Doing Good rather than Doing Well: What Stimulates Personal Data Sharing and Why?

Maclean Gaulin[a], Nathan Seegert[a], Mu-Jeung Yang[a], Stephen Alder[a,b], Adam Looney[a,b]

[a]*University of Utah, 1655 Campus Center Drive, Salt Lake City, UT 84108*
[b]*Author names are ranked according to contribution, with first three authors contributing equally.*

Abstract

Personal data markets have become ubiquitous. The non-rivalry of data suggests that the social returns to personal data sharing will often exceed its private returns. Using a unique sequence of Randomized Controlled Trials (RCTs) for COVID-19 testing among tens of thousands of households in Utah, we analyze different tools to stimulate personal data sharing. We contrast the effectiveness of incentives for data sharing with mechanisms suggested by behavioral economics, including moral engagement, image motivation, and identity. Our results suggest that incentives can easily backfire and are highly complementary with framing effects. Furthermore, image motivation and identity are an order of magnitude more effective in influencing data sharing than monetary incentives.

Keywords: Personal Data Markets, Complementarity of Incentives and Social Preferences, Motivational Crowding, COVID-19, Randomized Testing
JEL MSC: I15, I18, J68

Email addresses: mac.gaulin@eccles.utah.edu (Maclean Gaulin), nathan.seegert@eccles.utah.edu (Nathan Seegert), mjyang@eccles.utah.edu (Mu-Jeung Yang), salder@utah.edu (Stephen Alder), looney@eccles.utah.edu (Adam Looney)

April 18, 2021

1. Introduction

Recent years have seen an unprecedented rise in the use of ‘‘big data” and technologies to analyze such data, including machine learning and large-scale A/B-Testing. Data and information are now widely considered a key economic resource or ‘‘fuel for the future” (*The Economist*[1]). In many respects, the ascendancy of data as an economic resource is synonymous with the rise of personal data markets. Technology companies, such as Google, Facebook, Amazon, and others, rely on gathering personal data to inform a variety of data-driven decisions, from demand-forecasting and pricing to product design. A key feature that differentiates data from other economic factors is that it is fundamentally non-rival: use of personal data by one consumer, does not preclude usage of the same data by others, including firms, researchers or the government. This non-rivalry makes personal data sharing similar in nature to the problem of contributing to public goods. Social returns of such contributions are typically far larger than private returns. An important recent example is participation in randomized COVID-19 testing. The private benefits of getting such testing are much smaller than the public health benefits of gathering large-scale data from such tests to track the prevalence of the infection.

This raises the question of how to stimulate personal data sharing. There are two alternative and potentially conflicting approaches to stimulate data sharing. An economic approach to personal data sharing emphasizes the importance of (monetary) incentives. Indeed, studies on mechanism design analyze how incentives can optimally be used to induce the sharing of private information, with pioneering contributions by Hurwicz (1960), Vickrey (1961), Clark (1971), Groves (1973), and Myerson (1981). On the other hand, behavioral economics argues that different types of socially-based motivations, such as moral engagement, social approval, and identity offer a range of powerful alternative mechanisms to incentives, see Ariely et al. (2009), Akerlof and Kranton (2010a), or Bowles and Polanía-Reyes (2012).

In this study, we investigate the quantitative effectiveness of incentives and social motivations

[1] See: economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy

as tools to induce personal data sharing using field Randomized Controlled Trials (RCTs) among tens of thousands of households in Utah during the COVID-19 pandemic. There are two key reasons our empirical setup can be particularly useful When examnining the drivers of personal data sharing decisions. First, media coverage of and public attention to the COVID-19 pandemic was particularly high during the data gathering period of the study, which was Spring and Summer 2020. As a result, both the risk of COVID-19, and the potential need for testing were likely to be salient and the stakes were likely perceived to be high. Second, we can measure decisions on different types of personal health data that can be shared, which varies by both costs and benefits. The personal health data we measure includes an online health survey, viral tests that indicate infection status, and antibody tests that provide information about past infections. The antibody test is the most costly and invasive, requiring a blood draw, followed by the viral test, which costs half of an antibody test and requires a swab of the nasal cavity, and finally the online survey.

Evaluating the effectiveness of incentives in stimulating personal data sharing as opposed to social preferences is a classic program evaluation problem. For example, a naive policymaker who offers an incentive for personal data sharing risks paying people who would have been willing to participate without such an incentive. We therefore use a sequence of RCTs to evaluate the effectiveness of incentives and social motivations. For our analysis of incentives, we randomly provide monetary incentives for personal data sharing, varying the incentive from \$10 per person in a household to \$30 per person. This variation implies strong differences in incentives. A four-person household could expect an hourly compensation between \$20 per hour at the \$10 incentive with a 2 hour testing time to \$360 per hour at the \$30 incentive with a 20 minute waiting time. By varying the incentive amounts, we are able to detect the non-linear effects of incentive strength on personal data sharing.

We also evaluate three types of social preferences: moral engagement, image motivation and identity motivation. We define moral engagement, as a change in the perception of the moral virtue of an action. To trigger moral engagement, we frame the participation in COVID-19 testing as ‘‘helping to overcome a crisis” and randomly assign this framing to households recruited via a mailed letter. An alternative type of social preferences we analyze is image motivation, defined as the tendency to be motivated by the perception and approval of others (see Benabou and Tirole (2006); Ariely et al. (2009)). Image motivation requires observability to be effective, which is why we use the random assignment of in-person canvassers instead of recruitment letters for this purpose. Finally, we assess motivation based on identity, defined as a person's sense of belonging to a specific social group (see Akerlof and Kranton (2010a) and Akerlof and Kranton (2010b)). To investigate identity motivation, we rely on natural variation from detailed voter registration files for the state of Utah because we are unable to credibly randomly assign such identities.

Three key sets of results are worth highlighting. First, the strength of incentives impacts personal data sharing non-monotonically. At small levels of incentives, personal data sharing is either unaffected or declines. As incentive levels increase, personal data sharing tends to increase as well. This is consistent with Gneezy and Rustichini (2000), who find similar non-monotonicities in laboratory experiments. However, we go beyond the findings of Gneezy and Rustichini (2000) by showing that the net effectiveness of incentives to stimulate personal data sharing strongly depends on moral engagement effects. If people are morally engaged, incentives can significantly stimulate personal data sharing, while moral disengagement can render incentives completely ineffective.

