022). We highlight a selection of studies related to ours from there and other sources.

$125 USD reduce vaccine uptake intentions. Both studies show that after a certain threshold, around $ 100, vaccination rates do increase with direct financial incentives. Campos-Mercade et al. (2021b) find that, in Sweden, a direct $24 incentive resulted in higher vaccination rates. Similarly, Sprengholz et al. (2023*a*) find that reactance among the unvaccinated persists until direct financial payments are sufficiently large. These studies highlight that there is still no consensus on the usefulness of incentives in driving vaccinations.

Kuznetsova et al. (2022) is most similar to our study, as this work analyzes several incentive programs. Using an interrupted time series analysis, Kuznetsova et al. (2022) concludes that cash incentives have the highest impact. We extend this work in several important dimensions. Unlike Kuznetsova et al. (2022), we study a broader range of incentive types—direct monetary payments, lotteries, direct nonmonetary payments, and scholarship lotteries—and explore the effect of incentive amounts. Our approach, though focused on observational data, is aligned with the experimental work of Serra-Garcia and Szech (2023); Zhang and Lane (2023); Campos-Mercade et al. (2021b), who considered the incentive's value in studying the effects of direct, financial incentives. Additionally, while Kuznetsova et al. (2022) conducted their analysis at the country level within Europe, our study examines the county-level responses. By incorporating a rich set of control variables, we ensure that our estimates are driven by differences in policy effects, rather than local-level conditions. Further, these county-level variables allow us to better understand the heterogeneous responses to these incentives. Given the model we estimate, we conduct counterfactual policy analysis that provides more granular insights into the local-level impact of the studied incentives.

Our study also contributes to a second body of literature: the exploration of "backfiring" in prosocial contexts, particularly the unintended consequences of nudges. The "backfire effect" refers to a psychological phenomenon where people's beliefs become stronger and more resistant to change when challenged or presented with contradictory evidence (Banerjee et al. 2023). In their experiment designed to promote sustainable food choices, Banerjee et al. (2023) show that nudging people who were already committed to sustainable eating—those with high intrinsic motivation—led to a backfire. In similar contexts, financial incentives may cause individuals to question the overall value of engaging in prosocial behavior.[4]

In the context of vaccinations, poor alignment of financial incentives causes people to question the risks associated with vaccination, potentially reducing vaccine uptake (Cryder et al. 2010; Loewenstein and Cryder 2020). Our study builds on this work by examining the potential for backfiring across three dimensions. First, we assess whether different incentive types are more likely to backfire by quantifying their impact on intrinsic motivation (i.e., latent utility). Second, we investigate whether the incentive amount further moderates this effect, similar to the studies of Cryder et al. (2010); Loewenstein and Cryder (2020). Third, extending the work of Banerjee et al. (2023), we explore how individual characteristics might influence the effectiveness of incentives (i.e., nudges). Specifically, we examine whether local alignment with COVID-19 policies affects how receptive individuals are to vaccination incentives. This final step allows us to offer actionable insights for public policymakers, guiding them on how best to target incentives.

The final stream of literature that we contribute to examines the impact of political polarization on individual decision-making. While there exists a small, but growing, body of research in marketing that explores how political polarization affects consumer behavior (Jung and Mittal 2020; Fernandes et al. 2022; Liaukonytė et al. 2023; Schoenmueller et al. 2022), these studies focus on party affiliation as the key measure of polarization. For instance, Schoenmueller et al. (2022) find that political polarization after the 2016 U.S. election influences consumer preferences, intentions, and purchases. In addition, Bigsby et al. (2017) observed that liberal participants were more supportive of health incentives. Thus, there is importance in understanding how political leanings affect the choices individuals make.

In the context of COVID-19, Thirumurthy et al. (2022) focus on the effects of lottery-based incentives in the US, using state party affiliation as a proxy for political polarization. Our study advances this literature by constructing a more nuanced measure of polarization.

[4]Other studies documenting the backfire effect are Dimock and Wike (2020); Dixit and Weibull (2007).

Instead of relying solely on party affiliation, we examine two dimensions related to polarization: party affiliation and ideological extremism. This delineation builds on the work of Patkos (2020); Jungkunz et al. (2024); Schedler (2023). For example, Patkos (2020) uses government satisfaction data to assess partisan polarization, while Schedler (2023) uses a cluster-based approach to better quantify ideological extremism. By measuring ideological extremism, as proxied by responses to COVID policies, we provide a latent measure of polarization. This latent measure that allows us to assess the impact of ideological extremism on the effectiveness of incentives in encouraging vaccination. This approach allows us to be the first study, to our knowledge, that directly compares and contrasts the effects of party affiliation versus ideological extremism.

Conceptual Model

We begin building our conceptual model with the foundational work of Bénabou and Tirole (2003, 2006), which identifies two key drivers of prosocial behavior: intrinsic motivation, or the internal desire to help others, sometimes at personal costs, and social signaling, or the desire to appear heroic through prosocial acts. In the context of vaccination, intrinsic motivation may stem from both personal protection and contributing to the public good, with prosocial messaging shown to increase vaccination rates (Li et al. 2016). We explore the role of incentives, both on average and heterogeneous incentives, on modifying the utility (intrinsic motivation) of individuals. Specifically, we consider three contextual dimensions of each incentive and three contextual characteristics related to each county.

Bénabou and Tirole (2006) posit that incentives can have dual effects. On one hand, incentives may reduce the inherent cost of engaging in prosocial behavior, making it much more attractive. On the other hand, the external incentive may cheapen the social value of the behavior, which can dissuade participation. This latter behavior is known as "crowding out," where the intrinsic motivation is crowded out by extrinsic factors (i.e., rewards or incentives). This "crowding out" of intrinsic motivation if further supported by Frey and

Jegen (2001). Kamenica (2012) shows that monetary incentives failed to influence prosocial behavior. In our setting, Zhang and Lane (2023) suggest, based upon their experimental study, that incentives reduced the desire to get vaccinated by either signaling lower vaccine quality or crowding out intrinsic motivation. Thus, we hypothesize:

H1: Incentives reduce the utility (intrinsic motivation) of an individual to get vaccinated.

While our first hypothesis examines the general effect of incentives, we expect heterogeneity in the effect based on characteristics of the incentives. We examine three key dimensions: incentive type (monetary vs. nonmonetary), timing (direct vs. probabilistic), and the incentive amount.Monetary incentives involve direct cash payments with no restriction on use, while nonmonetary incentives may either involve no cash transfer or restrict the usage of funds (e.g., for educational purposes or to get a license). Direct incentives are given directly after vaccination occurs. While probabilistic incentives are paid at a future time and are not guaranteed. Last, the incentive amount is the stated monetary value, if known, of the incentive. Figure 1 illustrates these dimensions.

We begin with the first contextual dimension: monetary vs. nonmonetary incentives. According to Self-Determination Theory (Ryan and Deci 2000), monetary incentives may suppress intrinsic motivation by being perceived as controlling, thus, reducing a person's perceived autonomy to act. Murayama et al. (2010) show that monetary incentives can diminish intrinsic motivation in performance-based tasks, as individuals may feel less autonomy in performing the task. Gagné and Deci (2005) further support this finding, noting that individuals who perceive their actions as self-determined are more likely to engage in prosocial behavior. Thus, we hypothesize:

H2: Monetary incentives reduce the utility of an individual to get vaccinated.

Nonmonetary incentives need not be perceived as controlling. These incentives do not seem to reduce the tendency to engage in prosocial behavior. For example, small gifts have been shown to encourage blood donation (Goette and Stutzer 2019). Further, though they are still material, extrinsic rewards, such items are less likely to be considered diminishing the intrinsic value of prosocial activities (Gneezy et al. 2011). Thus, we hypothesize:

H3: Nonmonetary incentives perform better than monetary incentives and may increase the utility of an individual to get vaccinated.

The next dimension is psychological distance. We investigate how direct (immediate) versus probabilistic (future and uncertain) rewards moderate the effect of incentives. Kivetz (2005) suggest that rewards perceived as unrelated to the prosocial act (e.g., lottery payouts) may be viewed as extrinsic, reducing the intrinsic motivation. Further, a public announcement of the winners of a probabilistic incentive may lead others to question the intrinsic value of the winner's actions. This outcome is counter to the social signaling value discussed in Bénabou and Tirole (2006), where the knowledge of receiving an extrinsic reward cheapen the social status of the prosocial act. Thus, we hypothesize:

H4: Regardless of incentive type, probabilistic incentives reduce utility to get vaccinated compared to direct incentives.

The last contextual dimension of the incentive is the incentive amount. Although we explore this effect across all four incentive types, direct, monetary payouts have garnered the most attention in the experimental literature (Serra-Garcia and Szech 2023; Sprengholz et al. 2023*b*; Zhang and Lane 2023). A critical aspect of direct monetary payments as an external stimulus is its perceived value. Higher amounts of direct payments can be perceived as a justified reward for behavior rather than a controlling mechanism. Thus, we hypothesize:

H5: Increasing the amount of the direct, monetary incentive will increase the utility of an individual to get vaccinated.

An important consideration in understanding the effectiveness of incentives is the characteristics of the individuals receiving them. These characteristics likely moderate the perception of the incentive and, in turn, affect the intrinsic motivation (i.e., utility). We proxy for individual-level characteristics through the use of county-level characteristics. This approach allows us to study the heterogeneous response to incentives across three dimensions: income, prosocial beneficiaries, and political polarization. For each moderator, we examine its relationship with intrinsic motivation or, potentially, validation through external cues.

We begin with income, for which predictions in the theoretical literature are mixed. On the one hand, research indicates that high-status individuals (i.e., those with higher income) are less likely to engage in prosocial behavior or may even engage in more unethical behavior compared to lower-status individuals (Visser and Roelofs 2011; Piff et al. 2012). Frey and Jegen (2001) propose that crowding in behavior may occur when a reward is perceived as supportive, granting individuals greater autonomy and enhancing self-determination (i.e., intrinsic motivation). For lower-income individuals, financial constraints may lead them to perceive smaller incentives as offering greater autonomy. Fishman et al. (2023) support this notion, showing in an online randomized controlled experiment that financial incentives, of $200 or $1,000 increased the willingness to get vaccinated and posit that the effect is larger for individuals of lower income compared to higher income. However, the hypothetical incentives used in this study were much higher than those in our empirical setting.

On the other hand, some studies suggest that higher-income individuals may have stronger intrinsic motivations for prosocial behavior and are less likely to be affected negatively by extrinsic rewards. For instance, Suss (2023) shows that higher-income individuals are more likely to engage in prosocial behaviors when confronted with greater inequality. Macchia and Whillans (2021) find that higher-income individuals are more likely to donate to charities and volunteer, indicating that they are motivated more by intrinsic factors. Further, Judge and Hurst (2007) suggest that individuals with positive core self-evaluations—related to self-esteem and internal locus of control—tend to have higher incomes throughout their careers. These findings suggest that higher-income individuals are both more likely to engage in prosocial activities and be less influenced by the stigma around extrinsic incentives. Thus, we have two competing hypotheses:

H6a: The utility associated with getting vaccinated increases for lower-income individuals when incentives are offered.

H6b: The utility associated with getting vaccinated increases for higher-income individuals when incentives are offered.

Our second moderator, prosocial beneficiaries, is related to the vulnerability of the community to COVID-19. The primary goal of the incentive policies was to support communities that were most at risk of experiencing negative outcomes from the pandemic. Thus, it is critical to examine how these at-risk communities respond to different public policy incentives. For this moderator, we present two competing hypotheses.

On one hand, existing literature suggests that vulnerable communities may be more likely to engage in prosocial behavior. Byrne et al. (2023) conduct a meta-analysis showing that prosocial initiatives lead to improved health outcomes in vulnerable communities. Similarly, Xiao et al. (2016) demonstrate that individuals are more likely to engage in prosocial behaviors when they are confronted with the benefits to others, particularly those who are vulnerable. If incentives reduce barriers to vaccination, individuals may still choose to vaccinate because their intrinsic motivation to help others outweighs any perceived reduction in autonomy caused by the extrinsic incentive.

On the other hand, the vulnerability of a community is predicated on their skepticism towards the intervention—in this case, COVID-19 policies. Previous research has shown that vaccine skepticism, fueled by misinformation, leads to lower vaccination rates (Duquette 2020; Sallam 2021; Argote et al. 2021) and can potentially reduce the effectiveness of incentives(Serra-Garcia and Szech 2023). Chang et al. (2021) highlights that financial incentives are ineffective in motivating vaccine-hesitant individuals through an experimental study. For such groups, misaligned incentives may exacerbate existing reluctance, creating a backfire effect. Drawing from Self-Determination Theory, extrinsic incentives may be perceived as controlling by vaccine-hesitant individuals, thereby diminishing their sense of autonomy. This reduction in perceived autonomy could lead to reactance, where individuals become even more resistant to vaccination. Thus, we present the following competing hypotheses:

H7a: The utility associated with getting vaccinated increases for individuals with a high vulnerability to COVID when incentives are offered.

H7b: The utility associated with getting vaccinated decreases for individuals with a high vulnerability to COVID when incentives are offered.

Given that public policy incentives are inherently political interventions, understanding the moderating effect of an individual's political orientation is crucial. We consider two dimensions of political polarization: party affiliation and ideological extremism. Based on party affiliation, we hypothesize that individuals aligned with the party opposed to vaccination would become more entrenched in their views when incentives are offered. As shown in Bowles and Polania-Reyes (2012), individuals who are control averse would feel resistance if the incentive is perceived as reducing their autonomy. We argue that individuals associated with the party opposing COVID policies would likely exhibit this control aversion to maintain their autonomy, thereby increasing reactance. We make the same argument for individuals who are ideologically opposed to vaccination policies. Thus, we hypothesize:

H8: The utility associated with getting vaccinated decreases for individuals who oppose COVID policies when incentives are offered.

DATA SOURCES

In this section, we first introduce the setting. Then, we provide detail on the incentives, chosen moderators, and our other control variables. The full set of descriptive statistics for all variables in our study can be found in Table 1.

Empirical Setting

We study the COVID-19 vaccine diffusion in the United States from December 13, 2020, to December 1, 2021. The dependent variable for our analysis is the weekly total of individuals receiving at least 1 dose of a COVID-19 vaccine (Pfizer, Moderna, or Johnson and Johnson) at the county level provided by the CDC.[5] While the CDC provided measure is at the daily level, we aggregate it to the bi-weekly level. Though there were supply chain issues

[5]https://covid.cdc.gov/covid-data-tracker/datatracker-home

during the initial period of the COVID-18 vaccination roll-out in December 2020 and January 2021, our comprehensive set of control variables, shown later in this section, accounts for these latent supply chain issues in modeling the diffusion of vaccinations.

For our main analysis, we examine 2,269 counties—1,232 received an incentive (treatment), and 1,037 did not (control). The sample is smaller than the 3,112 counties in the U.S. for two reasons. First, approximately 10.5% counties lack Google Mobility data, which we used as a proxy for social mobility, necessitating their exclusion. Figure 3 shows in gray the counties that are eliminated from our sample due to this criteria. Mainly, we lose very rural counties with this step. Second, we remove counties from six states, resulting in a 16.3% reduction in the sample—Colorado, Hawaii, Ohio, Texas, Virginia, and Washington—due to insufficient reporting or large statewide re-statement of vaccination numbers during the study period. We conduct robustness exercises later on in this study to see the impact of the removal of this state and similar results of our main effect in terms of magnitude, directionality, and statistical significance. We make one final cut of counties that either did not report any vaccinations in the first 22 periods or reported negative vaccination numbers in that period, leaving us with a final sample of 2,269 counties.

Incentive Contextual Moderators

To encourage vaccination, governors across the U.S. allocated funds for vaccine incentives. We treat this emergency decree by state governments as a quasi-natural experiment, comparing vaccination adoption between counties receiving treatment (i.e., public policy incentive) and those not receiving an incentive. Using the National Governors Association memo,[6] which compiles incentive data across the U.S., we observe a total of 67 instances of periods where incentives occur. To align with our conceptualization, we categorize these incentives into four groups: Monetary (direct monetary), Lottery-based (probabilistic monetary), Nonmonetary (direct nonmonetary), and Scholarship (probabilistic nonmonetary).

[6] https://www.nga.org/publications/covid-19-vaccine-incentives/

Figure 1 provides details of these incentives. For further discussion, see Section A in our Web Appendix.

In quasi-natural experiments, there is always a concern that treatment and control groups differ. Table IA.5 in our Web Appendix Section E shows the covariate balance between treatment and control counties. As indicated, there is a lack of balance, particularly regarding reactions to COVID-19 policies. Without adjusting for this discrepancy, we risk biased inferences about the effect of treatment. We address this issue in our modeling section using both direct modeling techniques and a matching exercise to show robustness of our findings.

Individual Contextual Moderators

In addition to studying the effects of public policy incentives, we also explore how county-level heterogeneity influences both the underlying diffusion of vaccine adoption and moderates the effects of these treatments. Based on our conceptual model, we examine three key moderators: income, prosocial beneficiaries, and political factors.

The first moderator, income (*Income*), assesses how an individual's income influences their willingness to adopt prosocial behavior and their response to incentives. Income is the median county-level income data from the U.S. Census scaled by $10,000.

The second moderator, prosocial beneficiaries (*Risk*), accounts for individuals who would most benefit from the prosocial incentives. We use the Center for Disease Control's (CDC) CVAC measure, which quantifies the difficulty of a vaccination roll-out within a community, accounting for both resistance to vaccination and other structural factors in the region. A CVAC score of 0 indicates an easy roll-out, while a score of 1 indicates the most difficult roll-out. Our goal is to investigate whether incentives influence vaccination adoption in communities with a high CVAC score.