Second, social motivations are an order of magnitude more effective in stimulating personal data sharing than monetary incentives. For example, our image motivation treatment causes personal data sharing increases that are ten times larger than our most effective incentive treatment. At the same time, moral engagement stimulates personal data sharing to a degree similar to the largest monetary incentive, while exhibiting a marginal cost of zero.

Third, we document the importance of complementarities among incentives and social preferences, as well as across different types of social motivations. Among the most surprising results is that higher levels of incentives systematically depress personal data sharing by people with conservative identities. In contrast, both moral engagement and image motivation are strongly complement conservative identity. These suggests are consistent with the view that incentives can in certain contexts be perceived as coercive (Grant (2006)) and therefore provoke a contrarian response, leading to lost overall output Kranton and Sanders (2017).

Our findings have implications for personal data markets in the private and public sectors. In our setting, learning about an infectious pathogen in real time is critically important to governments. For example, the State of Oregon began a monumental study to track 100,000 Oregonians at a cost of $24 million. However, due to an inability to recruit a representative sample to participate in the study, the government cancelled it after spending more than a million dollars. In contrast, the State of Utah created the HERO group and they have successfully tested tens of thousands of individuals; it is the only study providing data in real time from a large-scale randomized sample (Samore et al., 2020). Said differently, the benefits to understanding and correctly setting up personal data markets are simply enormous.

Our work contributes to at least two related literatures. First, our work is related to the literature on public good contributions through non-monetary motivations. This includes empirical work on charitable giving, motivated by image motivation (Ariely et al., 2009), warm-glow altruism (Andreoni, 1989, 1990; List et al., 2019), and social pressure (DellaVigna et al., 2012; Andreoni et al., 2017). Charitable giving usually does not involve direct monetary incentives for giving, which is why we are able to analyze the interaction between incentives and social preferences. On the other hand, while some blood donation campaigns use monetary incentives, empirical results on the impact of incentives on blood donations are ambiguous. While studies such as Lacertera et al. (2014b) and other cited in Lacertera et al. (2014a) report a stimulating effect of incentives for blood donations, studies such as Mellstrom and Johannesson (2008) report the opposite result. Our context allows us to demonstrate that the effectiveness of incentives strongly depends on framing effects and the related moral (dis)engagement.

Although there is a sense in which our analysis is related to public good contributions, there is an important difference between participation in COVID-19 testing and purely pro-social contributions. While charitable contributions and blood donations primarily benefit others when not incentivized, people do receive a clear benefit in our context: test results indicating whether a person had COVID-19 in the past or is currently infected. In other words, we analyze a ‘‘personal data-for-benefit” exchange, which has also been recently analyzed by some recent studies, such as (Hui et al., 2007; Premazzi et al., 2010; Gabisch and Milne, 2014). Furthermore, the potential to appeal to moral engagement is similar in recruitment for medical research trials by pharmaceutical companies and other biomedical firms, such as for anchestry DNA tests, digital thermometers, etc. Our work is therefore related to recent work on public good provision in digital economic environments, such as Gallus (2016) and Burtch et al. (2018). These studies show the importance of non-monetary motivations and rewards in stimulating contributions to Wikipedia (Gallus, 2016) and online reviews (Burtch et al., 2018). To our knowledge, ours is the first study to explore the complementarity of incentives and social preferences for a "data-for-benefit exchange" in a biomedical context.

The remainder of the paper proceeds as follows. Section 2 provides a conceptual framework. In section 3 we provide details on the research design and data. In Sections 4 and 5, we report our main findings and the interactions of incentives and social preferences. We provide evidence on sample selection in Section 6. Finally, we provide a discussion in Section 7 and conclude in Section 8.

2. Conceptual Framework

To facilitate discussion of our empirical results, we begin by outlining a theoretical framework for how incentives and social preferences interact. We build on the discussion in Bowles and Polanía-Reyes (2012) and adjust it here for our purposes. Specifically, we consider the following utility function for a consumer deciding whether to share personal information:

$$U(p) = \beta_I \cdot p + \beta_m \cdot m \cdot p - c(p) + \left[\beta_{S,c} \cdot p + \beta_{S,m} \cdot m \cdot p\right] \cdot 1_{\{m>0\}} \tag{1}$$

where p captures the extent or probability that the consumer will share her personal data, with higher values of p corresponding to more personal data sharing.

Broadly, there are four different types of motivation for personal data sharing that this utility function formalizes. First, there is an intrinsic value of personal data sharing, captured by the coefficient β_I. In our context, this mostly likely includes the value a consumer assigns to the results of the test. But more broadly, this term could also include motivations, such as pure altruism, which are unaffected by monetary incentives. Second, consumers will be motivated by incentives, such as a monetary reward m for sharing their personal data. The strength of this effect will depend on the value of the monetary incentive m, as well as the marginal utility of money β_m. Third, personal data sharing will require consumers to incur effort costs $c(p)$. We think of this effort cost as capturing the opportunity costs of the consumer's time, as well as other potential non-monetary costs, such as effort spent in filling out a survey or driving to a testing location. For simplicity, we assume that this effort cost takes the form of a second order polynomial: $c(p) = c_e \cdot p + \frac{1}{2}p^2$. Fourth, we capture social preferences in the term $\left[\beta_{S,c} \cdot p + \beta_{S,m} \cdot m \cdot p\right]$. Whether and how social preferences interact with monetary incentives will depend on the values and signs of two coefficients, $\beta_{S,c} and \beta_{S,m}$. In this context, we distinguish two types of interaction effects between incentives and social preferences. On the one hand, the mere presence of a monetary incentive might change the extent of personal data sharing as captured by the coefficient $\beta_{S,c}$. Following Bowles and Polanía-Reyes (2012), we call this a "categorical" substitute or complement effect of social preferences and incentives. On the other hand, the social preferences might change the marginal utility of the incentive, captured by $\beta_{S,m}$. We refer to this as the marginal complementarity or substitutability of incentives and social preferences.