The third and final set of moderators focuses on the political landscape of the region. We account for politics along two dimensions: political affiliation (*Vote*) and ideological extremism (*Polar*). The *Vote* variable approximates the political affiliation of the individuals within a county. Consistent with existing literature (Thirumurthy et al. 2022), we measure political affiliation, *Vote*, as the difference between Democrat (Joe Biden) and Republican (Donald Trump) candidates in the 2020 U.S. presidential election.

In addition, we construct *Polar* measure to capture ideological extremism relate to COVID-19 policies. While a region's political affiliation may be influenced by socioeconomic factors or historical legacies, it may not fully reflect local attitudes toward specific policies. To address this issue, we create a measure of polarization on COVID-19-related policies; including vaccination, masking, and social distancing. Figure 3 shows the distribution of this constructed measure of ideological extremism across the United States. The constructed variable related to polarity takes the values of -5.078, which represents being against COVID-19 policies, to 4.942, which represents being for COVID-19 policies.[7]

Control Variables

In addition to our selected individual contextual moderators, other factors related to both intrinsic motivation (demand) and supply of vaccinations may influence vaccine adoption rates with a region. As noted, the beginning of our empirical setting encompasses the time when there were supply chain constraints on vaccinations. As such, we need to include control variables to account for these issues. In this section, we categorize these control variables into those affecting intrinsic motivation to get vaccinations and the supply of vaccinations.

Regarding factors related to intrinsic motivation, we already account for factors in our individual contextual moderator selection—income, prosocial beneficiaries, and political polarization. However, there may be other factors influencing the intrinsic motivation to get vaccinated that are not fully captured my these moderators. We collect four additional variables that may influence the intrinsic motivation of a region to get vaccinated. First, we collect census data on the percentage of minorities and females within a region, as these demographic characteristics may influence intrinsic motivation. Second, we collect data on

[7] Please see Section B in our Web Appendix for a description of the data and construction of this variable.

the county-level unemployment rate, as individual who are unemployed may have different incentives or barriers to vaccination. Third, we collect the Social Vulnerability Index (SVI) from the CDC,[8] which summarizes a community's vulnerability to disaster. The SVI integrates economic, demographic, and health-related factors such as education, housing, and access to health care. Ranging from 0 (least vulnerable) to 1 (most vulnerable), the SVI describes regions where both intrinsic motivation and supply chain issues may hinder vaccine adoption. Fourth, we collect the overall population size of the county, which may influence both the diffusion of vaccination and the total number of individuals choosing to vaccinate.

Our prior intrinsic motivation controls are static within a region. We include one final variable that is time varying as a control: the number of COVID-19-related deaths per period within each county. This data, obtained from John's Hopkins COVID-19 website that we aggregate to the weekly level, reflects the severity of COVID-19 in a region over time. As local news often covered COVID-19 deaths, this variable may influence the intrinsic motivation to get vaccinated. To standardize this variable across counties, we scale it by the population size and multiply it by 10,000. Additionally, any period with negative values is adjusted to 0.

On the supply side, the availability of vaccinations is a key concern. Our individual contextual moderator income is likely correlated with vaccine availability——wealthier regions typically have more infrastructure and capacity. We further approximate availability of vaccinations by incorporating a measure of the number of medical facilities within each county. Given that hospitals and outpatient clinics were major vaccination distribution points and medical personnel were prioritized in obtaining vaccines in the initial period of vaccinations, the number of medical facilities serves as a proxy for vaccine supply and targeting of distribution. Additionally, we use data from the Census to classify regions by their metropolitan status (M_{Code}), ranging from 1 (rural) to 6 (metro area). This variable further serves as a proxy for the availability of infrastructure to deliver vaccinations.

[8]https://data.cdc.gov/stories/s/Vaccine-Hesitancy-for-COVID-19/cnd2-a6zw/

ESTIMATION

In this section, we start by providing evidence that vaccine adoption follows a Bass Diffusion framework and then describe the utility-based, Bass Diffusion model of Cosguner and Seetharaman (2022) to estimate the underlying diffusion process. Then, we describe the framework that enables us to estimate the county-level treatment effects that are necessary to achieve our goal of examining heterogeneous effects of this public policy intervention.

Adoption of Prosocial Behaviors in a Bass Framework

We argue that the Bass Diffusion framework is well-suited to studying prosocial adoption. The Bass Diffusion model is typically used to describe the adoption of durable products. To map out the adoption curve, the model considers two groups of individuals. The first group, innovators, are primarily driven to adopt the usage of the product through either internal desire or external cues, like advertising. The second group, imitators, are thought to be motivated to adopt based upon social cues and communication with others who have adopted the product previously. The combination of these two groups then describes the trajectory of potential to actual adopters of a product over time.

We use the intuition of the Bass Diffusion model to explain the adoption of prosocial behavior, specifically the choice to get vaccinated. We start by categorizing the people that may choose to get vaccinated into two groups. The first group are intrinsically motivated and need no external cues to engage in prosocial behavior (Bénabou and Tirole 2003), which are akin to the innovators in the Bass framework. The second group are motivated through extrinsic factors and social cues (i.e., seeing their peers adopt prosocial behaviors), which are akin to the imitators in the Bass framework. Bollinger et al. (2020) show in their field experiment with solar panels, that the use of prosocial, community-oriented messaging leads to greater peer recommendations, which mimics the information flow from innovators to imitators in the Bass framework. Last, most prosocial behaviors, including vaccination, result in an absorbing state, meaning individuals who engage in the behavior are likely to continue. Our work directly builds on Kim and Rao (2021), who used the Bass Diffusion model to estimate the vaccination rates of individuals in Michigan.[9]

To provide further justification of using a Bass Diffusion model as the underpinnings of our structural model, we provide further evidence that vaccine adoption in our setting. Figure 4 illustrates the vaccination adoption amounts over time for counties in our sample. The top panel compares the average vaccination rates between counties that received an incentive (treatment) and those that did not (control). Both treatment and control counties show an initial slow start, then an acceleration period, followed by a peak, and then a decline. As shown, the levels between the diffusion curves are notably different. Below the aggregate figure, we display vaccination trends for 12 example counties (6 treatment and 6 control), showing variations in the timing and magnitude of the peaks. To ensure that we estimate unbiased effects of incentives, our estimation framework must directly account for the observed heterogeneity in diffusion patterns.

Model Framework

Having shown that the pattern of vaccine adoption fits the general pattern associated with Bass Diffusion, we now turn to our estimation strategy. The standard Bass Diffusion model does not easily account for contemporaneous effects of marketing interventions. The utility-based, Bass Diffusion model of Cosguner and Seetharaman (2022), specifically the Bass-Logit Diffusion Model, provides us with the needed structure to estimate the direct effect of contemporaneous shocks. However, the model proposed by Cosguner and Seetharaman (2022) assumes homogeneity in the Bass Diffusion parameters, as they study only a single market, and focus on time-varying shocks to the utility. Our setting, has heterogeneity in the diffusion process and time-invariant shocks in the form of treatment. Thus, we propose a modified framework on the work of Cosguner and Seetharaman (2022) to account for these

[9]Other studies have found similar diffusion patterns to vaccinations. For example, most provinces of Canada, Germany, Italy, and France follow the same inverted U-shaped pattern in vaccination rates (Karaivanov et al. 2022).

factors and allow for the estimation of heterogeneous effects of treatment.

Our target estimate of interest is $\tau_{P,m}$, which is the effect of a given public policy incentive P in market m. We can then build the utility for an individual in county m at time t to get vaccinated:

$$U_{m,t} = \alpha_{m,t} + \boldsymbol{X_{m,t}\beta_m} + \boldsymbol{P_{m,t}\tau_{P,m}} + \epsilon_{m,t}. \tag{1}$$

$\alpha_{m,t}$ is the market-specific intercept related to people within county m at time t. $X_{m,t}$ represents time varying shocks within the county that may cause an individual to choose vaccine adoption. In our context, this $X_{m,t}$ will be the number of COVID related deaths within the county during the observed period and β_m is the county-level response to those COVID deaths. $P_{m,t}$ are dummy variables that take the value of 1 if a specific incentive bundle is occurs in county m during week t and 0 otherwise. $\epsilon_{m,t}$ are the IID error terms that follow a logistic distribution with location parameter 0 and scale parameter 1.[10]

As noted, the intercept of this model varies over time and across markets. Specifically, the intercept is parameterized as follows:

$$\alpha_{m,t} = ln\left[ln[\frac{1-F_{m,t-1}}{1-F_{m,t}}]\right], \tag{2}$$

where $F_{m,t}$ is the cumulative distribution function that represents the total number of individuals who have been fully vaccinated at time t in market m. If we assume $F(0) = 0$, then we can solve $F_{m,t}$ directly with:

$$F_{m,t} = \frac{1-e^{-(p_m+q_m)t}}{1+\frac{q_m}{p_m}e^{-(p_m+q_m)t}} \tag{3}$$

where p_m represents the parameter of innovation from a Bass Diffusion process in market m. Similarly, q_m is the parameter of imitation for a Bass Diffusion process in market m.

[10]Since we estimate the model at the market level, the effect of time-invariant, control variables at the market level would not be identified. Thus, they are not included in the utility specification.

Thus, due to market-specific diffusion parameters, representing the latent adoption pattern of vaccinations within the market, we capture market-level heterogeneity across markets and time.

With the above equations, we can now construct the unconditional likelihood of the number of individuals in market m getting vaccinated at time t as:

$$L_{m,t} = \left[\prod_{s=1}^{t-1} 1 - pr_{m,s} \right] pr_{m,t} * M_m, \tag{4}$$

where $pr_{m,s}$ is defined as:

$$pr_{m,s} = \frac{e^{U_{m,s}}}{1 + e^{U_{m,t}}}. \tag{5}$$

Intuitively, the overall likelihood function is the product of the percentage of people not getting vaccinated in each prior period (i.e., $t-1$, $t-2$, etc.) multiplied by the percentage of people choosing to get vaccinated in the current period t. This variable is then scaled by the market size M to match the number of people getting vaccinated at time t. Estimation of the parameters can then be achieved through non-linear least squared optimization.

The model of Cosguner and Seetharaman (2022) is predicated on studying a singular market with time-varying covariates to estimate the effect on the observed diffusion process. Thus, directly transporting the aforementioned model to our setting is prohibited by two challenges. First, generating a pooled model across all counties may yield inconsistent estimates or an intractable model for the effects of treatment variables. Assuming one average set of diffusion parameters would likely over/under estimate the diffusion pattern at the county level. The inclusion of fixed effects at the county level for the utility parameterization would likely result in a biased outcome as well. Covariates included in the utility function would fail to properly capture heterogeneity in the local-level diffusion process.[11] Further, if the span of time an incentive is offered is sufficiently long in our observation window, it would be

[11]See Section C in our Web Appendix where, through simulation, we show that the inclusion of market-level covariates is unable to properly account for market-level heterogeneity in the underlying Bass Model.

challenging to separate the effect of the incentive from the county-level fixed effect. Second, extracting the heterogeneity in the policies using the direct utility specification would be challenging. The simultaneous inclusion of control variables that also serve as moderators on the treatment variables may lead to biased effects if the control counties are markedly different from the treatment counties. Thus, to properly estimate the heterogeneous effects of the incentives, we need a framework that allows us to properly estimate a counterfactual scenario where counties that received an incentive did not receive such an incentive. Then the difference between the observed diffusion pattern and the counterfactual one would then allow for the estimation of the heterogeneous incentives, the desired estimate of our study.

Causal Estimation of Treatment Effects

To estimate the heterogeneous effects of heterogeneous incentives under the utility-based, Bass Diffusion model, we propose a framework that leverage counterfactual estimation techniques. These techniques allow us to estimate the causal, local-level effects of treatment by drawing on information from the control group's observations.[12] To construct such counterfactuals, we would start with the estimation of effects market by market. This approach would allow us to directly account for the underlying diffusion pattern within each market and isolate the effects of the public policy incentive. However, we face a challenge: we need to estimate at least 5 parameters at the county level, but only have 25 observations per county. Fortunately, Cosguner and Seetharaman (2022) show the performance in estimating p, q, m, and β for a small number of observations. However, their study does not account for covariates with time-invariant effects, nor does it explore how to obtain the causal effect of said covariates. Thus, we need to leverage existing work in the Bass Diffusion literature to find a path forward.

The major challenge we must overcome is a way to systematically create a counterfactual outcome for each county that received treatment. The notion of creating this counterfactual

[12]Please see Liu et al. (2024) for a discussion of counterfactual estimators in linear regression models, which serve as a basis for our counterfactual framework.

estimate of diffusion is no different than a marketer trying to leverage information to forecast the diffusion pattern for an unknown product. The literature suggests using analogs of the focal product to estimate its diffusion pattern (Cosguner and Seetharaman 2022). If more than one analog is suitable, the researcher takes a weighted average of the analogs to estimate the parameters of interest p, q, and m to construct the Bass model of the new product. We adopt this intuition to estimate the counterfactual diffusion pattern for a given county as if it had not received the incentive (treatment). In our case, we treat counties in the control group as analogs for those in the treatment group. These control counties provide an opportunity to estimate the structural parameters of the Bass model, as they are unaffected by the incentive. The key challenge is determining the appropriate weights for each control county when imputing the unobserved Bass parameters for each treatment county. This step is crucial, as pooling over counties yields inaccurate estimates. By creating the counterfactual diffusion pattern using similar control counties, we obtain a more accurate estimate of the effect of the public policy incentive.

To systematically estimate the counterfactual Bass parameters, we need an accurate mapping from control county diffusion parameters to observed county characteristics. For this, we turn to machine learning techniques that enable non-parametric imputation. A benefit of machine learning methods are their ability to find non-linear relationships between variables and, thus, create more accurate predictions. We use an optimally tuned random forest with 10,000 trees as the method to link our observed covariates to estimated Bass parameters at the county level. Random forest is known for its strong performance in prediction/imputation tasks. Further, it is known as a method that prevents overfitting due to its underlying structure, which is necessary to increase the accuracy of our imputed Bass parameters.[13]

More formally, our estimation approach proceeds in the following steps:

Step 1. Estimation of $\hat{\theta}$. For each control (analog) county, we estimate the model

[13]For our study, each random forest is optimally tuned via cross-validation using the grf package.

primitives in Equation 1. Note, since these are control group counties, there is no shock from an incentive occurring; thus, this estimation provides estimates of $\hat{\theta} = \{\hat{p_m}, \hat{q_m}, \hat{M}_m, \hat{\beta_m}\}$, which are the county-level estimates of the underlying Bass Diffusion model.

Step 2. Estimation of $g_\theta(F_{m,c})$. Armed with $\hat{\theta}$, we now need to generate functions that link observed county characteristics of the control (c) counties, $F_{m,c}$, to the estimated Bass parameters. We call this set of four functions, one for each estimated Bass parameter, $g_\theta(F_{m),c}$. We view this exercise as one requiring prediction/interpolation of effects;thus, we leverage machine learning to generate the structural linkage between F_m and $\hat{\theta}$. Each function is generated via an optimally tuned random forest model.

Step 3. Estimation of $\dot{\theta}$. To complete the estimation of our counterfactual diffusion process, we need estimates of the Bass parameters for the treatment group counties, $\dot{\theta}$. For each treatment (T) group county, we use their county characteristics and then obtain the predicted Bass parameters from $g_\theta(F_{m,T})$.

Step 4. Estimation of $\hat{\tau}_{P,m}$. Using Equation 1 with $\dot{\theta}$ at the county level, the only free parameters are then $\tau_{P,m}$ and M_m. We then estimate $\hat{\tau}_{P,m}$ for each incentive P in market m. The estimate of $\hat{\tau}_{P,m}$ is then the difference between the observed outcome in period t for market m and the assumed counterfactual outcome for the same period described by $\dot{\theta}$ for market m based upon the counterfactual diffusion parameters. For example, if the observed vaccination rates are lower than the predicted rates, then the incentive is suppressing the diffusion process, resulting in a negative value for $\hat{\tau}_{P,m}$. A secondary benefit of this approach is we can compare $\dot{M}_m$ versus $\hat{M}_m$ for each market in treatment to estimate the effect of public policy incentives on the total vaccination amounts.

Focusing on $\hat{\tau}_{P,m}$ in our subsequent analysis offers two main advantages. First, by imputing the underlying Bass parameters at the county level based upon observable features, we effectively difference out the influence of all control variables. Consequently, $\hat{\tau}_{P,m}$ reflects only the effect directly attributable to public policy incentives and their interaction with county-level characteristics. Second, using $\hat{\tau}_{P,m}$ as our construct of interest allows for various forms of aggregation, enabling us to explore heterogeneity in effects based on treatment type/value, county-level characteristics, and their interaction.[14]

We rely on two key assumptions to estimate the effects of incentives. The first assumption is that our set of county-level characteristics adequately captures the variation in Bass parameters, and that there are sufficient similar analogs to construct the imputed Bass parameters for treated counties. We argue that our collection of variables is sufficient in describing variation of the Bass Diffusion process at the county level. To support this assumption, Web Appendix Section E shows that there are sufficient counties in our control group to create a matched sample of our treatment counties. To flexibly impute the Bass parameters, we use machine learning to capture non-linearities and interactions that would otherwise be difficult to capture.