To understand the potential consequences of complementarity or substitutability of incentives and social preferences, we take the first-order condition for (1) and solve for the optimal personal data sharing decision:

$$p = \beta_I + \beta_S + \beta_m \cdot m - c_e + \left[\beta_{S,c} + \beta_{S,m} \cdot m\right] \cdot 1_{\{m>0\}}. \tag{2}$$

Equation (2) highlights the connection of monetary incentives and the personal data sharing, as well as the interaction of incentives with social preferences. Figure 1 shows how the categorical and marginal interactions of incentives and social preferences impact the optimal personal data sharing decision. While categorical complementarity or substitability will shift the level of the optimal data sharing curve, marginal complementarity or substitutabilty will affect the slope. Categorical substitutability has been shown to be important in understanding the non-monotonic effects of incentives on behavior (Gneezy and Rustichini (2000)).

3. Research Design and Data

The following sections give an overview of data collection and econometric specifications.

3.1. Background on the HERO Project

The Utah Health and Economic Recovery Outreach (HERO) Project was established to estimate the actual rate of community-based SARS-CoV-2 (the cornonavirus that causes COVID-19) infection and to help guide decision making about public health and Utah's economy. The HERO Project was a collaboration between the University of Utah David Eccles School of Business and University of Utah Health and is supported by the Governor's Office of Management and Budget.

To track the prevalence of COVID-19, thousandts of households were recruited per letter or per canvasser to participate in COVID-19 testing and an online health survey. Phase 1 part A of the HERO Project was aimed at measuring the proportion of people who have SARS-CoV-2 antibodies in Davis, Salt Lake, Summit, and Utah counties and understanding the factors associated with having SARS-CoV-2. Phase 1 part B extended this same work beyond these four counties to assess communities that may have high viral activity and monitor changes in antibody prevalence over time.

Phase 1 part A of the HERO Project has 10,996 participants from 5,130 households and has the results of 9,351 serology tests for SARS-CoV-2 antibodies. Phase 1 part B of the HERO Project has 8,590 participants from 4,180 households with results of 6,791 serology tests. For more details on the testing and recruitment see Samore et al. (2020).

3.2. Personal Health Data

Our main outcome variables of interest capture the willingness of potential respondents to share different types of personal health data, such as symptoms, social distancing behavior, and the results

Table 1: Positives by Activity

This table reports respondents from the survey's answers to what symptoms do you have. We scale the answers to report them as number of people per 100,000 people.

Symptoms	Symptom per 100,000 people
Fever	966
Cough	2,267
Shortness of breath	1,453
Chills	1,093
Repeated shaking	229
Muscle pain	3,793
Headache	8,474
Sore throat	4,728
New loss of taste or smell	534
None of the above	81,459

of two different types of COVID-19 diagnostic tests. This personal health data can be understood to differ in both the benefits and costs for potential participants. In turn, these cost and benefit differences are potentially informative, since respondents were free to choose which type of data they wanted to share and which type of data to withhold.

The personal health data with the lowest participant costs and also the lowest benefits is information we gather through an online health survey. The goal of this survey was to gather information on potential COVID-19 symptoms, as well as potential exposures to the virus and social distancing behavior. Filling out this online survey took about 20 minutes on average, which is a moderate time cost for respondents. On the other hand, completing the survey for all household members tended to increase the necessary time cost, especially for large households. Note that filling out the health survey by itself did not generate any tangible private benefit for respondents, since people had to come to a testing sight and get tested to receive any compensation.

Data from the online survey provide a foundation for understanding the spread of COVID-19. Table 1 reports the number of respondents reporting different symptoms per 100,000 people. Not surprisingly, many people reported headaches and muscle pain. Somewhat interestingly, 534 people per 100,000 reported a new loss of taste or smell. This symptom has been shown to be correlated with COVID-19 infection (Yang et al., 2020b).

After filling out the online health survey, participants were then assigned a nearby testing location, which either consisted of a bus that was parked in their neighborhood or a drive-through testing location. In other words, providing health data in terms of COVID-19 diagnostic tests was considerably more costly for potential respondents, due to the need to set aside a time to visit the testing location and due to the effort required to reach the testing location. Pictures of this recruitment are given in Figure 2. At the same time, conducting the tests offered two types of intrinsic benefits for participants. On the one hand, they would get the results of a PCR test, which indicates whether a person is currently infected with COVID-19. This type of test tends to have a smaller private benefit, since it can only diagnose whether a person currently carries the SARS-CoV-2 virus. Early on, a nasal-pharyngeal swab was used to for sample collection for the PCR test. This collection method tends to be uncomfortable as it involves a nasal swab, which needs to reach relatively deep into the nasal passages.

In contrast, the antibody (or CIA) tests offer higher private benefits at higher personal costs. The antibody tests require participants to provide veinous blood samples, which participants on average perceive as been more painful than nasal swabs. However, antibody tests also provide individuals more valuable information, such as whether they have antibodies against COVID-19 and were therefore exposed to the virus in the past. Since media reports emphasize the possibility of asymptomatic transmission of the virus, antibody tests can potentially provide the highest private benefit.[2]

The tests also provide public value. For example, Table 2 reports the ratio of people who report doing an activity and test positive for antibodies over the number of people who report doing an activity, per 100,000 people. This table provides the risks of different activities in the population we examined. Perhaps unsurprisingly, going to a medical facility is relatively low risk, while going to a restaurant or work was relatively higher risk.[3]

In Phase 1 parts A and B of the HERO project, 19,586 people from 9,310 households took

[2]However, the PCR test was a better indicator of current infection, so likewise a better indicator of infectiousness.

[3]For a look at how people respond to data on COVID-19 and mask orders particularly, see Seegert et al. (2020).

Table 2: Positives by Activity

This table reports the positive rate per 100,000 people reporting they are more likely than others to do the following activities: go to the grocery store, restaurant, retail store, medical facility, go for a walk, go to work, hang out with family, or go to church. Specifically, we calculate this number as 100,000 × (num people say go to activity and infected)/(num people go to activity).

Activity	Positive per 100,000
Groceries	5.58
Restaurant	10.29
Retail	6.02
Medical	2.97
Walk	4.99
Work	10.18
Family	8.50
Church	4.21

a survey, 14,050 people from 6,131 households took a PCR test, and 16,142 people from 7,125 households took an antibody test.