The second assumption is the absence of anticipation effects. If individuals alter their behavior anticipating the incentive (e.g., delaying vaccination until the incentive takes effect), then our counterfactual diffusion pattern would not align with the observed diffusion pattern before treatment. We argue that, since the incentives were implemented by state governors via emergency powers, the timing left little opportunity for individuals to adjust their behavior strategically. In Section D of the Web Appendix, we formally test this assumption by comparing the vaccination rates in the period before the incentive was offered versus prior periods for counties that received treatment. From this test, we find no significant difference in behavior; thus, our assumption of no anticipation holds.

ESTIMATION RESULTS

In this section, we present the results from our estimation of heterogeneous incentives on vaccine adoption timing and overall vaccination uptake.[15] We start by presenting the

[14]See our Web Appendix Section C where we conduct a series of simulation exercises to show that our approach recovers unbiased estimates of treatment, while other candidate approaches do not. Further, we show through a second simulation exercise that joint estimation of time-invariant treatment along with the Bass Diffusion parameters yields biased results.

[15]The focus of our paper in understanding the implications of public policy incentives on prosocial adoption. However, we do include in Web Appendix Section D a discussion of our results from the underlying Bass parameters

aggregate effect on utility of the public policy incentives. Along with these estimates, we provide a series of robustness checks on the aggregate estimated effect. Next, we explore the heterogeneity of effects in three steps: aggregate effects of incentives on the estimated treatment effect parameters, aggregate effects of county-level variables on estimated treatment effect parameters, and then the effect of county-level variables on the heterogeneous treatments. We end the section by exploring the combination of heterogeneous treatments and county-level variables on changes in aggregate vaccination adoption.

Average Effect of Incentives

We start with the pooled estimate of the effect of each public policy incentive and county pairing within our sample, $\hat{\tau}_{P,m}$, which will also serve as the dependent variable for our analysis. Given that there may be counties that do not have enough support in the analog counties based upon observed covariates, we trim our treatment effect estimate at the 1% level. We estimate the pooled effect of incentives to be $-.518$ with a standard error of .021 (p-value of this estimate compared to 0 is less than .001) across 2,195 treatment observations, as each county may receive more than one incentive during our observation window. Thus, on average, incentives are shown to slow the diffusion of vaccination adoption.

To see if our estimates are robust, we proceed with a series of exercises that either adjust the sample or the utility function. To start, if we estimate the effect of incentives using our untrimmed sample, the mean of our estimated effects is $-.556$ with a standard error of .031 (p-value $< .001$), which is statistically indifferent to our prior estimate. Next, we estimate the effect of incentives by adding back counties that were removed to revisions in the vaccination records. Using this new sample, our estimate is $-.604$ with a standard error of .038 (p-value $< .001$), showing a consistent effect. Finally, we test to see if our contemporaneous shock, COVID-19 Deaths, provides meaningful variation to our model. We estimate our model without the inclusion of COVID-19 deaths and find the effect to be $-.795$ with a standard

from the analog counties.

error of .043 (p-value $< .001$), showing a significantly more pronounced negative effect. Thus, the inclusion of deaths within the county serves as an important control; yet, we still find consistency of the estimated effect.

One last robustness exercise tests if the use of machine learning to generate the counterfactual Bass parameters is providing consistency in estimates. Inherently, the machine learning method is finding partitions of the data to group like counties together to estimate the Bass parameters. An alternative method would be to direct match treatment counties with control counties and use the matched county's Bass parameters for the treatment county. Using our set of control variables, we find a one-to-one match analog using the Mahalanobis distance measure with replacement and nearest neighbor matching.[16] The estimated effect of treatment, after trimming at the 1% level is $-.428$ with a standard error of .028 (p-value $< .001$). Once again, we show consistent negative effects of incentives, regardless of sample, and imputation, thus bolstering confidence in our results and providing support for our **H1**.

Heterogeneous Effects of Heterogeneous Treatments on Vaccination Diffusion

To examine the role of both public policy incentive and county level moderators on the overall utility of vaccine adoption, we proceed in three stages. First, we regress the individual-level effects on covariates related to the incentive type and value. Second, we perform the same analysis, but examine the county-level characteristics as regressors to understand how they moderate the effectiveness of the incentives. Third, we discretize the treatments into their groups and look at the interaction between county-level characteristics and public policy incentives. These exercises allow us to formally test our hypotheses.

Table 2 presents results from a regression of the estimated treatment effects on characteristics of the incentives. Recall that we group incentives into four categories: Monetary, Lottery, Nonmonetary, and Scholarship. In our estimation, we assume that Monetary is the base case and compare the latent utility of each other type to that base case. We include fur-

[16]Web Appendix Section E provides covariate balance measures and a replication of our homogeneous results presented in the next subsection. We show consistency in direction and magnitude across most covariates.

ther variables in our regression related to the incentives: the number of weeks the incentive is offered (Duration), the incentive's value (Cash, which is logged to account for skewness in the values), and the number of concurrent promotions occurring at the time (Num Promo).

The results from this regression provide the following findings. The incentive that offers the highest latent utility, holding the value of the incentive fixed, is Monetary payouts. The dummy variable for each other incentive type is negative and statistically significant. Congruent with other studies, we find that incentives with a higher value are more likely to accelerate vaccine adoption ($\beta = .054$, p-value $< .001$). This finding suggests that direct incentives, particularly direct monetary payouts of sufficient size, accelerates vaccination adoption, showing support for **H5**. Similarly, given the large negative effect on probabilistic incentives, **H4** holds as well. We also note that offering too many concurrent incentives may suppress vaccination adoption ($\beta = -.13$, p-value $< .1$).

Examining differences between incentives only shows some interesting aspects of heterogeneity. Table 3 presents the results of regressing the estimated incentive treatment variables on our moderators of interest and other control variables. To check the stability of our effects, we recursively add covariates from left to right, as shown in the columns of the table.[17] In all specifications, we use the natural log of income. In the first panel, we include Polarity directly. In subsequent models, we move towards the Polar (For) and Polar (Against) specification, bottom 25% and top 25% of the Polarity measure, to see if the effect of polarity is driven by pooling of effects from more extreme counties.

We start by examining the effect of our vulnerability and income moderators on whether incentives accelerate or suppress vaccination diffusion. Those regions that are most vulnerable accelerate vaccinations when incentives are offered ($\beta = .249$, p-value $< .05$). Similarly, higher income regions accelerate vaccination rates when incentives are offered ($\beta = .405$, p-value $< .01$). These findings support our arguments for hypotheses **H6B** and **H7B**.

[17] We do not include population, the number of medical facilities, and the type of the region in these models as, we argue, they will affect the diffusion patterns and aggregate number of vaccinations, not market-level reactions to incentives.

Polarity is shown to be consistently statistically significant across all specifications. We focus on the results in column four. The results indicate that counties who are against COVID-19 policies respond negatively to incentives by slowing the diffusion of vaccinations within those counties ($\beta = -.183$, p-value $< .01$). For those counties that are for COVID-19 policies, incentives accelerate diffusion ($\beta = .130$, p-value $< .05$). Results for the other control variables show important drivers of treatment effect heterogeneity. Vote, a measure of political affinity, is in the expected direction as Republicans are generally considered against all COVID related protocols ($\beta = -.414$, p-value $< .001$). This finding provides support for hypothesis **H8**. We find a similar effect with those regions having more minorities do not get vaccinated as quickly when incentives are offered ($\beta = -.383$, p-value $< .01$). Higher unemployment rates show an acceleration in vaccination adoption when incentives are offered ($\beta = .057$, p-value $< .001$). This last result adds nuance to our investigation of the moderating effect of income. Regions with more unemployed may see benefit in the incentives aligned with vaccination, and thus choose to get vaccinated to obtain the offered incentive.

Thus far, our examination has been piecemeal to first see the overall effects of the heterogeneous incentives and then heterogeneous reactions to incentives governed by county-level characteristics. We now turn to examining the interactive effects between incentive type and county-level characteristics to see if there is richer heterogeneity at play. For this estimation, we isolate treatment observations that only include a single one of our incentive types. For each incentive type, we regress the estimated treatment effects on our covariates of interest. The results of this exercise are shown in Table 4. Due to a lack of observations for scholarship incentives, we can only focus on these three: Monetary, Lottery, and Nonmonetary.

The results in Table 4 show the importance of examining the heterogeneous responses to these different treatments. Our prior homogeneous results suggest that entrenched anti-COVID policy beliefs result in negative effects of incentive. These heterogeneous results show the opposite, as the Polar (Against) show positive and statistically significant gains on both the Monetary and Lottery incentives ($\beta = .999$, p-value $< .001$) and ($\beta = -.187$, p-value $< .1$), respectively. Conversely, this group reacts negatively to Nonmonetary incentives, ($\beta = -.412$, p-value $< .001$). Those regions with more vulnerability to COVID (Risk) show negative effects of Lottery ($\beta = -1.325$, p-value $< .001$) and positive effects of Nonmonetary ($\beta = .735$, p-value $< .001$), highlighting the importance of providing tangible benefits to this group that encourage prosocial behavior. Vote shows a negative effect across all categories, which highlights the divide between party affiliation and ideological extremism. Thus, we modify our original assertion that there is support for **H8** when considering party affiliation, but not necessarily ideological extremism.

Impact of Incentives on Aggregate Vaccination Adoption

In this section, we explore the effect of incentives on the total percentage change of people choosing to get vaccinated within a county. Our analysis in the prior section focuses specifically on how incentives affect intrinsic motivation to get vaccinated, which relates to the speed at which diffusion happens in a market. However, policymakers care both about when prosocial adoption occurs and how many people then adopt that policy. Thus, understanding how incentives may affect overall vaccination rates is important as well.

To estimate the effect of incentives on overall vaccination rates, we proceed in the following steps. Our imputation approach earlier allowed us to construct a counterfactual $\dot{M}$, the market size, for each market in treatment. When we estimate the diffusion model with the imputed Bass parameters, we estimate a second $\hat{M}$, which is the realized number of people getting vaccinated in the market. We subtract the realized $\hat{M}$ by the counterfactual $\dot{M}$ and then divided by the population size of the county to quantify the percentage gains or loss in overall vaccination rate by county. From this exercise, after trimming the top and bottom 1%, we find that the average effect of incentives within a county is a loss of approximately 3%, with a p-value less than .001. Thus, we argue that incentives reduced the speed at which people got vaccinated and the overall level of vaccinations within the United States.

To understand what variables, either incentive or county characteristics, impact the overall percentage change in vaccination within a county, we modify our original estimation scheme. Our dependent variable is the culmination of actions taken by the states. States often ran multiple incentives during our observation window; thus, we need to account for the value and frequency of each incentive offered. Our independent variables are then our chosen moderators, other control variables, and information on the incentives being offered across the full sample window. The results of this exercise are presented in Table 5.

We start our analysis by examining our chosen moderators. Similar to our prior analysis, we find differential effects between party affiliation and ideological extremism. Examining the results in column 6, we find that Polar (For) is both negative and significant ($\beta = -.047$, p-value $< .001$). This result shows that regions who support COVID policies are actually less willing to get vaccinated when an incentive is offered, showing a backfire effect for the most entrenched group. This outcome is a different version of the backfire effect discussed earlier, but is similar to the findings in Banerjee et al. (2023), where the individuals most likely to engage in the behavior choose not to do so when an incentive is offered. However, there is no statistical effect for Polar (Against). Examining party affiliation, we find a negative and significant effect ($\beta = -.115$, p-value $< .001$), which is different than the directionality of Polar (For). Once more, there is support that **H8** is supported when considering party affiliation, but not necessarily ideological extremism.

Examining our other two moderators, we find consistency in their effects when compared to our prior utility analysis. Starting with the vulnerability of a region (Risk), we find a positive and statistically significant effect ($\beta = .065$, p-value $< .01$). Similarly, for income, we find a positive and statistically significant effect ($\beta = .093$, p-value $< .01$). Both of these results are congruent with our analysis examining the effect of interventions on the utility in the market. Thus, we find further support for hypotheses **H6B** and **H7B**.

Turning to the effects of the incentives on changes in vaccination rates, the results show additional support for our hypotheses. Direct incentives (monetary and nonmonetary) both show positive and statistically significant main effects ($\beta = .443$, p-value $< .001$) and ($\beta = .059$, p-value $< .001$), respectively. While indirect incentives (monetary and non-monetary) show mixed, but statistically significant, outcomes, ($\beta = .171$, p-value $< .001$) and ($\beta = -.097$, p-value $< .001$), respectively. Examining the incentive's amount, Cash, both nonmonetary incentives have positive and statistically significant coefficients while the opposite is true for the monetary incentives. The positive main effect and effect of value for the direct, nonmonetary incentive suggests support for our **H3**.

POLICY COUNTERFACTUAL

In this section, our goal is to better understand the role that incentives play in encouraging prosocial behavior. Our prior analysis focused on simple parametric models for interpretability and to provide direct evidence for our hypotheses. To enhance our findings and provide more meaningful guidance to public policymakers, we proceed with a series of counterfactual policy exercises to better understand the dimensions of heterogeneity. First, we investigate if there are non-linearities in the effect of an incentive's value allows for more optimal decision-making. Second, we investigate if there are differences in the characteristics of the counties that benefit from different incentive policies (both type and amount).

To generate the effects of different counterfactual policies, we proceed in the following stages. First, it is likely that the relationship between incentive type, incentive value, and county characteristics is likely non-linear. To generate a response surface that links the set of variables to an outcome, either the effect of the policy on utility or the overall effect on changing the vaccination total, we need a flexible estimation procedure. Our model of the response surface, regardless of dependent variable, uses LASSO. Using LASSO as a model allows us to include the full set of explanatory variables alongside, second and third order polynomials of each covariate, and the interactions between all variables. Thus, we create a flexible representation of the response surface. However, the inclusion of all these variables simultaneously would likely result in overfitting, which would not be useful in the counterfactual policy exercise. To guard against overfitting, LASSO identifies the optimal lambda (penalty factor) that minimizes overfitting. We use cross validation to find the optimal lambda and apply its value to obtain the final set of parameters for our counterfactual exercises. These parameters are then used to create our counterfactual exercise where a singular policy is applied to all counties in our sample.

Our first counterfactual policy exercise examines a set of incentives and incentive values as they relate to changes in both the effect of the incentive on utility and the overall change in vaccination amount. We examine all four incentive types, and choose incentive values that have support within our observed data. For each incentive policy, we collect a range of descriptive statistics: the overall change in vaccinations across the U.S., the number of counties that see a positive increase in vaccinations, the gain in vaccinations if only those counties who benefit are targeted with the policy, the average treatment effect, and the average percentage of people choosing to get vaccinated. The comparison of these descriptive statistics across policies provides greater insight into the usage of incentives to drive vaccinations. Table 6 presents the results of our first counterfactual exercise.

The results of this counterfactual policy exercise provide further support to our hypotheses. First, the results show that probabilistic incentives lead to substantial declines in vaccination uptake. For example, the use of lottery promotions encourages less than 2.5% of the counties to get vaccinated, regardless of the amount offered. Similarly, as the value of the scholarship increases, there is a decline in the number of counties benefiting from the incentive, from 8.8% to 4.2%. In contrast, direct incentives (both monetary and nonmonetary) are more effective at encouraging prosocial behavior. Specifically, our results show an approximately $100 direct monetary payout is where there is a tipping point in overall vaccinations being net positive if the incentive is not targeted. This finding aligns with experimental studies (Mardi et al. 2022; Serra-Garcia and Szech 2023; Klüver et al. 2021), which indicate that low-value monetary incentives dissuade individuals from vaccinating, but higher payouts lead to increased vaccination rates and encourages quicker vaccinations through the average treatment effect. Direct nonmonetary incentives show a similar increase in vaccinations as value increases. Direct nonmonetary incentives perform better than their direct monetary counterparts for the same value. Similarly, scholarship incentives perform better than lottery incentives. Through counterfactual analysis, we find final support for **H2**, **H3**, **H4**, and **H5**.

Our second counterfactual policy exercise, explores the heterogeneity in the characteristics of counties that benefit or suffer based on different incentive policies. The Web Appendix Section F details this analysis and presents a table of results. The findings of this exercise support our prior analysis. For example, counties that are ideologically opposed to COVID policies are more inclined to capitalize on any incentives offered, while party affiliation creates greater reactance to monetary payouts. The differential in characteristics of counties that benefit from the incentives show that income does not seem to be a major differentiator. In contrast, counties that would benefit most from COVID policies skew more towards benefiting from all policies. The findings show that even low values of monetary payouts and nonmonetary payouts both benefit at-risk communities. Examining the five different incentive plans in detail show clear targeting implications in the careful deployment of them.

DISCUSSION

We use the COVID-19 vaccination incentives implemented across the United States as a quasi-natural experiment to quantify how various moderators influence the diffusion of prosocial behavior. Additionally, we evaluate the impact of different incentives on both the diffusion process and overall changes in vaccination rates. By employing a new framework that blends the utility-based, Bass Diffusion model with a counterfactual estimation framework, we obtain county-level estimates for the effects of four distinct incentive types and assess how observed factors moderate their effectiveness.

Our findings indicate that, overall, the incentives slowed the progress of COVID-19 vaccinations, reducing the intrinsic motivation (utility) on average by 0.5 and resulting in a

3% reduction in vaccination rates. Specifically, we observe that probabilistic (Lottery-style and Scholarship) incentives generated significant backfiring, rendering them ineffective. In contrast, direct incentives—whether monetary payouts of sufficient value or nonmonetary rewards—proved most effective in encouraging both initial vaccination uptake and accelerating individuals' decisions to get vaccinated.