3.3. Recruitment and Empirical Design

The main endogeneity issue is that individuals are likely to have a wide variety of unobserved motivations to participate in COVID-19 testing. The unobserved differences across individuals could for example include differences in beliefs of the risk of being infected with the SARS-CoV-2 virus, due to private information on their own travel history and contact with potentially symptomatic people. Other unobserved factors include differences in the valuation of one's own life and the private opportunity costs of time. To address the presence of unobserved motivational differences and potential sample selection in personal data sharing, we conduct randomized evaluations of different factors. These program evaluations take the following form:

$$y_{vikst} = \alpha_k + \alpha_{st} + \beta_1 \cdot T_{1,i} + \beta_2 \cdot T_{2,i} + \beta_x \cdot T_{1,i} \times T_{2,i} + \varepsilon_{vikst}, \tag{3}$$

where y_{vikst} captures personal data sharing decisions, such as participating in the online health survey, taking a viral test or providing a blood sample for an antibody test. At the same time, we use randomized treatments $T_{k,i}$ to ensure that our results are not driven by unobserved individual differences, including potential interactions across treatments. These interactions are crucial for identification of the complementarity or substitability of monetary incentives and social preferences in personal data sharing decisions. Note that k denotes a stratum, t denotes sampling day, v denotes a sampling location, and i denotes a household. We include stratum fixed effects, as well as sample day dummies interacted with sample location. The variables $T_{k,i}$ are defined based on the randomization, irrespective of whether for example recruitment letters were returned or not. These estimates, therefore, represent intent-to-treat effects.

3.4. Incentive, Effort, and Information Treatments

Our empirical design comprises six different types of variation, four of which are randomly assigned treatments, while two are exploiting natural variation in the data. We start by discussing exogenous variation in factors that standard economic theory would predict to impact personal data sharing. The sample group consists of household, which we contact per letter or canvasser to participate in COVID-19 testing. All household members are invited to participate in testing and we record demographic information on all participants within the household.

The most obvious of the treatments are monetary incentives for participation in COVID-19 testing. For most of our treatments, these incentives are announced using letters that are sent to households' physical addresses. Incentive treatments vary in their strength and are disbursed to random households in increments of \$10, from a value of \$0 to \$30 per person. These incentives were only paid if subjects agreed to testing, which could take between 20 minutes amd 2 hours, depending on demand at a specific testing location. A household of four people could therefore expect an incentive payment all the way from \$20 per hour (with a \$10 incentive and 2 hour testing time) to \$360 per hour (with a \$30 incentive and a 20 minute testing time). This potential range of incentive effects therefore is sufficiently large to rule out ”small stakes” problems in inducing personal data sharing decisions. We also note that the use of different incentive levels allows us to detect potential non-monotonic effects of monetary incentives, such as the categorical motivational crowding out discussed in section 2, see Gneezy and Rustichini (2000).

A useful contrast to monetary incentives is private effort costs. While private effort costs are likely to directly impact personal data sharing decisions as captured in equation (2), they are less likely to interact with social preferences, the same way that monetary incentives do. Specifically, we exploit random variation in distance of households from the testing sites to proxy for the degree of private effort necessary to participate in data sharing.

Another potentially important factor that rational households are likely to consider when deciding whether to participate in COVID-19 testing is their beliefs about the prevalence of COVID-19 and therefore their risk of having been exposed to the virus. In this context, people might focus on two different pieces of information that are implicitly reported in case counts that are regularly reported by the media. On the one hand, higher case counts can be indicative of higher viral prevalence, which increases infection risk and therefore increases the benefit of testing β_I in equation (2). On the other hand, higher case counts could reflect more testing, which itself might suggest more likely quarantining of infected persons, thereby reducing infection risk and the benefit of testing. See also the discussion in Acemoglu et al. (2020) and Yang et al. (2020a).

We identify these two types of information, which change the intrinsic benefit of testing, by providing two different pieces of information in recruitment letters. To capture the effects on information on viral prevalence, we follow Yang et al. (2020b) and provide information on current positivity rates of COVID-19 testing in recruitment letters. Yang et al. (2020b) show that Bayesian updating can be used in combination with these positivity rates to provide sample-selection corrected estimates of the active prevalence of COVID-19. The recruitment letter highlights this information by providing the following passage:

> ‘‘Your participation is important for us to better understand and counteract COVID-19. Because of people like you, we have **found 35,000 positive cases (or 7% of those tested)**. Help Utah recover, come take our drive through test.”
> (Emphasis as in original letters.)

In contrast, an alternative information treatment provides information on the quantity of testing instead of the positivity rate or prevalence:

> ‘‘Your participation is important for us to better understand and counteract COVID-19.

> Because of people like you, we have **tested 470,000 people (or 15% of Utahns)**. Help Utah recover, come take our drive through test."
>
> (Emphasis as in original letters.)

3.5. Social Preference Treatments

We focus on identifying the three types of social preference effects discussed in the introduction. The most "light-touch" treatment to test for social preferences is the framing in the recruitment letters that were sent out to households. Our goal was to engage respondents morally and motivate them to participate based on this moral engagement. There were two treatment groups, which differed in text, we provided at the end of the first paragraph of their recruitment letter. For our moral engagement treatments, we therefore invoked the severity of a "health crisis" that COVID-19 triggered in the respondent's community with an appeal to help. The "health framing" referred to COVID-19 as a health crisis in the recruitment letter:

> ‘‘Covid-19 is a public health crisis. More than 1.3 million Americans have tested positive for COVID-19. In Utah, 553 residents have been hospitalized because of the disease. Your participation will help Utah stay healthy."

As an alternative framing, we also referred to COVID-19 as an economic crisis and appealed to sentiments expressed by several politicians that the economic consequences of COVID-19 are more important than its public health consequences[4]:

> Covid-19 is not just a public health crisis, but also an economic crisis. In April, the unemployment rate in the U.S. was 14.7 percent. This week 106,377 Utahans filed for unemployment benefits. Your participation will help Utah get back to work.

We expected these different frames not just to directly impact participation in COVID-19 testing, but to also interact with out monetary incentives. For example, a stimulative effect of the health

[4]See: https://www.nytimes.com/2020/03/23/us/politics/trump-coronavirus-restrictions.html

framing may be at least partially offset by the offer of monetary incentives if there is motivational crowding.

To induce image motivation we randomized canvassers instead of letters to recruit participants. To recruit canvassers, the David Eccles School of Business (where the authors work) cooperated with the University of Utah and the Utah Community Builders to hire college students to form the ”Hope Corps”. The mission of the Hope Corps was defined as ”assisting and lifting small businesses, nonprofits, and people of Utah” in response to the COVID-19 pandemic. It is therefore plausible that we were able to recruit particularly idealistic canvassers. Canvassing field teams were tasked to walk door-to-door in the neighborhoods selected for sampling to encourage participation.

The costs of these canvassers was about $25 per household, which allowed us to calculate the return on investment of this recruitment measure, compared to the monetary incentives. We also note that the average household size in our sample was two persons, so that the average cost of canvassers per person was $12.5. We also note that canvassers passed out survey forms for the health survey if households agreed to participate. On the other hand, canvassers were not involved in any kind of follow-up with households that stated they were interested in participating in COVID testing. This observation is helpful in confirming that the most likely impact of canvassers on COVID-testing was due to initial observability of household responses and not for example through any other enforcement or commitment effects later on.

3.6. Voter Registration Data

Our third social preference factor is identity. To test the impact of identity on the propensity to share personal data, we use natural variation in registered political affiliation. Previous empirical work shows that political affiliation is strongly correlated with the tendency to identify with a group. For example, Kranton and Sanders (2017) show that the percentage of Republicans, who strive for group identification is particularly high. Furthermore, Kranton et al. (2020) highlight in an Amazon MTurk experiment how some people with strong group identifications are willing to destroy income to reduce benefits for out-group members. This evidence suggests that political affiliation might be a good proxy for social identity and group-based perception of benefits of personal data sharing.

To measure social identity, we therefore obtain detailed voter registration data from the state of Utah. The provided file contains voter addresses, ZIP code, registered party affiliation, name, and participation in past elections. We merge it with our health survey and experimental data using ZIP code, address and, last name.

4. Results for Main Effects

We present the results in four steps. First, we present evidence on monetary incentives, which provides both novel evidence of personal data markets and a benchmark for our other treatments. Second, we present evidence on moral engagement and evidence and how different framings affect participation. Third, we present evidence on pro-social behavior by contrasting in-person and letter recruitment strategies. Finally, we use voter registration data to quantify the effects of social identity on personal data sharing decisions.

We present the results in this manner to focus on one aspect at a time and note that the isolated discussion consistent with the fuller model with interactions. In this section and the next, we focus on participation---what motivates people to provide their data through surveys and diagnostic tests. In section 6 we discuss sample selection of people with different propensities of infection.

4.1. Monetary Incentives

The top panel of Figure 3 presents the empirical counterpart to our theoretical graphs in Figure 1. In particular, the x-axis shows the strength of incentives on a per-person basis, while the y-axis shows percentage point participation relative to the control group with no incentive treatment. The different black lines capture participation in the online health survey, the virus test and the antibody test. The figure documents a number of key findings. It should first be noted that generally personal data sharing increased in the strength of incentives. This broad finding is consistent with basic economic logic and is reassuring, since it suggests that stronger incentives are effective in inducing personal data sharing. In addition, the point estimate for the weakest incentive payment of $10 is negative, implying a non-monotonic response of households' personal data sharing in response to weak incentives. This type of non-monotonic response is consistent with a categorical crowding-out effect, displayed in the top panel of Figure 1: the presence of monetary incentives induces respondents to perceive the situation more strongly as market transaction instead of prosocial behavior, see Gneezy and Rustichini (2000). However, as the strength of incentives increases, personal data sharing increases as well, consistent with at most weak substitability of incentives and social preferences in the bottom panel of Figure 1. Figure 3, is also instructive about the different types of personal health data that are shared. Specifically, monetary incentives seem to be more effective in stimulating symptom sharing, viral exposure, and social distancing data through the online health survey. This is unsurprising, since the effort cost to fill out this survey is comparably low. We obtain the second highest responses for antibody testing, presumably driven by the fact that the informational value of these antibody tests is the highest, even if they are very inconvenient.

In Table 3, we present the quantitative magnitudes of personal data sharing given the different incentive levels. Column (1), shows that \$20 and \$30 incentives increased survey participation by 1 and 2.1 percentage points relative to \$0, respectively, while a \$10 incentive had no statistically significant effect. These results suggest a slight convexity of the incentive treatments, as a 50% increase in the incentive amount from \$20 to \$30 implies a more than twice as large treatment effect. These effects are also economically significant in inducing data sharing. Relative to the baseline, these effects suggest that a \$20 and \$30 incentive increased participation by 7.6% and 16%, respectively.

The increasing and convex effect of incentives on participation is similar in the virus and antibody tests, which we report in columns (2)-(4). Specifically, the \$30 incentive increases participation in the virus and antibody tests by 1.6 and 2.0 percentage points relative to no incentive. In contrast, the \$20 and \$10 incentives had little or no effect on participation. The damped response for virus and antibody tests relative to survey participation is consistent with the notion that the effort cost to participate in the virus and antibody tests is greater. For example, many participants had to drive to a location near them and have a swab taken and blood drawn to participate in the two tests.

4.2. *Effort Costs and Distance*

The bottom panel of Figure 3 gives an overview of our basic findings on the direct effect of distance to the testing location on personal data sharing. As mentioned in Section 3, distance as a proxy for the effort costs of data sharing is an informative contrast to monetary incentives. Higher effort costs should reduce personal data sharing, but in a different way than lower monetary incentives. In particular, higher distance increases the effort costs of sharing personal health data without necessarily changing the perception of the participation from a prosocial motivation to a market exchange. As a consequence, one should not expect the same type of non-monotonic response in distance, as we observe for different monetary incentive levels. And as the bottom panel of Figure 3 shows, two out of three measures of personal health data do not exhibit any non-monotonicities. While both the propensity to share personal health data via the online survey and taking the antibody test steadily decline in distance, the propensity to take the virus test exhibits a non-monotonicity. However, it should also be noted that overall, the propensity to take the virus test declines much more steeply than the other types of personal health data sharing. This is potentially driven by the fact that the benefits from a virus test tend to be lower than the benefits of the antibody test, while the effort costs of taking the virus test is higher than the costs of filling out the online survey. This general pattern is quantitatively confirmed in table 4. The table also shows that distance effects tend to be slightly convex for people taking the online health survey. For instance, the distance coefficient of -0.035 for 3.25 miles implies that one additional mile has about the same effect as reducing monetary incentives by \$20. Under linearity, this would imply that each additional mile would increase monetary costs by \$20 to encourage people to the same degree to share their personal health data. However, the distance coefficient for 4.25 miles is -0.054, implying that 1 mile discourages data sharing as much as an additional cost of \$25.4.