In this section, we delve deeper into our results, discussing the implications for both the academic literature and public policy decision-makers.

Academic Implications

Our study makes at least four novel contributions to the academy. First, we contribute to the growing body of work showing that nudges intended to enhance social good can backfire. Specifically, we demonstrate that the backfire effect can manifest in two key ways. The first dimension involves offering low-value incentives that may be perceived as trivial by recipients. Our findings show that lottery-based incentive—whether through direct monetary prizes or the chance of winning a scholarship—reduce the intrinsic motivation to get vaccinated. This effect is evident in both a slowing of the adoption process and a decrease in overall vaccination rates. The second dimension involves the moderating role of county-level characteristics. We find that counties with a large proportion of the population supportive of COVID-19 protocols (measured by ideological polarization on policies) exhibit a reduction in vaccination rates when incentives are introduced. This aligns with existing literature on backfiring. Conversely, in counties where entrenched political beliefs (beyond COVID-19-related views) dominate, there is resistance to vaccination even in the presence of incentives.

Second, our study highlights the importance of distinguishing political polarization by issue rather than simply by party affiliation. We create a measure linked to ideological extremism, *Polar*, which is the congruence in response to county-level preferences and outcomes regarding COVID-19 policies. Using this measure, we show that political polarization based on policy preferences yields different outcomes than polarization based solely on party lines.

We find that *Polar* correlates positively, albeit not strongly, with 2020 presidential voting patterns. When both *Polar* and party affiliation are included in our model, we observe divergent effects, which highlights the value of using a richer measure of political polarization, as relying solely on party affiliation would overlook critical nuances in vaccine adoption.

Third, by leveraging the utility-based, Bass Diffusion framework and existing practices that use analogs to forecast, we propose a novel approach to causal inference within the Bass model. Treating state-level incentive variations as a quasi-natural experiment, our approach extends the work of Cosguner and Seetharaman (2022) by enabling the direct estimation of heterogeneous effects through counterfactual estimation. Our simulations show that our approach yields unbiased estimates of treatment effects. In our empirical study, we validate that our method provides consistent and reliable estimates of the causal effects of incentive interventions through a series of robustness checks. Thus, we offer a new tool that allows for the estimation of treatment effects in settings where the Bass Diffusion model is appropriate.

Fourth, our study contributes to the literature on COVID-19 vaccine incentives by examining a broad range of incentives combined with local-level conditions. Previous research has focused either on a limited set of treatments (Walkey et al. 2021; Taber et al. 2021; Sehgal 2021) or employed experimental designs with narrow scope (Schneider et al. 2023; Kachurka et al. 2021; Klüver et al. 2021; Robertson et al. 2021). In contrast, our study considers the diverse landscape of treatment types and their interaction with local conditions, enabling us to assess the perceived changes in utility and subsequent vaccination outcomes. Our findings suggest that, overall, the COVID-19 incentives were largely ineffective, with many strategies backfiring due to their perceived low value and local-level resistance. However, we also identify certain incentive types that show promise in enhancing vaccination rates. Our results demonstrate that prosocial behavior can be nudged, but only through the careful and strategic application of appropriate interventions that consider local-level characteristics.

Managerial Contributions

This research provides at least four actionable insights that are useful for public policymakers. First, we propose that prosocial adoption can be modeled within the Bass Diffusion framework. By framing the adoption of prosocial behaviors in this way, policymakers can leverage established tools to forecast the diffusion of different incentives within a region. This predictive ability allows policymakers to better time the introduction of interventions, optimizing the chances of swaying the median individual in a given region.

Second, our findings suggest that outreach efforts for prosocial incentives should focus on core values, minimizing emphasis on party affiliation. As shown in our analysis, the most entrenched individuals—those most resistant to COVID-19 policies, as measured by our polarization metric—were most easily persuaded to vaccinate when offered an incentive. However, strong party affiliation stymied efforts. This duality is crucial: by reducing the focus on party lines and highlighting the intrinsic benefits of prosocial behavior, public policy advocates may successfully persuade individuals who are initially opposed to the policies.

Third, policymakers should exercise caution when implementing incentives to encourage prosocial behavior. Our study found that most incentives failed to drive significant increases in vaccination rates, often slowing the diffusion process. However, direct cash payments of sufficient value were the most effective in quickly encouraging prosocial adoption. Direct nonmonetary incentives also yielded positive results, although caution is warranted. Because we grouped all nonmonetary incentives into a single category due to identification concerns, there is potential variation in their effectiveness. Despite this, the overall effect was positive, suggesting that targeted nonmonetary incentives are more effective with further refinement.

Fourth, our study highlights potential avenues for messaging to encourage prosocial adoption. By examining the latent diffusion process in the absence of treatment, we observed that regions with strong opposition to COVID-19 policies accelerated vaccination adoption when confronted with rising COVID-19 death rates. The takeaway for policymakers is that presenting salient, evidence-based information on the consequences of inaction may sway those who are opposed to the behavior. Demonstrating the direct impact of the pandemic on their community may prompt faster adoption in regions that are most resistant.

ACKNOWLEDGMENT

The authors would like to thank Candace Jens, Nawar Chaker, and Murali Mantrala for their valuable feedback. The authors disclose the use of ChatGPT-o1 as an AI assisted writing tool to provide light copy editing suggestions on a few parts of the original written work of the authors.

REFERENCES

Argote, Pablo, Elena Barham, Sarah Daly, Julian Gerez, John Marshall and Oscar Pocasangre (2021), Messages that increase covid-19 vaccine willingness: Evidence from online experiments in six latin american countries. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3812023.

Banerjee, Sanchayan, Matteo M. Galizzi, Peter John and Susana Mourato (2023), 'Immediate backfire? nudging sustainable food choices and psychological reactance', *Food Quality and Preference* **109**, 104923.

Bigsby, Elisabeth, Holli H. , Scott D. Halpern, Kevin Volpp and Joseph N. Cappella (2017), 'Estimating acceptability of financial health incentives', *Health Education & Behavior* **44**(4), 513–18.

Bollinger, B, KT Gillingham and M. Ovaere (2020), 'Field experimental evidence shows that self-interest attracts more sunlight', *PNAS* **117**(34).

Bowles, Samuel and Sandra Polania-Reyes (2012), 'Economic incentives and social preferences: Substitutes or complements?', *Journal of Economic Literature* **50**(2), 368–425.

Byrne, Margaret, Rayner Kay Jin Tan, Dan Wu, Gifty Marley, Takhona Grace Hlatshwako, Yusha Tao, Jennifer Bissram, Sophie Nachman, Weiming Tang, Rohit Ramaswamy and Joseph D. Tucker (2023), 'Prosocial Interventions and Health Outcomes: A Systematic Review and Meta-Analysis', *JAMA Network Open* **6**(12), e2346789–e2346789.

Bénabou, Roland and Jean Tirole (2003), 'Intrinsic and Extrinsic Motivation', *The Review of Economic Studies* **70**(3), 489–520.

Bénabou, Roland and Jean Tirole (2006), 'Incentives and prosocial behavior', *American Economic Review* **96**(5), 1652–1678.

Böhm, Robert and Cornelia Betsch (2022), 'Prosocial vaccination', *Current Opinion in Psychology* **43**, 307–311.

Campos-Mercade, P., A. N. Meier, F. H. Schneider and E. Wengström (2021b), 'Prosociality predicts health behaviors during the covid-19 pandemic', *Journal of Public Economics* **195**, 104367.

Centola, Damon (2019), 'The 25 percent tipping point for social change'. Accessed: 2024-10-29. **URL:** *https://www.psychologytoday.com/us/blog/how-behavior-spreads/201905/the-25-percent-tipping-point-social-change*

Chang, Tom, Mireille Jacobson, Manisha Shah, Rajiv Pramanik and Samir B. Shah (2021), 'Financial incentives and other nudges do not increase covid-19 vaccinations among the vaccine hesitant', *NBER* **Working Paper**(29403).

Cosguner, Koray and Seethu Seetharaman (2022), 'Dynamic pricing for new products using a utility-based generalization of the bass diffusion model', *Management Science* **68**, 1904–1922.

Cryder, C. E., A. J. London, K. G. Volpp and G. Loewenstein (2010), 'Informative inducement: Study payment as a signal of risk', *Social Science & Medicine* **70**(3), 455–464.

Dimock, Michael and Richard Wike (2020), 'America is exceptional in the nature of its political divide', *Pew Research Center* .
URL: *https://www.pewresearch.org/short-reads/2020/11/13/america-is-exceptional-in-the-nature-of-its-political-divide/*

Dixit, Avinash K. and Jörgen W. Weibull (2007), 'Political polarization', *Proceedings of the National Academy of Sciences* **104**(18), 7351–7356.

Duquette, Nicholas (2020), 'Heard immunity: Effective persuasion for a future covid-19 vaccine'.

Fernandes, Daniel, Nailya Ordabayeva, Kyuhong Han, Jihye Jung and Vikas Mittal (2022), 'How political identity shapes customer satisfaction', *Journal of Marketing* **86**(6), 116–134.

Fishman, J., DS Mandell, MK Salmon and M Candon (2023), 'Large and small financial incentives may motivate covid-19 vaccination: A randomized, controlled survey experiment', *PLoS ONE* **18**(3), e0282518.

Frey, Bruno S. and Reto Jegen (2001), 'Motivation crowding theory', *Journal of Economic Surveys* **15**(5), 589–611.

Gagné, Marylène and Edward L Deci (2005), 'Self-determination theory and work motivation', *Journal of Organizational behavior* **26**(4), 331–362.

Gayer, Ted and Emily Parker (2013), 'Cash for clunkers: An evaluation of the car allowance rebate system'. Accessed: 2024-10-29.
URL: *https://www.brookings.edu/articles/cash-for-clunkers-an-evaluation-of-the-car-allowance-rebate-system/*

Gneezy, Uri, Stephan Meier and Pedro Rey-Biel (2011), 'When and why incentives (don't) work to modify behavior', *Journal of economic perspectives* **25**(4), 191–210.

Goette, Lorenz and Alois Stutzer (2019), 'Blood donations and incentives: Evidence from a field experiment'.

Hendren, Nathaniel and Ben Sprung-Keyser (2020), 'A unified welfare analysis of government policies', *Quarterly Journal of Economics* **135**(3), 1209–1318.

Hoekstra, Mark, Steven L. Puller and Jeremy West (2017), 'Cash for corollas: When stimulus reduces spending', *American Economic Journal: Applied Economics* **9**(3), 1–35.

Judge, Timothy A. and Charlice Hurst (2007), 'Capitalizing on one's advantages: Role of core self-evaluations', *Journal of Applied Psychology* **92**(5), 1252–1262.

Jung, Jihye and Vikas Mittal (2020), 'Political identity and the consumer journey: A research review', *Journal of Retailing* **96**(1), 55–73. Understanding Retail Experiences and Customer Journey Management.

Jungkunz, Sebastian, Marc Helbling and Nina Osenbrügge (2024), 'Measuring political radicalism and extremism in surveys: Three new scales', *PLOS ONE* **19**(5), 1–16.

Kachurka, R., M. Krawczyk and J. Rachubik (2021), 'Persuasive messages will not increase covid-19 vaccine acceptance: Evidence from a nationwide online experiment', *Vaccines (Basel)* **9**(10), 14–20.

Kamenica, Emir (2012), ‘Behavioral economics and psychology of incentives’, *Annual Review of Economics* **4**(1), 427–452.

Karaivanov, Alexander, Dongwoo Kim, Shih En Lu and Hitoshi Shigeoka (2022), ‘Covid-19 vaccination mandates and vaccine uptake’, *Nature Human Behaviour* **6**, 1615–1624.

Kim, Hwang and Vithala R. Rao (2021), ‘Vaccination diffusion and incentive: Empirical analysis of the us state of michigan’, *Frontiers in public health* **9**, 740367.

Kivetz, Ran (2005), ‘Promotion reactance: The role of effort-reward congruity’, *Journal of consumer research* **31**(4), 725–736.

Klüver, Heike, Felix Hartmann, Macartan Humphreys, Ferdinand Geissler and Johannes Giesecke (2021), ‘Incentives can spur covid-19 vaccination uptake’, *Proceedings of the National Academy of Sciences* **118**(36), e2109543118.

Kuznetsova, Lidia, Elizabeth Diago-Navarro, Rachel Mathu and Antoni Trilla (2022), ‘Effectiveness of covid-19 vaccination mandates and incentives in europe’, *Vaccines (Basel)* **10**, 1714.

Li, Meng, Eric G. Taylor, Katherine E. Atkins, Gretchen B. Chapman and Alison P. Galvani (2016), ‘Stimulating influenza vaccination via prosocial motives’, *PLOS ONE* **11**(7), 1–14.

Liaukonytė, Jūra, Anna Tuchman and Xinrong Zhu (2023), ‘Frontiers: Spilling the beans on political consumerism: Do social media boycotts and buycotts translate to real sales impact?’, *Marketing Science* **42**(1), 11–25.

Liu, Licheng, Ye Wang and Yiqing Xu (2024), ‘A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data’, *American Journal of Political Science* **68**(1), 160–176.

Loewenstein, G. and C. Cryder (2020), ‘Why paying people to be vaccinated could backfire’, *The New York Times* p. 810323.

Macchia, Lucía and Ashley V. Whillans (2021), ‘The link between income, income inequality, and prosocial behavior around the world’, *Social Psychology* **52**(6), 375–386.

Mardi, Parham, Shirin Djalalinia, Reza Kargar, Mahnaz Jamee, Zahra Abdar Esmaeili and Mostafa Qorbani (2022), ‘Impact of incentives on covid-19 vaccination; a systematic review’, *Frontiers in Medicine* **9**, 810323.

Mian, Atif and Amir Sufi (2012), ‘The effects of fiscal stimulus: Evidence from the 2009 ’cash for clunkers’ program’, *The Quarterly Journal of Economics* **127**(3), 1107–1142.

Milkman, Katherine L., Linnea Gandhi, Sean F. Ellis, Heather N. Graci, Dena M. Gromet, Rayyan S. Mobarak, Allison M. Buttenheim, Angela L. Duckworth, Devin Pope, Ala Stanford, Richard Thaler and Kevin G. Volpp (2021), ‘An experiment evaluating the impact of large-scale, high-payoff vaccine regret lotteries’, *SSRN Electronic Journal* p. Advance online publication.

Murayama, Kou, Madoka Matsumoto, Keise Izuma and Kenji Matsumoto (2010), ‘Neural basis of the undermining effect of monetary reward on intrinsic motivation’, *Proceedings of the National Academy of Sciences* **107**(49), 20911–20916.

Pasquini, Giancarlo and Emily Saks (2022), ‘Partisan differences are common in the lessons americans take away from covid-19’. Accessed: 2024-10-29.
URL: *https://www.pewresearch.org/short-reads/2022/09/06/partisan-differences-are-common-in-the-lessons-americans-take-away-from-covid-19*

Patkos, Veronika (2020), ‘Measuring partisan polarization with partisan differences in satisfaction with the government: the introduction of a new comparative approach’, *Quality & Quantity* **57**, 39–57.

Piff, Paul K, Daniel M Stancato, Stéphane Côté, Rodolfo Mendoza-Denton and Dacher Keltner (2012), 'Higher social class predicts increased unethical behavior', *Proceedings of the National Academy of Sciences* **109**(11), 4086–4091.

Queensland Government (2024), 'Climate smart energy savers rebate'. Accessed: 2024-10-29.
URL: *https://www.qld.gov.au/housing/home-modifications-energy-savings/climate-smart-energy-savers-rebate*

Robertson, E, KS Reeve, CL Niedzwiedz, J Moore, M Blake, M Green, SV Katikireddi and MJ Benzeval (2021), 'Predictors of covid-19 vaccine hesitancy in the uk household longitudinal study', *Brain Behav Immun.* pp. 41–50.

Ryan, Richard M and Edward L Deci (2000), 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.', *American psychologist* **55**(1), 68.

Sallam, M (2021), 'Covid-19 vaccine hesitancy worldwide: A concise systematic review of vaccine acceptance rates', *Vaccines* **9**(2), 160.

Schedler, Andreas (2023), 'Rethinking Political Polarization', *Political Science Quarterly* **138**(3), 335–359.

Schneider, Florian H., Pol Campos-Mercade, Stephan Meier, Devin Pope, Erik Wengström and Armando N. Meier (2023), 'Financial incentives for vaccination do not have negative unintended consequences', *Nature* **613**(7944), 526–533.

Schoenmueller, Verena, Oded Netzer and Florian Stahl (2022), 'Frontiers: Polarized america: From political polarization to preference polarization', *Marketing Science* **42**(1).

Schwalbe, Nina, Layth Hanbali, Marta C. Nunes and Susanna Lehtimaki (2022), 'Use of financial incentives to increase adult vaccination coverage: A narrative review of lessons learned from covid-19 and other adult vaccination efforts', *Vaccine: X* **12**, 100225.

Sehgal, Neil K.R. (2021), 'Impact of vax-a-million lottery on covid-19 vaccination rates in ohio', *The American Journal of Medicine* **134**(11), 1424–1426.

Serra-Garcia, Marta and Nora Szech (2023), 'Incentives and defaults can increase covid-19 vaccine intentions and test demand', *Management Science* **69**(2), 1037–1049.

Sprengholz, Philipp, Luca Henkel, Robert Böhm and Cornelia Betsch (2023*a*), 'Different interventions for covid-19 primary and booster vaccination? effects of psychological factors and health policies on vaccine uptake', *Medical Decision Making* **43**(2), 239–251. PMID: 36404766.