4.3. *Framing, Moral Engagement and Information*

Table 5 reports the results of using different types of framing in recruitment letters, as discussed in Section 3. In this table, we focus on respondents who were offered \$30 incentive, which is why we will primarily focus on the differences between the different framings. A complete analysis of the interaction of framing and monetary incentives, which is consistent with the results presented here, is provided in section 5.1.

Panel A in Table 5 shows that framing and the related moral engagement is significant both statistically and economically. Qualitatively our results confirm that the use of a ”health crisis” in the recruitment letter triggered moral engagement and made people more likely to share their personal heath data, irrespective of whether through the online health survey or the COVID tests. At the same time, the use of the reference to an ”economic crisis” seems to have led to moral disengagement, resulting in depressed participation. The implied effect sizes are large. Our estimates suggest that the difference between economic and health framing are equivalent to about a $60 incentive in terms of Table 3. This is a huge effect, given that the only difference in the treatments consisted of a few lines of text in the recruitment letter, with a marginal cost of zero.

One potential explanation for what is driving the effects of the health framing as opposed to the economic framing is that the former might reveal information about the prevalence of COVID-19 instead of triggering moral engagement. To investigate this potential issue, we test the two additional information treatments described in Section 3.4. As Panel B of Table 5 shows, providing information on prevalence from the number of positive test, and the positivity rate does not come close to explain the responses to the health letter. In fact, all the coefficients on information about COVID-19 test positivity have the wrong sign, even as they are statistically insignificant. We contrast this information effect with providing information on the number and fraction of tested individuals without any mention of the associated positivity rates. As we argue in Section 3.4, information on total testing conducted can be viewed as indicating more quarantining of infected persons, which in turn might reduce the perceive risk of being exposed to the virus. Our results in Panel B of Table 5 are consistent with this interpretation. However, it is worthwhile emphasizing that this information effect again cannot explain the positive impact that our health framing on personal health data sharing.

4.4. *Image Motivation*

We examine the effect of image motivation on personal data sharing. Table 6 provides our results from the use of paid canvassers in recruitment. The results show that, social approval associated with the observability of personal data sharing decisions by canvassers has a powerful motivating effect. In fact, image motivation seems to boost participation for all outcomes by an order or magnitude more than incentives. For example, the results suggest that the use of canvassers increased the response rate for the online health survey by 26 percentage points, while boosting participation in virus testing by 15 percentage points and antibody testing by almost 18 percentage points.

These magnitudes are eight to ten times larger than the associated effects we document in Table 3 for monetary incentives. For the online health survey, they suggest that in order to obtain a similarly strong quantitative effect using monetary incentives, one requires a payment of over \$380 per person $(= \frac{0.266}{0.021} \times \$30)$. Taking into account that wages, transportation, and logistics costs for canvassers is about \$25 per household, and that there are on average two persons in each household, the return on investment from using canvassers is over 30 times higher $(= \frac{\$380}{\$12.5})$. Similarly, to induce a similar number of households to take the antibody test, one would need a monetary incentive of over \$268 per person $(= \frac{0.179}{0.02} \times \$30)$ with a 21 times higher return on investment $(= \frac{\$268}{\$12.5})$. These magnitudes are consistent with the view that image motivation and the corresponding social approval are both far more powerful and more cost effective than the use of direct monetary incentives to induce personal data sharing.

4.5. *Identity*

Table 7 provides the direct correlations of identity and participation in personal health data sharing. As discussed in Section 3.6, we use naturally occurring variation in voter registrations for this purpose, because we believe they signal personal group affiliations to the wider community. One of the disadvantages of using this naturally occurring variation is that we cannot rely on randomized treatments to exclude the importance of omitted variables. One candidate for such an omitted variable is geographic location, as it is possible that rural populations may be more conservative and less likely to participate in COVID-19 testing. We control for this issue directly by using local fixed effects that absorb differences in geographic location. Additionally, we believe that since voter registration is mostly pre-determined, we can broadly rule out that reverse causality might drive any correlations.

As Table 7 shows, registered conservatives are significantly less likely to participate in personal data sharing. One possible explanation for this finding is that they recognize that at least part of the benefits from data sharing might accrue to out-group members and are therefore hesitant to support COVID-19 testing. Alternatively, this finding might reflect scepticism of some conservatives about the magnitude of the COVID-19 health crisis and mistrust of scientific or government institutions[5].

5. Results for Interactions of Incentives and Social Preferences

In this section, we investigate interactions across several of the determinants we analyzed in the last section. The analysis of interaction effects is crucial to understand complementarity and substitutability between different tool, to facilitate data sharing, but is also more generally of interest. Specifically, the analysis of the interaction between incentives and different types of social preferences allows an investigation of how market domains and social domains interact. Since Smith (1759, 1776), economists have recognized that the demands of market logic and community spirit can potentially conflict (Hayek (1988); Smith (2002)). Our analysis here not only sheds light on how these two types of domains affect each other, and offers a direct quantification of these interactions in the context of personal data sharing.

Our interaction analysis also allows us to investigate how different types of social preferences interact. This is especially interesting when contrasting moral engagement and image motivation on the one hand and identity on the other hand. While moral engagement and image motivation typically involve altruism including towards anonymous strangers, identity by definition differentiates

[5] See, https://www.brookings.edu/blog/fixgov/2020/05/01/destroying-trust-in-the-media-science-and-government-has-left-america-vulnerable-to-disaster/

between in-group and out-group members. How these different types of social preferences interact can therefore offer interesting and novel insights into the nature of social preferences.