Sprengholz, Philipp, Luca Henkel, Robert Böhm and Cornelia Betsch (2023*b*), 'Historical narratives about the covid-19 pandemic are motivationally biased', *Nature* **623**, 588–593.

Suss, Joel H. (2023), 'Higher income individuals are more generous when local economic inequality is high', *PLOS ONE* **18**(6), 1–17.

Taber, JM, CA Thompson, PG Sidney, A O'Brien and J Updegraff (2021), 'Promoting vaccination with lottery incentives', *PsyArXiv* p. Advance online publication.

Thirumurthy, H., Katherine L. Milkman, Kevin G. Volpp, Alison M. Buttenheim and Devin G. Pope (2022), 'Association between statewide financial incentive programs and covid-19 vaccination rates', *PLoS One* **17**(3), e0263425.

Visser, Michael S and Matthew R Roelofs (2011), 'Heterogeneous preferences for altruism: Gender and personality, social status, giving and taking', *Experimental Economics* **14**, 490–506.

Walkey, Allan J., Anica Law and Nicholas A. Bosch (2021), 'Lottery-Based Incentive in Ohio and COVID-19 Vaccination Rates', *JAMA* **326**(8), 766–767.

World Bank (2010), 'Bolsa família: Changing the lives of millions'. Accessed: 2024-10-29.
URL: *https://www.worldbank.org/en/news/feature/2010/05/27/br-bolsa-familia*

Xiao, Fengqiu, Zhiwei Zheng, Heyi Zhang, Ziqiang Xin, Yinghe Chen and Yiwei Li (2016), 'Who are you more likely to help? the effects of expected outcomes and regulatory focus on prosocial performance', *PLOS ONE* **11**(11), 1–15.

Zhang, Xinrui and Tom Lane (2023), 'The backfiring effects of monetary and gift incentives on covid-19 vaccination intentions', *China economic review* **Epub**.

Table 1: Descriptive Statistics of Treatment and Control Counties. This table contains the pooled averages of our variables of interest for the 2,269 counties in our study (1,232 treatment and 1,037 control). Our dependent variable is the bi-weekly number of new COVID vaccinated individuals, we present the median of that value. We have a time-varying moderator, the number of bi-weekly deaths per 10,000 people in the county, presented as the median across the county. Promo is an indicator if the county ever received a public policy incentive. The remaining variables are all static variables that describe the demographics and psychographics of the county. Sev. Hesitant and Hesitant are both measures of vaccine hesitancy. Mask is the percentage of people wearing a mask in the county. Mobility is a measure of the average mobility of residents before the start of our observation window. Polarity is our constructed measure of political polarization. Risk is CVAC measure from the CDC and Income is the median Income of the county, scaled by $10,000. Vote is the vote share differential in the 2020 election. Population is the number of people in the county, scaled by 10,000. Medical is the number of designated medical facilities within the county. Unemploy is the average unemployment rate of the county. Minor is the percentage of minorities in the county. Female is the percentage of people identifying as female in the county. SVI is the social vulnerability of the region. Last, M Code is an indicator of the county being rural (1) or a metro area (6).

Group	Name	Mean	St. Dev.	Minimum	Median	Maximum
Dependent Variable	Vax	611.805	2201.645	3	126	63890
Time Varying Moderator	Deaths	0.122	0.155	0	0	0.831
Treatment	Promo	0.543	0.498	0	1	1
COVID Policy Metrics	Sev. Hesitant	0.090	0.033	0.019	0.088	0.182
	Hesitant	0.138	0.045	0.037	0.139	0.267
	Mask	0.503	0.147	0.115	0.490	0.884
	Mobility	-23.931	6.315	-67	-22.7472	-4.0565
	Polarity	-0.237	1.539	-5.078	-0.403	4.942
Other Moderators	Risk	0.490	0.279	0	0.5	1
	Income	5.546	1.416	2.635	5.329	13.523
County-Level Variables	Vote	0.239	0.284	-0.738	0.216	0.829
	Pop	11.518	35.581	0.363	3.069	1003.911
	Medical	641.216	5570.398	1	89	244455
	Unemploy	6.844	2.108	2.2	6.6	22.5
	Minor	0.178	0.178	0.011	0.102	1.000
	Female	0.501	0.020	0.311	0.504	0.570
	SVI	0.511	0.282	0	0.51	1
	M Code	4.537	1.494	1	5	6

Table 2: Effects of Incentive Type and Incentive Amount on Treatment Effect Estimates This table presents the effects of the different incentive types in our study on the estimated treatment effects. For this exercise, direct monetary payments are considered the base case and the other three types of incentives are then shown as differential from that base effect. We also include variables for how long the incentive is run, the total monetary value (if any) that is logged, and how many incentives are included in that effect. As shown, monetary has the most positive effect among all types, longer incentives tend to have higher effects, and having larger payouts have a more positive effect as well.

	EST	SE	Sig
Intercept	-0.266	0.065	***
Lottery	-0.931	0.071	***
Nonmonetary	-0.245	0.063	***
Scholarship	-0.199	0.080	*
Duration	0.022	0.006	***
Cash	0.054	0.006	***
Num Promo	-0.130	0.068	.
N Obs	2195		
Adj R squared	0.097		

.p <.10; *p <.05; **p <.01; ***p <.001.

Table 3: Effects of County-Level Moderators and Control Variables on Treatment Effect Estimates This table presents the effects of moderators of interested and demographic/psychographic control variables. Pooling across all incentive types shows that Polarity (Against) and Vote move in the same direction (i.e., against the incentive). Polar (For) has higher effects across the incentives. Both Risk and Income are statistically significant and positive, showing that communities with higher risk and/or higher incomes both view incentives positively. These results are robust with the inclusion of our moderator variables related to community demographics/psychographics.

	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig
Intercept	-0.996	0.211	***	-1.147	0.209	***	-0.888	0.215	***	-0.711	0.652	
Polarity	0.138	0.015	***									
Polar (Against)				-0.289	0.053	***	-0.258	0.053	***	-0.183	0.056	**
Polar (For)				0.292	0.052	***	0.185	0.056	**	0.130	0.058	*
Risk	0.251	0.104	*	0.245	0.105	*	0.168	0.106		0.249	0.118	*
Income	0.198	0.111	.	0.292	0.107	**	0.226	0.107	*	0.405	0.127	**
Vote							-0.382	0.081	***	-0.414	0.088	***
Minor										-0.383	0.127	**
SVI										-0.016	0.126	
Female										-1.706	1.129	
Unemploy										0.057	0.012	***
N Obs		2195			2195			2195			2195	
Adj R Squared		0.064			0.064			0.073			0.085	

.p <.10; *p <.05; **p <.01; ***p <.001.

Table 4: Evaluation of County-Level Heterogeneity on incentive Effectiveness in Driving Diffusion Patterns. This table presents the evaluation of our moderators and control variables effect on different incentive types. We isolate observations where only a single incentive is occurring at a time that is found in multiple states. As such, we do not have enough data support for the Scholarship incentive type to evaluate in this manner. These results show a consistent negative, and mostly statistically significant, effect of political affiliation on all incentive types. This result is in stark contrast to Polar (Against), where there is a positive, statistically significant effect on Monetary and Lottery incentives. Risky populations react negatively to Lottery incentives and react positively to Nonmonetary incentives. The value of the incentive, if indicated, is logged and used for our Cash variable, which shows a positive association across all three incentive types (though it is not significant for the direct Monetary payments). Last, higher income areas show a positive association with Monetary and Lottery incentives. These results highlight that different pools of people react differently to the context of the inducement.

	Monetary			Lottery			Nonmonetary		
	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig
Intercept	-2.703	1.836		-0.187	1.099		0.045	1.335	
Polar (Against)	0.999	0.204	***	0.170	0.101	.	-0.412	0.124	***
Polar (For)	0.121	0.205		0.060	0.101		-0.145	0.127	
Risk	0.400	0.499		-1.325	0.215	***	0.973	0.312	**
Income	1.166	0.401	**	0.735	0.219	***	0.135	0.260	
Cash	0.068	0.146		0.057	0.009	***	0.179	0.093	.
Vote	-0.594	0.359	.	-0.688	0.170	***	-0.152	0.173	
Minor	-0.742	0.302	*	-0.267	0.222		-0.304	0.250	
SVI	0.430	0.355		1.685	0.203	***	-0.798	0.263	**
Female	-0.579	2.584		-4.639	1.867	*	-2.498	2.363	
Unemploy	0.016	0.037		-0.002	0.021		0.087	0.025	***
N Obs	262			587			727		
Adj R Squared	0.254			0.335			0.073		

.p <.10; *p <.05; **p <.01; ***p <.001.

Table 5: Effect of County-Level Characteristics and Incentive Heterogeneity on Changes in Vaccine Adoption This table presents the effects of both the moderators and the incentive information on changes in total vaccine adoption at the county level. We compute the total percent lift in vaccine adoption within a market as the difference between the estimated market size from the treated observation and the imputed market size from the analog (control) observations. This variable is then scaled by the population size of the county and serves as our dependent variable. As shown, Polar (For) mitigates the effect of incentives, showing a backfire effect of the people most in favor of COVID-19 policies. Additionally, those with higher Risk and Income are more likely to get vaccinated due to incentives. This result is in stark difference to the Vote metric, where more Republic counties react negatively to incentives in getting vaccinated. The difference in outcomes between Polarity and Voter Identity one again highlights that they are two different mechanisms. Monetary incentives drive more vaccination uptake, but it is moderated by the cash value of the incentive. Note, all Cash measures are logged to test if there are diminishing returns to the face value of the incentive. Further, Hospitals is logged as well.

	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig	EST	SE	Sig
(Intercept)	-0.089	0.033	**	-0.030	0.034		-0.092	0.041	*	-0.122	0.040	**	-0.234	0.038	***	-0.203	0.092	*
Polar (Against)	0.029	0.008	***	0.037	0.008	***	0.037	0.008	***	0.008	0.009		-0.006	0.008		0.001	0.008	
Polar (For)	-0.054	0.008	***	-0.084	0.009	***	-0.090	0.010	***	-0.089	0.009	***	-0.059	0.009	***	-0.047	0.009	***
Risk	-0.025	0.016		-0.037	0.016	*	-0.013	0.017		0.022	0.018		0.069	0.017	***	0.061	0.019	**
Income	0.044	0.017	**	0.031	0.017	.	0.050	0.018	**	0.055	0.018	**	0.069	0.016	***	0.093	0.020	***
Vote				-0.098	0.013	***	-0.100	0.013	***	-0.111	0.013	***	-0.067	0.013	***	-0.115	0.015	***
Num Monetary							-0.010	0.008		0.014	0.008	.	0.529	0.039	***	0.443	0.040	***
Num Lottery							0.005	0.008		0.107	0.017	***	0.171	0.016	***	0.153	0.016	***
Num Nonmonetary							0.016	0.005	**	0.039	0.006	***	0.059	0.006	***	0.048	0.006	***
Num Scholarship							0.019	0.009	*	-0.005	0.009		-0.097	0.027	***	-0.099	0.027	***
Cash										-0.009	0.001	***						
Cash - Monetary													-0.136	0.010	***	-0.116	0.010	***
Cash - Lottery													-0.011	0.001	***	-0.010	0.001	***
Cash - Nonmonetary													0.047	0.005	***	0.043	0.005	***
Cash - Scholarship													0.008	0.002	**	0.009	0.002	***
Unemploy																-0.005	0.002	**
SVI																0.006	0.019	
Female																0.148	0.153	
Minor																0.028	0.018	
Hospitals																-0.019	0.003	***
Num Obs		1206			1206			1206			1206			1206			1206	
Adj R Squared		0.051			0.096			0.110			0.144			0.299			0.336	

.p <.10; *p <.05; **p <.01; ***p <.001.

Table 6: Counterfactual Analysis of Incentive Effectiveness at Different Incentive Value Levels. This table presents the summary results of counterfactual analysis that examines the four different types of incentives and different dollar values associated with each one. To construct the counterfactuals, we estimate the response surface between incentives and county characteristics via LASSO for both the effect of treatment (Treat) and percentage change in population based on our structural estimates for all counties in our study. Once the response surface is estimated, we choose to include 1 incentive of each type at a time and change its value. We show four descriptive statistics. No Target is the total change in population if the incentive is broadcast across all counties in our study. Yes Target assumes we only give the incentive to those counties where the change in population is positive. % Counties is the percentage of all counties in our study where the effect in vaccination rates is positive. Avg. Treat is the average treatment effect in the diffusion model, which represents if the incentive either accelerates (positive) or decelerates (negative) the diffusion process. Avg. M Diff is the percentage difference change in vaccinations due the promotion, on average. As shown, only Monetary and Nonmonetary incentives result in gains in the total vaccinations overall. Higher direct payments (monetary) result in both more vaccinations and an acceleration in the diffusion process. Non-direct payments (Lottery and Scholarship) both result in reduced vaccination rates overall. We note that, on average, the value of the treatment changes with increases in the Monetary payment, not the other promotions. The numbers in No Target and Yes Target are in 100,000 increments.

Promo Type	Value	No Target	Yes Target	% Counties	Avg. Treat	Avg. M Diff
Monetary	$10	-54.765	1.869	23.9%	-0.413	-0.028
Monetary	$25	-52.004	2.620	23.1%	-0.413	-0.030
Monetary	$50	-44.467	5.421	23.4%	-0.397	-0.031
Monetary	$75	-29.951	10.654	25.1%	-0.332	-0.028
Monetary	$100	-5.548	24.533	27.4%	-0.187	-0.020
Monetary	$125	28.206	49.052	35.1%	0.060	-0.007
Monetary	$150	86.090	100.323	49.4%	0.446	0.014
Lottery	$1 Million	-97.465	0.019	2.4%	-0.549	-0.061
Lottery	$2 Million	-98.289	0.016	2.2%	-0.557	-0.062
Lottery	$ 3 Million	-99.536	0.012	1.7%	-0.581	-0.063
Nonmonetary	0	-59.059	1.656	30.0%	-0.523	-0.021
Nonmonetary	$10	-26.158	10.498	53.7%	-0.523	-0.001
Nonmonetary	$20	61.174	77.132	88.6%	-0.526	0.070
Scholarship	$50K	-66.824	2.223	8.8%	-0.227	-0.048
Scholarship	$100K	-69.993	1.798	8.1%	-0.226	-0.049
Scholarship	$150K	-78.504	0.882	6.3%	-0.226	-0.054
Scholarship	$200K	-95.827	0.392	4.2%	-0.224	-0.062

Figure 1: Representation of Incentives Studied based on Contextual Characteristics. This figure presents the four types of incentives we examine in our study. We delineate each incentive on two dimensions: type (monetary vs. nonmonetary) and psychological distance (direct vs. probabilistic). For each type of incentive, we provide an example from our study, descriptive statistics of the amount associated with the incentive, and the number of states using such an incentive.

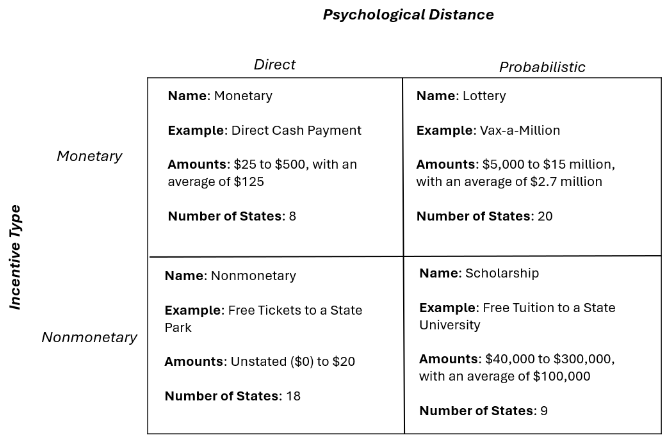


Figure 2: Conceptual Model. This figure presents our conceptual model, which estimates the latent utility to adopt prosocial behaviors (vaccinations) in two pieces. First, there are individual characteristics that drive intrinsic motivation for prosocial behavior adoption (vaccination), which describe a set of our Bass Diffusion parameters linked to part of the latent utility function. Second, incentives function as an extrinsic reward that may further impact the utility to engage in prosocial behavior, represented as our estimated τ and change in market size. Our conceptual model contains hypotheses for the direct effect of incentives (H1), the moderating effect of incentive contextual features (H2 - H5), and the moderating effects of individual contextual features (H6a - H8).