5.1. Incentives and Moral Engagement

Figure 4 provides a summary of the main results on the interactions of incentives and different types of recruitment letter framings, we discuss in Section 3.4. The figure shows participation on the y-axis, and different levels of monetary incentives on the x-axis. We display the impact of incentives on data sharing with three lines, denoting the interaction of three different framings in the recruitment letter: the ”health crisis” framing, the ”economic crisis” framing and the placebo, which had neither framing. The dashed blue line captures the placebo, which has no particular framing other than the basic information about the randomized COVID-19 testing and offered monetary incentives per person at the values shown on the x-axis. The red line denotes our economic framing, which focuses attention on recovery from the economic crisis. The solid black line shows responses to letters that included the health framing as well as different incentive levels.

Figure 4 shows that without any particular framing, there seems to be a strong negative response to offering low levels of incentives. This negative response eventually recovers at higher incentive levels and returns to the baseline at about $30 per person. This type of pattern is reminiscent of the laboratory findings of Gneezy and Rustichini (2000), who find similar effects in a sample of Israeli undergraduate students asked to answer IQ test questions. The effects can be explained by what we previously referred to as categorical crowding out in Section 2: the offer of incentives leads to moral disengagement by participants, leading to a systematic decrease in data sharing. However, once such moral disengagement has taken hold, the logic of incentives applies and higher incentives lead to more data sharing.

However, the results in Figure 4 also show that moral engagement and incentives are not always substitutes. Indeed, framing the recruitment letter in terms of an ongoing health crisis seems to be strongly complementary with incentives. At small values of our monetary incentive, categorical crowding out and an increase in incentives through compensation seem to offset each other. But it should also be noted that for higher values of monetary incentives, the slope the curve with health framing is steeper than the slope of incentive treatments without any framing. In other words, the health framing seems to not only reduce the amount of moral disengagement, but it also seems to help with moral engagement. Indeed, the steeper data sharing curve indicates that this moral engagement through the health framing is an example of a marginal complementarity of incentives and social preferences − see Figure 1.

Interestingly, the moral engagement effects of the economic framing for recruitment letters do not simply fit between the no framing and the health framing treatment. Instead, the results in Gigure 4 suggest that with the economic framing, is successful in stimulating data sharing, as could be explained by a categorical crowding-in effect in figure 1. However, at the same time, higher monetary treatments seem to increase data sharing less than for the no-framing baseline. This is consistent with the economic framing and incentives being weak marginal substitutes as in the bottom panel of Figure 1.

We formalize these results with an interaction effect analysis in Table 8. The table broadly confirms the qualitative patterns shown in Figure 4. The table also allows us to highlight some important quantitative results. For example, health framing and a $30 per person incentive stimulate data sharing by around 4.7% for the online survey and by 4.9% for antibody testing. These effects are over two times larger than the average effects of this $30 incentive across framings, from Table 3. As a result, much of the overall stimulating effects of incentives can be attributed to treatments that morally engaged potential respondents with the health framing. In contrast, the alternative economic framing or no framing led to insignificant data sharing responses, even for relatively high incentive amounts of $30 per person.

These results highlight the importance of complementarity between framing and incentives. The moral engagement or disengagement effects triggered by different types of framing can either lead incentives to backfire or can promote the effectiveness of incentives.

5.2. Incentives (or Effort Costs) and Image Motivation

We show our baseline results for the interaction of image motivation and monetary incentives in the bottom panel of Figure 4. The x-axis shows cost per person, which is the direct monetary cost of our interventions. For letter recruitment, these costs mainly capture the monetary incentives per person. In contrast, for the canvassers, we drew in the estimates cost per household of $25, which implies a per person cost of $12.5, as explained above. Additionally, a random subset of households were visited by a canvasser, who offered them $10 per person for their participation, which we mark in the figure at the $22.5 line. The bottom panel of figure 4 shows that image motivations and incentives are strong substitutes. Since we did not vary the incentive amounts across canvassers, we cannot distinguish here between categorical or marginal crowding out effects. But the basic presence of crowding out of image motivation by incentives is consistent with laboratory and small-scale field evidence, such as Ariely et al. (2009). However, the sheer magnitude of the crowding out effect is notable: a $10 per person incentive reduces data sharing by roughly 10 percentage points or a third of the total stimulative effect of image motivation on participation. This is a huge effect, especially compared with the stimulative impact of incentives. Remember from Table 8, that even in the most effective case of a $30 per person incentive combined with health framing, these incentives increased data only by about 4.7%. In contrast, the substitutability of image motivation and incentives implies that the crowding out effect of just a $10 per person incentive is larger by about 30%. As a result, providing incentives strongly undermines the stimulative effect of image motivation and social approval on personal health data sharing.

Table 9 reports the effects between image motivation and a $10 per person incentive. As can be seen, these effects are consistent even across different framings.

To investigate image motivation further, we analyze the interaction of image motivation and effort, as measured by distance to the testing location in table 10. If distance is unobservable by canvassers, one might not expect the presence of canvassers and distance to interact. On the other hand, if higher distances allow people to more credibly signal a virtuous image, we predict that sufficiently high distances might stimulate data sharing.

Indeed, as the results in Table 10 shows, small increases in distance exhibit negative interaction effects with the presence of canvassers. This finding is consistent with the view that small distances are insufficient to signal a virtuous image to canvassers. Furthermore, with increasing distance, data sharing becomes more likely if canvassers are used for recruiting. This finding is especially strong for distances of about 3-4 miles, which implies a very strong positive interaction effect between image motivation and effort. This result suggests, that people might exert a moderate effort to signal a virtuous image to canvassers and then follow-through with this signalling by participating in COVID-19 testing.

5.3. Incentives and Identity

In this section, we analyze the interaction between incentives and identity, as measured by voter registrations. Table 11 reports two key results. First, as highlighted by the baseline effects of incentives, these incentives are highly effective in inducing non-conservative people to share their personal health data. Across the different columns of table 11 incentives increase data sharing by between 4% and over 20%. For non-conservative voters, the results in the table also show that larger incentive payments induce more data sharing, as expected from basic economics and our discussion in section 2.