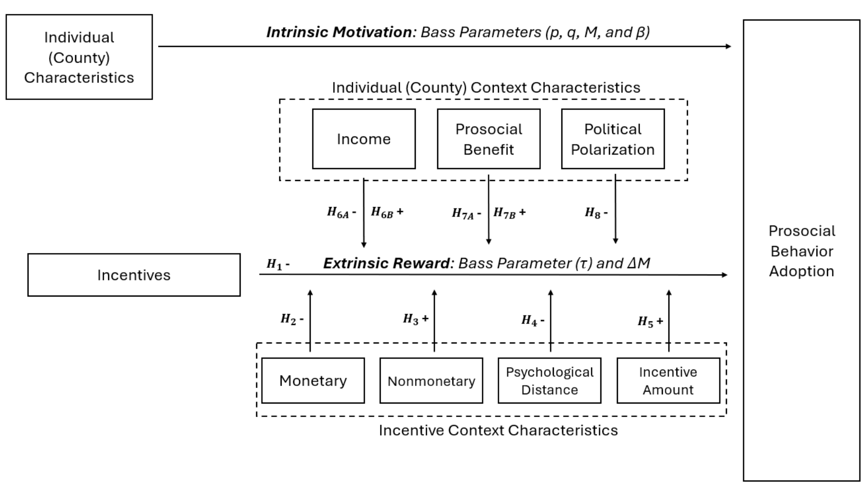


Figure 3: County-Level Measure of Polarization Metric in the United States This figure presents the polarization metric for each county in our study. Polarization is calculated via PCA reduction on three key dimensions: vaccine hesitancy, willingness to mask, and willingness to engage in social distancing. For each county that has all measures, we then calculate the polarization score, where polarization is defined as congruence in response across a series of policies that may be interrelated. In our context, a lower polarization score (red) represents being against COVID policies (i.e., vaccine hesitant, unwilling to mask, and not engaging in social distancing), while higher polarization scores (blue) represents the opposite. Counties that are gray are excluded from our analysis as they may be missing a key variable, generally the variable related to social distancing (Google Mobility metric). Thus, many rural counties are excluded from our analysis. Please note, even though Connecticut is shaded in gray, that is due to a mismatch between their counties in our sample period versus the new commonwealth structure adopted post our observation period.

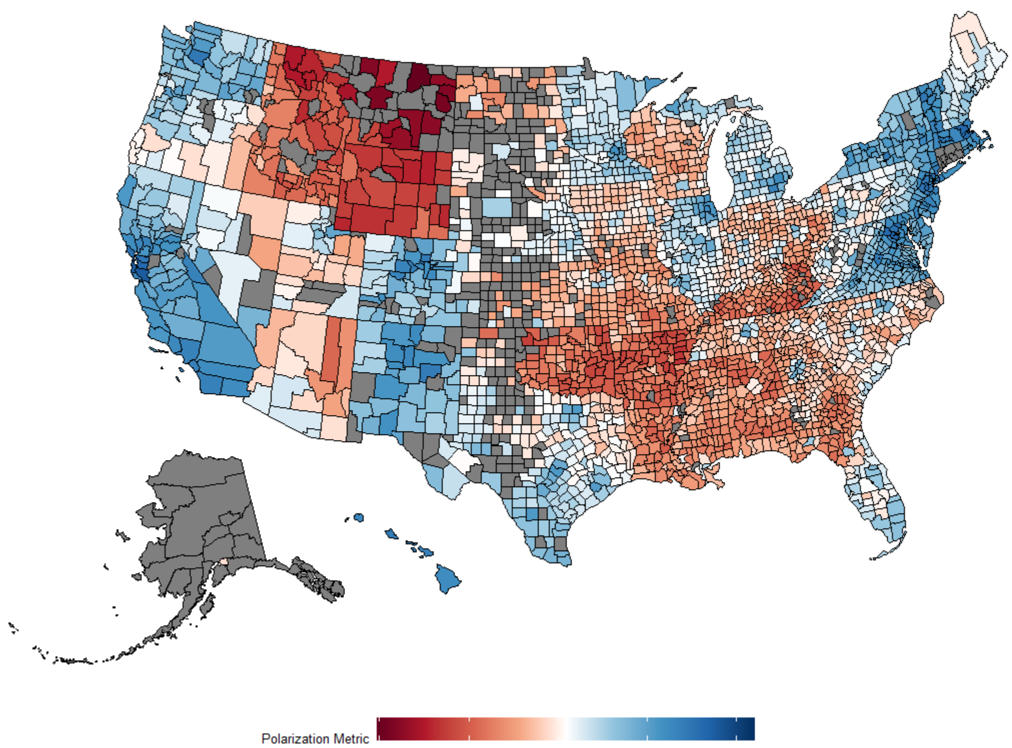
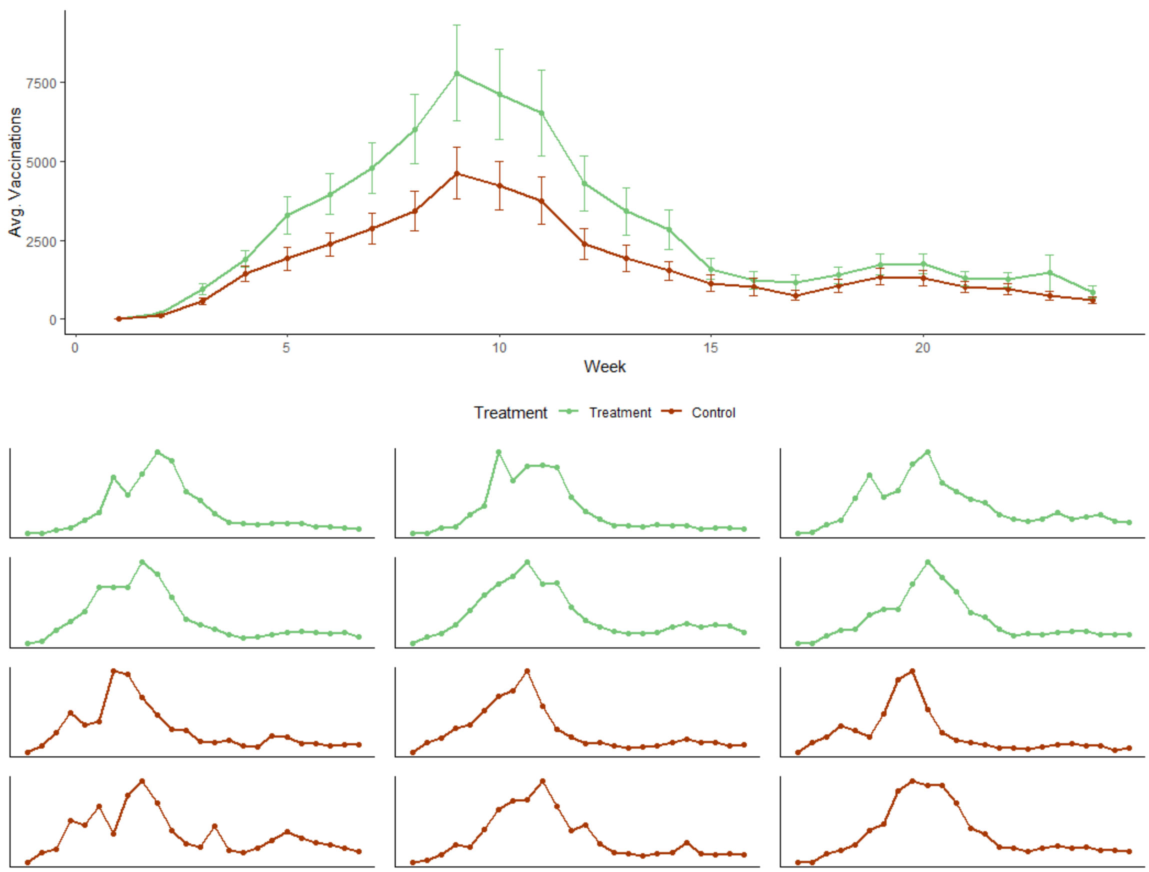


Figure 4: Aggregate Diffusion Patterns for Treated and Control (Analog) Counties. This figure presents the average diffusion curves of vaccination adoption, discretized by two-week intervals, separated by both counties that received an incentive (treatment in green) and those that did not (control in brown). Also presented in the figure are bars associated with the 95% interval. As shown, both curves show the classic diffusion pattern where, there is a starting portion of low vaccinations, that then accelerate and taper off. Further, this figure shows that the treatment counties from periods 5 to 15 show larger amounts of heterogeneity than the control counties, which speaks to the potential for the incentives to have effects during this time. An examination of median values (not shown in the figure) within groups shows a similar diffusion pattern. Below the main plot, we have a sample of 12 counties from our study. The top six counties are from the treatment group (in green) and the bottom six counties are from the control (analog) group (in brown). As shown, there are differences in the acceleration of adoption across counties, we use this variation to estimate the effects of incentive on prosocial adoption.



Web Appendix

Heterogeneous Incentives and Their Impact on Prosocial Behavior: Evidence from COVID-19 Vaccinations

These materials have been supplied by the authors to aid in the understanding of their paper. The AMA is sharing these materials at the request of the authors.

Table of Contents

ANCILLARY DOCUMENTATION OF INCENTIVES

In this section, we provide further information of the incentives we study. We provide more details on the incentives themselves, the amount associated with the incentives, the timing of the incentives, and which states offered the incentives.

Monetary incentives are those that offer a direct cash payment to a vaccine recipient. While the terms may differ (only government employees versus all vaccine recipients), these incentives offer a small cash payment. The range of payments is from $25 to $500 with an average of $125 and a mode of $100. As shown in Figure IA.1, Monetary incentives are found in 8 states.

Lottery incentives are those where the individual receives a ticket for a lottery being run with a chance to win a large cash prize. The difference between Lottery and Monetary incentives is that Monetary has guaranteed terms for a small cash payment, while Lottery relies on only having a chance at a large cash prize. For those with posted payout amounts, the values range from $5,000 to $15 million dollars. The average payout within the sample is approximately $2.7 million dollars. As shown in Figure IA.1, Lottery incentives are found in 20 states.

Nonmonetary incentives are a heterogeneous mix of incentives where the states offer an immediate, but non-direct monetary benefit for getting a vaccination. For example, these could be free drink tickets, reduced license fees, access to state parks, trips around a racetrack, admittance into amusement parks, and other associated items. In two instances, we see direct mentions of the cash value of this incentive, which is $20, otherwise, we do not operationalize a direct value for this incentive. As shown in Figure IA.1, Nonmonetary incentives are found in 18 states.

Scholarship incentives are those where the individual receives an opportunity to win a scholarship for vaccination. Generally, these incentives rely on the individual being under the age of 18 to get access to the incentive. For those with posted scholarship amounts, the values range from \$40,000 to \$300,000. Where possible, we imputed the value of a one-year college education at the flagship state school for those with enough details to pin down the nature of the incentive. The average payout (either stated or imputed) within the sample is approximately \$100,000. As shown in Figure IA.1, Lottery incentives are found in 9 states.

As shown, within incentive type, there is heterogeneity in the value. Further, incentive windows may eclipse each other as states ran incentives for varying lengths of time. The shortest incentive ran for 2 weeks (1 week for the announcement week and then one week thereafter). The longest incentive window is 18 weeks. The average is approximately 7 weeks total. Figure IA.2 shows the incentive window for the incentives in each state. Note, this figure shows only the presence of an incentive occurring, there are situations where multiple incentives are occurring at the same time. We use the differential in timing windows to then estimate the effects of each incentive type on the overall vaccination rates and whether the incentive accelerated or decelerated vaccination adoption.

Figure IA.1: Use of Incentives by State and Type. This figure presents four panels that show if a state used a specific type of incentive to encourage COVID-19 vaccinations. The top left panel is Monetary, direct cash payment, incentive. The top right panel is Lottery (Lotto), a lottery ticket for a chance to win a large prize, incentive. The bottom left panel is Nonmonetary, any incentive that is not directly tied to direct cash payments, incentive. The bottom right panel is Scholarship, a chance for an individual to get a scholarship at a state university, incentive. As shown, there are various states using different types of incentives and, sometimes, even multiple incentives.

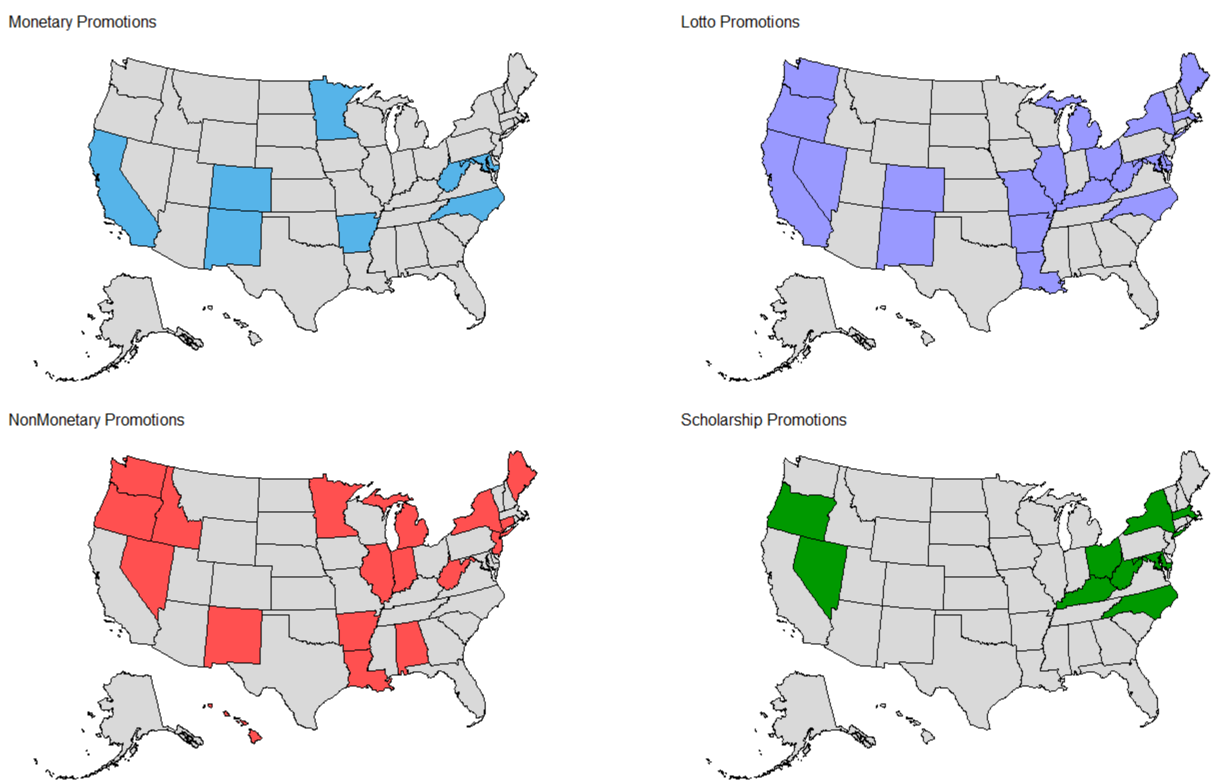
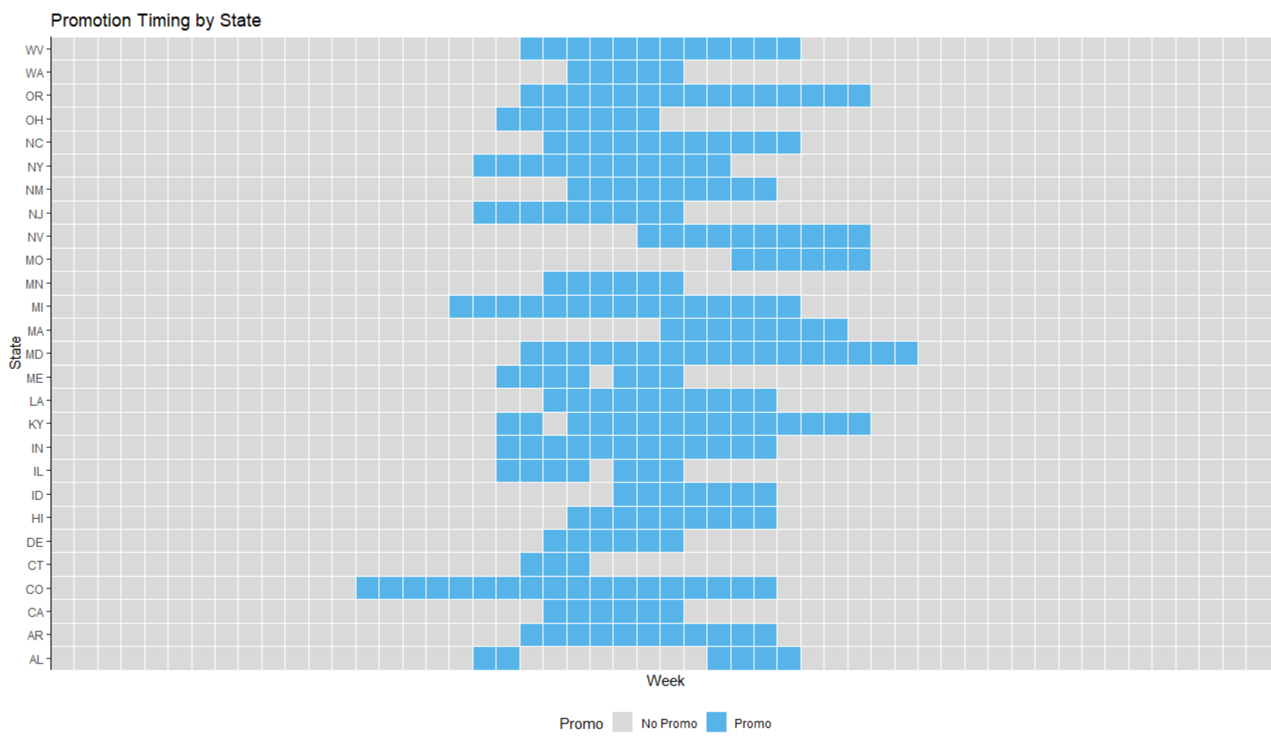


Figure IA.2: Incentive Timing by State. This figure presents the timing of incentive offerings by state. We show the presence of an incentive, of any type, being offered during a week by each state. Note, some states have overlapping intervals of different incentives or concurrent incentives being offered. As shown, there is heterogeneity in the timing and duration of incentives, which allows us to better estimate the effects of the different incentives.



POLARIZATION METRIC COMPUTATION

In this section, we provide further detail in how we calculate our measure of ideological extremism, a proxy for political polarization. We start by detailing the various sources of data used to construct this measure. Next, we explain how the variable is constructed. Last, we provide evidence that, while correlated with political affiliation, this construct is different than the more commonly used measure to test political polarization.

We identify three major policy categories for COVID-19 and collect variables for each one. Those three policy categories are: vaccination, masking, and social distancing. Our measures of vaccination hesitancy come from the CDC report, where we collect two measures: if the individuals within the region are hesitant and/or strongly hesitant to receiving a vaccination. We choose these two measures as, a priori, we do not know which would be the best to describe polar views of vaccine hesitancy. Our measure of masking comes from a survey administered by The New York Times and Dynata survey in July 2020 that asked 250,000 people the frequency they would wear a mask.[1] From this survey, the researchers construct a percentage estimate of individuals who engage in mask wearing. Mask wearing, as another form of COVID-19 prevention, may also be relevant to individuals' vaccination decisions and predictive of the vaccination rates in a county. Our measure of social distancing comes from the Google Mobility report, which provides a series of measures that describe how much individuals are moving within a county.[2] We use the Workplaces measure as a suitable proxy for social mobility within a county and average the measure for each county until the week prior to our sample window.

Each county may have varying degrees of adherence to any group of these policies. Our goal is to map out the latent variable that describes the overall willingness of the county to engage with COVID-19 policies. As such, we conduct a PCA analysis on these four variables to construct that latent measure, which we call POLAR. Table IA.1 shows the pairwise

[1] https://github.com/nytimes/covid-19-data/tree/master/mask-use
[2] https://www.google.com/covid19/mobility/

correlations between each of our variables used to construct our PCA reduced variable, as well as our measure of political affiliation. Hesitant and Sev. Hesitant show strong correlations by definition. For example, Mask shows negative correlation with the other variables, indicating that as people are more willing to mask, they are less vaccine hesitant and reduce their social distancing (mobility). We also see that VOTE is correlated in the appropriate directions with expectations (i.e., Republicans are generally against masking, vaccinations, and social distancing); however, there is not a strong correlation. This speaks to there being a potential underlying, latent variable that may capture polarization on issues that is not directly captured by political affiliation.

Using PCA reduction, we find that one variable explains 61.8% of the variation across the four variables used to construct POLAR. The correlation between POLAR and VOTE is -.457, which is the expected direction, but still not a particular strong correlation. Examining the PCA Loadings, Sev. Hesitant is -.577, Hesitant is -.59, Mask is .432, and Mobility is -.364. We show, by county, the value of POLAR across the United States in Figure 3. As shown, the most polar against COVID policies are in North Dakota, while California and Massachusetts show polar scores that favor COVID policies. In states like Florida, we see a mix between polar on both sides of the spectrum. Since political polarity is defined as congruence of polar opinions across inter-related issues, we use this PCA reduced variable as our proxy for political polarization related to COVID as congruence across the three dimensions results in either more or less positive POLAR values, depending on if the region is for or against those policies. While counties in the middle of the spectrum either show weak affinity to each issue or a mix of being for some policies and against others. We use this variation across treated and control counties to establish if polarity issues rather than political affiliation plays a role in both vaccination adoption and the strength of the incentive's effect.

Table IA.1: Correlation between Policy Variables. This table contains correlation between the variables used to create our measure of Political Polarization. We focus on three dimensions of COVID policy: vaccinations, masking, and social distancing. The table shows significant and moderate correlations between the variables of interest. We include the Vote variable to show there is some, but not perfect, correlation between the voting behavior of people in a county and adherence to these chosen policy variables.

	Sev. Hesitant	Hesitant	Mask	Mobility	Vote
Sev. Hesitant	1.000	0.971	-0.337	0.265	0.250
Hesitant	0.971	1.000	-0.390	0.284	0.273
Mask	-0.337	-0.390	1.000	-0.400	-0.529
Mobility	0.265	0.284	-0.400	1.000	0.423
Vote	0.250	0.273	-0.529	0.423	1.000

SIMULATION RESULTS OF ESTIMATION STRATEGY

In this section, we explore two simulation exercises to highlight the usefulness of our estimation approach. First, we show a simulation study that contrasts estimates from a pooled estimation strategy versus our counterfactual estimation strategy. We show that the pooled estimation strategy yields a biased estimate, while our approach shows very little bias. Second, we show a simulation study that highlights the need for separate estimation rather than estimation at the county-level.

Validation of Estimation Approach

To validate our estimation approach, we provide the following simulation study. In this study, our goal is to recover the estimate of treatment $\beta = -1$. We consider a sample of 1,000 markets and simulate the pooled estimate of our object of interest across 100 simulations.

We start with the utility expression shown in Cosguner and Seetharaman (2022):

$$U_{m,t} = \alpha_{m,t} + P_{m,t}\beta + \xi_{m,t} + \epsilon_{m,t}, \tag{1}$$

, where $\alpha_{m,t}$ is defined as:

$$\alpha_{m,t} = ln\left[ln[\frac{1-F_{m,t-1}}{1-F_{m,t}}]\right]. \tag{2}$$

$\alpha_{m,t}$ is the time-varying intercept for market m, which is defined by time t and Bass parameters p_m and q_m as shown in:

$$F_{m,t} = \frac{1-e^{-(p_m+q_m)t}}{1+\frac{q_m}{p_m}e^{-(p_m+q_m)t}}. \tag{3}$$

$P_{m,t}$ is a dummy variable representing that a promotion is occurring in market m at time t, and β is then the coefficient of our variable of interest. $\xi_{m,t}$ is a normally distributed, IID shock with mean 0 and varying standard deviations. This variable is present to create simulation error, so that our simulation study is not perfectly deterministic. Since the utility-based, Bass Diffusion model has dynamics involved, draws from $\xi_{m,t}$ then propagate forward throughout the estimation. Last, $\epsilon_{m,t}$ is the structural error, assumed to be mean 0 and standard deviation 1, with a distribution of Type I extreme value. Under this assumption of the error structure, then we can estimate the probability of the market choosing to engage based upon a logit formulation, as shown in Cosguner and Seetharaman (2022).

For our simulation study, we consider the possibility of three market types. A market can be described by two variables $X1$ and $X2$. Both $X1$ and $X2$ are uniformly distributed, where $X1$ directly dictates the market type, described by Bass parameters. For market type 1, $X1$ is distributed between 0 and 4.99, with $p = .05$ and $q = .2$. For market type 2, $X1$ is distributed between 5 and 7.99, with $p = .025$ and $q = .3$. For market type 3, $X1$ is distributed between 8 and 10, with $p = .015$ and $q = .4$. $X2$ is uncorrelated with market type and uniformly distributed from 0 to 10.

For our simulation study, we assume that 70% of the markets are type 1, 20% of the markets are type 2, and 10% of the markets are type 3. If a market is designated as type 2 or type 3, then there is a 40% chance that the indicated market received a promotion. If a market receives a promotion, then it occurs on the 9th week of the sample and runs for 6 weeks. Under these set of assumptions, we have selection on observables $X1$ with common support for both treated and untreated, which mirrors the assumptions in our paper.

For each of the 1,000 markets, using the specified utility and simulation draws, we then forward forecast the outcome variable Y for each period across 24 periods. For illustrative purposes, we consider two estimation strategies. The first strategy assumes a pooled estimation strategy so that there is one common estimate of p, q, M, and β within the sample. This strategy would not account for selected targeting of promotions, nor would it account for local level heterogeneity. Though, we do modify this approach, with the inclusion of market level controls to see if the pooled approach is able to accurately estimate the effect of treatment.

The second strategy is our chosen approach, outlined in the paper. Briefly, we estimate p, q, and M for each market M that did not receive a promotion within the sample window. Based upon these estimates, we train two different random forests where the estimated p and q serve as the target (dependent) variables and then $X1$ and $X2$ are feature variables that explain variation in the target variables. Once each random forest is estimated, we then interpolate the p and q for each market in our treatment condition and then estimate only β, the coefficient on our variable of interest, market by market.

We start our exercise by assuming that ξ has a standard deviation of .2. Given the small values within the utility function and that variation in prior periods propagates to future periods, this variable provides sufficient variation to our simulation study. Assuming a pooled estimation strategy, the estimate of β across the 100 simulations is -0.825, showing a bias of 17.5%. Though directionally consistent with the known estimate of β, this bias occurs due to having one singular estimate of p and q across the sample, rather than heterogeneous effects at the market level. We extend this test by incorporating $X1$ and $X2$ as variables within the indirect utility equation. With these additional parameters, we see a decline in performance of the estimates, resulting in an estimated effect of β as -0.417 with an overall bias of 58.3%. The inclusion of these market-level control variables in the utility specification is unable to properly reconcile the heterogeneity in diffusion pattern among the simulated markets.

Unlike the prior methods, our approach is able to properly estimate both the market level heterogeneity sufficiently and, thus, the effect of treatment. Recall that our approach estimates the Bass parameters (p, q, and M) market by market for any market that did not receive treatment. Then, we impute the Bass parameters for those markets that did receive treatment and estimate the effect of treatment holding the imputed Bass parameters fixed. Using our estimation strategy, the estimate of β is -1.009, showing a bias of .9%. Our estimation strategy is able to accurately estimate the effect of treatment in this simulation study. To further test the robustness of our approach, we increase the variation of ξ standard deviation up to .5. At a standard deviation of .3, the estimate is -1.017, with a bias of 1.7%.

At a standard deviation of .4, the estimate is -1.034, with a bias of 3.4%. At a standard deviation of .5, the estimate is -1.063, with a bias of 6.3%. As expected, as the amount of variation in ξ increases, we see more bias. However, the directionality and magnitude of the estimates are consistent across these further simulations and the bias remains 70% smaller than our estimated pooled model at the smallest standard deviation.

Observation Level Estimation

While our prior simulation shows the pitfalls associated with a pooled estimation and the benefits of our estimation strategy, a secondary strategy could be estimating the model at the county level. In this simulation study, we perform a variety of simulations to see if it is possible to estimate the effects of treatment using a localized model, rather than our imputation approach. For the simulation study, we proceed in the following steps.

We consider simulations across both p and q. For each combination of these Bass parameters, we forward simulate outcomes across 35 periods. We use the aforementioned utility specification from our prior simulation and fix the distribution of ξ to be normally distributed with mean 0 with a standard deviation of .2.

For the purposes of this simulation and to test the robustness of our approach, we hold fixed the value of treatment at -.5. Further, we consider the treatment being active for 2, 4, 6, or 8 periods. We also consider the timing of the treatment, where it can occur anywhere from period 1 to period 25. For each combination of parameters, incentive timing window, and incentive start period, we simulate 500 times. We then, for each combination of p and q compute the percentage difference between the estimated effect of treatment and the true treatment parameter. We consider both jointly estimating the full set of Bass parameters, versus only estimating the market size and effect of the incentive. Table IA.2 presents the results of this simulation. As shown, joint estimation yields much higher bias than the imputation approach.

Table IA.2: Bias Exercise - Comparison of Bias from a Fully Estimated Model versus a Partially Estimated Model with time-invariant shocks. This table presents the summary results from a series of simulated exercises meant to approximate the bias from estimating a full Bass model versus an imputed Bass model when there are time-invariant shocks, such as an incentive. We hold the effect of incentive fixed at -.5, since that is the average seen in our data. We construct incentive windows of 2, 4, 6, and 8 periods. For each incentive window, we start at period 1 and then sweep across 25 periods, where the total Bass model has 35 periods in total. Given the incentive window timing, we simulate a Bass diffusion process with p and q values shown. Each cell in the table represents the simple average of bias (estimated versus true, divided by true) of all periods across all incentive spells for a given set of Bass parameters. In this case, we look at the absolute bias across bootstraps, to ensure we prevent canceling, to see where bias occurs the most. We compare the bias from truth in the situation where the Bass model estimates all 4 parameters (p, q, M, and β) jointly versus one where the values from p and q are assumed to be imputed at their true values, thus, only estimating two free parameters (M and β). As shown, the imputation approach yields lower bias in nearly all cases, where the average bias is approximately 3.1%. For reference, the average p and q seen in our data is approximately .01 and .4, which would result in a bias of 10.9% for the full model. Though the point estimates show bias, the Full Model and Imputed Model are always consistent directionally. Bias increases for the Imputed Model in situations not seen in the data (higher values of p and high values of q), shown in the bottom right of the panel.

	Full Model					Imputed Model				
			q					q		
p	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
0.005	5.480	2.792	0.246	0.716	0.403	0.010	0.016	0.019	0.019	0.037
0.01	0.379	0.063	0.232	0.109	0.320	0.012	0.015	0.020	0.018	0.055
0.025	0.011	0.013	0.024	0.086	0.175	0.014	0.015	0.021	0.040	0.078
0.05	0.012	0.019	0.041	0.201	0.211	0.015	0.017	0.023	0.075	0.108

BASS DIFFUSION MODEL - ASSUMPTIONS AND ESTIMATED PARAMETERS

Anticipation Assumption Testing

Our modeling requires that individuals did not change their behavior before the incentives are offered. We argue that, due to the quick announcement of the incentives, there was a lack of time for individuals to strategically change their behavior. Here, we formally test this assumption. We compare the vaccination totals in the two weeks preceding the first incentive roll-out in each county in a treated state on the full sample (t-1,t-2) with those in the two weeks before that period (t-3, t-4). We choose two-week samples to be parsimonious with our estimation by smoothing out the reporting measure. We then compute the difference between these totals and perform a paired-sample t-test across states offering incentives. The test statistic .574, with a mean value of 80.36 and a standard error of 139.817. This test shows there was no strategic change in behavior before the incentive was offered. For robustness, we alternatively compare the total vaccinations of the preceding week against the average vaccination rates of the prior three weeks at the county level. The test statistic is -1.49, which again shows no significant anticipation effect. Since there is no statistical difference in the vaccinations immediately before the incentive and the preceding period, we find little evidence that individuals altered their behavior before the incentive was implemented.

Bass Diffusion Model Estimated Parameters

In this section, we explore the effects of our moderators and control variables on the latent, bass diffusion parameters. We start with a linear projection of the county-level estimates onto these aforementioned variables to understand how these variables may moderate the latent utility. Then, we explore the feature importance of the same variables, since our imputation strategy relies on the use of an optimally-tuned, random forest to impute the counterfactual bass diffusion parameters.

We start our discussion by examining the role of our county-level variables in moderating the four primitives of the Bass Diffusion model on vaccine adoption: the parameter of innovation, the parameter of imitation, the effect of COVID-19 deaths, and the overall market size. Table IA.3 presents the results of this analysis, where the dependent variable of each column is noted at the top of the columns. Since our goal is to understand how political polarization may moderate vaccine adoption rates, we discretize our PCA constructed variable into two values. Polar (Against) is the bottom quartile of the Polar variable's distribution. Polar (For) is the top quartile of the Polar variable's distribution. In addition, we omit our variables of vaccine hesitancy, mobility, and masking, as the Polar variable is representative of the underlying latent factor of these variables.

First, we note that the Polar variable has different effects contingent upon the parameter evaluated. Polar (Against) increases the parameter of innovation ($\beta = .003$, p-value $< .05$) and decreases the parameter of imitation ($\beta = -.061$, p-value $< .001$), both effects being statistically significant. Compared to the median values of both for the sample of analog counties, $p = .0106$ and $q = .409$ respectively, being strongly against COVID-19 policies results in a 28% higher parameter of innovation and 15% lower parameter of imitation compared to the baseline. This would result in an acceleration in the initial period of diffusion, but a longer process overall. Acknowledging this pattern change is important, as the timing of when to offer an incentive is key to enhancing its success.

We see a further effect of Polar (Against) in increasing the effect of β_{Deaths}, with a positive and statistically significant coefficient ($\beta = .284$, p-value $< .01$). This finding signals that, when presented with factual evidence of the benefits of a prosocial outcome (i.e., deaths due to not being vaccinated), even those entrenched on the underlying policies are now more inclined to act than moderates on the policies.

Last, Polar (For) directly impacts the total number of people being vaccinated ($\beta = 20328.124$, p-value $< .001$). In regions where they are for COVID-19 policies, approximately 20,000 more people choose to get vaccinated, while Polar (Against) shows no statistical difference between COVID moderates. The combination of these findings suggest that, while those entrenched against similar policies are quicker with the adoption curve, the result is a lower-level overall. While the latter finding is intuitive, the former is surprising as one would posit that being for inter-related policies would, most likely, accelerate the adoption pattern of the prosocial behavior, we find the opposite.

Examining the roles of Risk and Income, we find a similar pattern to that of Polar (Against), though the effect is much more pronounced. In regions where there is greater risk of COVID-19 vaccination roll-out, we find a much quicker diffusion pattern. The quicker pattern can be thought of as there being less marginal people considering vaccination, which is why the parameter of innovation is suppressed ($\beta = -.426$, p-value $< .001$) and the parameter of imitation is enhanced ($\beta = .019$, p-value $< .01$). We find a similar quick pattern of diffusion in regions with higher income, with the effect on innovation being ($\beta = .013$, p-value $< .01$) and imitation being ($\beta = -.129$, p-value $< .01$).

One last finding to highlight is that the effect of political affinity is different from that of political polarization on the issues. In our model, we show that political affinity based on party (Vote) has a positive effect on both innovation ($\beta = .008$, p-value $< .01$) and imitation ($\beta = .060$, p-value $< .01$). While Vote shares the same directionality on innovation, it is in the opposite direction on imitation as Polar (Against). This differential highlights the importance of how polarization is classified, as classifying purely on party affiliation may lead to a biased interpretation of how the market will react compared to polarization on the underlying issues related to the prosocial activity. We continue with this theme throughout our analysis.

It is worth briefly highlighting that some of our other control variables do have an effect on the Bass Diffusion parameters. For example, SVI, how vulnerable a community is to disasters, moderates the effect of imitation ($\beta = -.066$, p-value $< .1$) and the final market size of vaccination adoption ($\beta = -13570.517$, p-value $< .1$). Unemployment ($\beta = .001$, p-value $< .001$), female percentage ($\beta = .050$, p-value $< .1$), and minority percentages

($\beta = -.011$, p-value $< .01$) affect the parameter of innovation, while only minority percentage ($\beta = .070$, p-value $< .05$) affects the parameter of imitation. The location of the county (rural versus metro) influences the entire bass diffusion chain. Last, population ($\beta = 3595.144$, p-value $< .001$) and the number of medical facilities directly impacts the total number of people vaccinated ($\beta = 34.590$, p-value $< .001$). These findings support the use of these control variables to help inform the parameters related to intrinsic (innovation) and extrinsic (imitation) populations within the market.

Our findings from the linear projection resonate with those in the feature importance shown in Table IA.4.

Table IA.3: Dynamic Parameter Estimates from Analog (Control) Counties This table presents the results of regressing the estimated dynamic parameters for all 1037 analog (control) counties on our moderators of interest and control variables. To account for political polarization, we create dummy variables for the counties based on them being in the bottom quartile of the Polar metric (Polar - Against) or the top quartile of the Polar metric (Polar - For) across all counties in our sample. As shown, Polar (Against) impacts the underlying diffusion parameters and Polar (For) impacts the final Market Size, thus showing differential effects. Also of note is that the effect of political identification (Vote) does not have consistent effects with our Polar measure, which indicates that polarization on issues and political affiliation need not have the same outcomes. Last, we note our other moderators of interest, Risk and Income, also have effects on the parameters of innovation and imitation. For Income, we take the natural log of the value.

		$\hat{p}$ (Parameter of Innovation)			$\hat{q}$ (Parameter of Imitation)			$\hat{\beta}_{deaths}$			$\hat{M}$ (Market Size)		
		EST	SE	Sig.	EST	SE	Sig.	EST	SE	Sig.	EST	SE	Sig.
	Intercept	-0.057	0.019	**	0.806	0.153	***	-1.172	1.151		3000.551	33552.458	
Moderators	Polar (Against)	0.003	0.001	*	-0.061	0.012	***	0.284	0.090	**	530.197	2617.291	
	Polar (For)	-0.002	0.002		0.005	0.020		-0.076	0.149		20328.124	4336.950	***
	Risk	0.019	0.004	***	-0.426	0.030	***	0.233	0.224		12280.102	6535.086	.
	Inc	0.013	0.005	**	-0.129	0.041	**	0.174	0.307		211.420	8941.675	
Control Variables	Vote	0.008	0.003	**	0.060	0.022	**	-0.212	0.163		843.947	4740.891	
	Pop	-5.44E-05	7.62E-05		0.001	0.001		-0.004	0.005		3595.144	136.462	***
	Medical	1.13E-06	2.02E-06		-3.06E-07	1.65E-05		8.25E-05	1.24E-04		34.590	3.622	***
	Unemploy	0.001	3.33E-04	***	4.62E-04	0.003		0.017	0.020		-614.475	596.025	
	Minor	-0.011	0.004	**	0.070	0.031	*	0.256	0.230		7858.901	6690.181	
	Female	0.050	0.027	.	-0.106	0.224		1.285	1.686		-31126.077	49140.687	
	SVI	0.007	0.004		-0.066	0.035	.	0.311	0.260		-13570.517	7584.763	.
	M Code	0.001	0.001	*	0.014	0.004	**	-0.096	0.034	**	2220.758	984.696	*
	Num Obs	1037			1037			1037			1037		
	Adj. R-Squared	0.150			0.464			0.049			0.925		

.p <.10; *p <.05; **p <.01; ***p <.001.

Table IA.4: Variable Importance Measures from Random Forests used to Impute Bass Diffusion Coefficients. This table presents the feature importance measures of each variable used to generate the Random Forest functions that relate county characteristics to Bass Diffusion parameters. For each forest, we optimally tune all random forest variables for each forest of 10,000 trees. Unlike our analysis in the paper, we use all features here, rather than excluding those associated with the Polarity variable because it is a combination of COVID policy related variables. As shown, there are strong associations between the most important features and those that are significant in the linear projections of features on the estimated Bass diffusion coefficients from the analog (control) counties.

Variable	$\hat{p}$	$\hat{q}$	$\hat{\beta}_{deaths}$	$\hat{M}$
Sev. Hesitant	0.136	0.130	0.074	0.001
Hesitant	0.031	0.065	0.083	0.002
Mask	0.018	0.018	0.071	0.006
Mobility	0.010	0.013	0.055	0.007
Polarity	0.028	0.027	0.065	0.014
Risk	0.544	0.484	0.074	0.003
Income	0.057	0.053	0.060	0.002
Vote	0.037	0.008	0.060	0.002
Pop	0.007	0.020	0.081	0.796
Medical	0.010	0.012	0.059	0.152
Unemploy	0.031	0.015	0.076	0.005
Minor	0.012	0.006	0.058	0.005
Female	0.011	0.005	0.059	0.001
SVI	0.059	0.141	0.071	0.001
M Code	0.009	0.005	0.055	0.004

ALTERNATIVE ESTIMATION SPECIFICATION AND RESULTS

To test the robustness of our findings, we modify our framework to use matching, rather than imputation. In this section, we detail the results from our matching. We start by highlighting the results from matching. Then, we explore the results of our initial heterogeneity exercises by comparing the estimates from our imputation approach against those from the matching approach.

Estimation Through Matching and Covariate Balance

To estimate the effects of incentives on vaccination adoption, we choose to use an imputation approach. However, as noted in our main paper, we test the robustness of our findings using other samples of the data, as well as a matching approach. To complete the matching approach, we proceed in the following steps.

First, since we have more treatment observations than control observations, we employ matching with replacement. To find the appropriate matched sample, we use the Mahalanobis distance measure, where we include the full set of moderators and control variables used in our study. We consider a one-to-one matching to find the closest county from the control pool to each observation in the treatment pool. Once a county from the control pool is found, we use its estimated Bass parameters, as explained in our main text, for the counterfactual diffusion pattern to estimate the effect of incentives (treatment) at the county and incentive level.

Table IA.5 shows the before and after means and standardized mean difference (SMD) for each variable within our sample. As shown, matching provides much better balance to the sample. Using the commonly agreed threshold of .25 as a measure of good balance between treatment and controls, we find that 4 covariates violate balance before matching and there are no unbalanced covariates post matching.

Table IA.5: Covariate Balance - Before and After Nearest Neighbor Matching This table presents the covariate balance of treatment and control counties before and after nearest neighbor matching. We employ the Mahalanobis distance measure, with one to one matching and replacement, to generate a matched sample of control counties. As shown, the matching is a success as the standardized mean difference (SMD) is below .25 across all covariates in the matched sample, which is a determinant of strong balance. Before matching, the samples were unbalanced on several dimensions, specifically on the variables related to COVID policy polarity and unemployment.

	Treated	Control - Before Matching		Control - After Matching	
Variable	Mean	Mean	SMD	Mean	SMD
Sev. Hesitant	0.085	0.096	-0.303	0.090	-0.153
Hesitant	0.130	0.148	-0.366	0.138	-0.167
Mask	0.533	0.469	0.424	0.512	0.135
Mobility	-24.542	-23.205	-0.199	-23.577	-0.144
Polarity	0.076	-0.608	0.406	-0.240	0.187
Risk	0.485	0.496	-0.044	0.505	-0.084
Income	5.580	5.505	0.049	5.532	0.031
Vote	0.247	0.229	0.058	0.267	-0.067
Pop	13.570	9.079	0.105	11.510	0.048
Medical	922.061	307.560	0.082	416.535	0.067
Unemploy	7.312	6.288	0.504	6.919	0.193
Minor	0.177	0.178	-0.006	0.163	0.078
Female	0.501	0.500	0.070	0.502	-0.062
SVI	0.526	0.492	0.130	0.511	0.058
M Code	4.421	4.674	-0.165	4.467	-0.030

Comparison of Estimates

Using a matched sample, we find the pooled effect of promotions to be $-.428$ with a standard error of .028. By comparison, using our approach, the pooled estimate is $-.518$ with a standard error of .021. Thus, we have consistent directionality and very similar magnitude of effects between methods. We do see that our imputation approach does yield a lower standard error compared to the matched sample. This finding is consistent throughout the exploration of heterogeneous effects, where the imputed value provides a more precise estimate of the underlying diffusion pattern, thus, more accurate estimates of treatment compared to one-to-one matching.

Table IA.6 replicates the results from the main paper, where we project the heterogeneous effects of the incentives onto characteristics of the incentives. The left two columns are those shown in the main paper, and the right two columns are those obtained using the estimates via matching. We note that there are comparable estimates across all variables, in terms of directionality and magnitude, save for the number of concurrent promotions. The major difference between the two is in terms of statistical significance. The standard errors are larger and the adjusted r-squared is smaller for the matching estimate model, suggesting there is much more variation in estimates using this approach.

Table IA.7 replicates the results from the main paper, where we project the heterogeneous effects of the incentives onto county-level characteristics. Again, the left two columns are those shown in the main paper and the right two columns are those obtained using estimates via matching. Most estimates show similar directionality and magnitude between the estimation types. The most notable differences are Polar (For), SVI, and Female. The major shift, as mentioned earlier, is in terms of statistical significance. The standard errors are larger and the adjusted r-squared is smaller when using the one-to-one matching estimates versus the imputation method.

Table IA.6: Comparison of Heterogeneous Effects of Incentive by Incentive Type - Imputation vs. Matching This table presents comparison of the Imputation Results (our main paper's results) versus the results from matching (Matching Results). As shown, regardless of method, the directionality remains the same between methods with similar point estimates. The main difference comes from significance, as the matching method shows larger variance in effects (noted by the lower adjusted r-squared). This finding shows the benefit of using the imputed coefficients from random forest, as those estimates are learned from non-parametric matching and aggregation for more stable and informative estimates.

	Imputation Estimates			Matching Estimates		
	EST	SE	Sig.	EST	SE	Sig.
Intercept	-0.266	0.065	***	-0.302	0.089	***
Lottery	-0.931	0.071	***	-1.031	0.097	***
Nonmonetary	-0.245	0.063	***	-0.268	0.087	**
Scholarship	-0.199	0.080	*	-0.254	0.110	*
Duration	0.022	0.006	***	0.018	0.009	*
Cash	0.054	0.006	***	0.055	0.008	***
Num_Promo	-0.130	0.068	.	0.025	0.093	
Adj R squared		0.097			0.056	

.p <.10; *p <.05; **p <.01; ***p <.001.

Table IA.7: Comparison of Heterogeneous Effects of Incentives by County Characteristics - Imputation vs. Matching This table presents comparison of the Imputation Results (our main paper's results) versus the results from matching (Matching Results). As shown, regardless of method, the directionality remains the same between methods with similar point estimates. The main difference comes from significance, as the matching method shows larger variance in effects (noted by the lower adjusted r-squared). This finding shows the benefit of using the imputed coefficients from random forest, as those estimates are learned from non-parametric matching and aggregation for more stable and informative estimates.

	Imputation Estimates			Matching Estimates		
	EST	SE	Sig	EST	SE	Sig
Intercept	-0.711	0.652		-0.885	0.887	
Polar (Against)	-0.183	0.056	**	-0.252	0.077	**
Polar (For)	0.130	0.058	*	0.010	0.078	
Risk	0.249	0.118	*	0.240	0.162	
Income	0.405	0.127	**	0.330	0.174	.
Vote	-0.414	0.088	***	-0.468	0.119	***
Minor	-0.383	0.127	**	-0.258	0.173	
SVI	-0.016	0.126		-0.285	0.173	.
Female	-1.706	1.129		-0.561	1.535	
Unemploy	0.057	0.012	***	0.059	0.016	***
Adj R Squared		0.085			0.049	

.p <.10; *p <.05; **p <.01; ***p <.001.

HETEROGENEITY IN RESPONSE TO INCENTIVES COUNTERFACTUAL

We explore our counterfactuals on additional time by focusing on a smaller subset of incentives and exploring the shift in county-level characteristics based upon the promotion.

Table IA.8 shows a more granular level of the prior exercise by focusing only on a small selection of incentives. Further, rather than examining the aggregate response to the incentives, we focus on the county-level characteristics of the counties that see an increase in vaccinations due to the incentive. We split the descriptive statistics of the counties in the sample into Yes for those that see an increase in the vaccinations and No for those that do not. Our selection of treatments are common ones seen in the data sample as a useful example.

We note the following five trends from the aforementioned counterfactual analysis. First, the counties that see benefits from most incentives are those with lower income, higher risk, and more polar to being against COVID-19 protocols. Effectively, this grouping of characteristics signals that incentives do indeed work in stimulating the targeted group (i.e., those who would benefit the most and who are of lower income). For example, examining the $1 million dollar lottery incentive, it has the lowest polar score for the Yes column among all policies (-1.56), so the most strongly, ideologically opposed actually choose to get vaccinated with this policy. From our exercise in the main paper, we see that only a few counties actually benefit from lottery incentives, but they share the common characteristics of being more ideologically opposed, having a higher risk profile, and being generally poorer.

Second, ideological extremism and party affiliation show diverging outcomes. The counties that benefit the least from the incentives are those with an affiliation to the Republican Party, not necessarily adhering to being strongly opposed to the underlying policies. Thus, when considering enacting public policy initiatives to drive prosocial behavior, believing that party affiliation is a true proxy for political polarization on the underlying issues leads to a mismatch in outcomes and may, in fact, lead to suboptimal decisions. For example, a direct monetary payment of $50 sees benefit for the more ideologically opposed (-.792 vs. -.095) for the Polar score between those that enhance vaccinations versus those that do not. If we look at the Vote scores, we see the opposite effect, with .028 for Yes and .312 for No. The lower score for Yes suggests more democratic counties are choosing to vaccinate do to the incentive, which is counter to being ideologically opposed to COVID policies.

Third, related to the prior point, entrenchment based on party affiliation is much more difficult to sway with incentives. Comparing the policy of giving a $50 pay out to a $100 pay out, we see more movement in balancing the polarity metric between the Yes and No groups (meaning that more weakly polar counties now benefit from the policy); yet, the vote metric barely moves between the two treatments for those not choosing to get vaccinated based on the offered incentive (.312 vs. .339). We argue that this finding shows political affiliation, in this hyper-partisan environment, it much more difficult to dislodge than meaningfully interacting with individuals based on their policy preferences.

Fourth, Nonmonetary promotions of $20 seems to be an effective way to target republican communities. Out of all our targeting exercises, this results in the highest Vote (.251) out of all the incentive plans. The more moderate risk communities and the higher income areas choose to get vaccinated when this incentive is offered. Thus, our analysis shows clear targeting opportunities for the deployment of these incentive types in a targeted way.

Fifth, when direct monetary payments become sufficiently high, we see that people who are for COVID-19 policies embrace this incentive, as opposed to the other polices show, as the value of Polarity for the Yes column is -.144 versus -.301 in the No. This finding shows that offering low valued COVID incentives caused a backfiring effect on the population that was willing to embrace COVID policies.

In summary, this exercise shows at a more granular level that effective targeting of policies is possible. By leveraging different incentive strategies combined with the incentive's value, there is a route to engage different groups to get vaccinated. This analysis, combined with the counterfactual analysis shown in the main paper highlight that, short of large, direct monetary payments or tangible nonmonetary incentives, singular policies are likely to fail to engage people sufficiently to adopt prosocial behaviors.

Table IA.8: **Counterfactual Analysis of County Characteristics Conditional on Gains from incentives.** This table presents the summary results of counterfactual analysis that examines the four different incentives, where we separate the counties on whether the vaccination rate increased due to the presence of the incentive. Our prior analysis suggests heterogeneous responses to the incentives, where most incentives see little benefit in both acceleration of vaccinations and overall vaccination rates. Here, we explore this more granular by examining the county characteristics for those counties that benefit (Yes) versus not (No) from the indicated incentive. We examine the values of our three chosen moderators (Polarity, Risk, and Income), as well as two additional variables (Vote and Unemployment Rate). Our findings show that, the counties that show positive benefit from incentives (i.e., a positive increase in vaccinations) are those with Polarity against COVID policies, higher risk, and lower income. More interestingly, these counties are also more moderate or democrat leaning than republican counties, where republican slanted counties seemingly never benefit from any policies. These findings suggest that there is a divide between our understanding of political polarity based on issues versus party affiliation.

	Monetary - $50		Monetary - $100		Lottery - $1 Mill		Nonmonetary - $0		Nonmonetary - $20	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Polarity	-0.792	-0.095	-0.144	-0.301	-1.560	-0.219	-1.055	0.100	-0.374	0.365
Risk	0.635	0.441	0.590	0.451	0.547	0.485	0.648	0.417	0.493	0.474
Income	5.128	5.670	5.471	5.548	5.253	5.555	5.054	5.764	5.558	5.015
Vote	0.028	0.312	-0.002	0.339	0.066	0.248	0.155	0.280	0.251	0.214
Unemploy	6.716	6.817	7.333	6.594	5.500	6.846	6.362	7.018	6.471	9.210