Second, people with conservative identity become less likely to share their personal data if offered incentives to do so. This result is consistent with the previously mentioned findings by Kranton et al. (2020), which show that people with strong in-group identities are even willing to destroy income if this reduces benefits for out-group members. Table 11 shows that higher incentive levels strongly discourage conservatives to share their personal health data. In the context of our theoretical discussion in Figure 1, this suggests that incentives and identity can be strong marginal substitutes. These results are surprising for at least two reasons. First, the logic of Kranton et al. (2020) can only explain our results if most conservatives think that the use of their personal health data primarily benefits out-group members (in other words: non-conservatives). This is surprising, since a majority of Utah's state population consists of conservatives and the state government is reliably conservative as well. On the other hand, COVID-19 infection risks are higher in areas with high population densities, such as cities. Cities in turn tend to be much less conservative than rural areas. Second, even if conservatives are willing to forgo the benefits of incentives to reduce benefits from sharing their data for non-conservatives, it seems surprising that willingness to share data declines for higher incentive levels. This result suggests that offering monetary incentives to introduce market logic does not just simply undermine identity effect, but might negatively reinforce it. It is consistent with the view that incentives might be perceived as a type of coercion, as emphasized by Grant (2006), which then triggers a contrarian response. In other words, offering ”side payments” to achieve pareto-improvements might backfire if identity effects are strong.

These results highlight how easily the use of incentives can be counterproductive in stimulating data sharing. Yet, they do show that identity-based incentives can be highly effective in stimulating data sharing.

5.4. Social Identity and Moral Engagement

In this section, we analyze the interaction between moral engagement and social identity. This analysis is facilitated by political developments during the US presidential campaign of 2020, which took place during the COVID-19 pandemic. Specifically, President Trump, who had an approval rating of over 80% among conservatives, repeatedly either played down the severity of the pandemic or spun the narrative that the health crisis was a ”liberal hoax”[6]. At the same time, Trump emphasized the need for economic recovery[7] and opposed health measures that could potentially dampen the speed of recovery. In contrast, presidential candidate Joe Biden emphasized the need to manage the health crisis and even avoided holding campaign rallies to not further spread the virus. As a consequence of these differences in priority, it seems reasonable to expect that identity will likely interact with moral engagement triggered by the framing of recruitment letters. Specifically, one might expect that conservatives respond more positively to the economic framing, while non-conservatives respond more positively to the health framing.

Table 12 show that this is in fact not the case. Starting with absolute response levels, as columns (1) and (2) show, non-conservatives respond to economic and health framings by systematically reducing data sharing. In other words, framing the decision to share personal health data in terms of

[6]For example, at a Trump campaign rally in February, President Trump said, ‘‘Now the Democrats are politicizing the coronavirus ... This is their new hoax.”

[7]Trump did so repeatedly, see:https://www.nytimes.com/2020/03/23/us/politics/trump-coronavirus-restrictions.html; https://www.nytimes.com/2020/05/06/us/politics/trump-coronavirus-recovery.html

helping to overcome a health or economic crisis, leads to moral disengagement by non-conservatives. These effects are the largest for filling out the online health survey and taking the virus COVID-19 test and to a lesser extent to the antibody test. The magnitudes of this reduction in data sharing are large and comparable to the effect of the $30 per person incentive with the health framing from table 8. Furthermore, these negative responses by non-conservatives do not differ significantly when comparing health with economic framings.

In contrast, conservatives respond positively to both economic and health framings. The moral engagement effect is quantitatively large. For the online health survey the stimulative effect equals the effect of the $30 per person incentive for the health framing from table 8. At the same time, the results in 12 show, that the health framing boosts conservatives' personal data sharing by 47% more than the $30 per person incentive with the health framing from table 8 (coefficient of 0.067 instead of 0.047). It is surprising that conservatives respond more positively to the health framing than the economic framing. However, these effects are consistent with the view that conservatives feel morally engaged by both the economic and health crisis framing. At the same time, it should be noted that such a moral engagement suggests that conservatives are willing to potentially help out-group members. This effect stands in contrast to the strongly negative effects of incentives for conservatives.

5.5. Social Identity and Image Motivation

How do image motivation and social approval interact with social identity? The results in 13 highlight that conservatives are substantially more motivated by social approval than non-conservatives. The interaction effect of conservative voter registration and the presence of canvassers for recruitment indicates that image motivation increases personal data sharing by around 30 percentage points. Furthermore, the magnitude of the interaction effects are similar across our different types of personal health data, such as the online health survey, the virus test and the antibody test.

These results highlight two separate implications. From the perspective of our theory discussion in section 2, the results in Table 13 show that image motivation and social identity are complementary tools to stimulate personal data sharing. This is especially interesting when contrasting these results to how moral engagement and social identity interact. In both cases, conservatives seem to respond stronger to a combination of different types of social preferences, while non-conservatives systematically respond less to both social approval and moral framing.

A related insight is that the complementarity in table 13 suggests that the use of canvassers to trigger image motivation and social approval should optimally be targeted towards conservatives. In other words, social approval is a less potent tool to motivate non-conservatives to share personal data.

6. Sample Selection in Randomized Testing?

This section provides a discussion of sample selection for randomized COVID-19 testing. A potential issue for randomized testing of emerging infectious diseases is sample selection of sick people. In any democracy, monitoring of disease prevalence necessarily requires voluntary participation. However, such voluntary participation can lead to a biased measurement of disease prevalence. For example, people who suspect that they are sick might be more likely to participate in COVID-19 testing, thereby biasing estimates of the prevalence of COVID-19 upwards. On the other hand, people who are more health conscious and more willing to aggressively socially distance might be more likely to volunteer for COVID-19 testing. Since health conscious people are less likely to be infected, this might bias estimates of the prevalence of the disease downward.

A simple economic model can be used to illustrate these sample selection effects and how our analysis might address this issue. Let state-dependent utility for person i be given by $u_{i,\tau,h}$, where $\tau \in \{0,1\}$ is an indicator that is one if a COVID-19 test was taken and $h \in \{0,1\}$ is an indicator that is one if person i is infected. People do not know whether they are infected, but form priors based on symptoms they observe and other information, such as travel history, exposure to infected people, own social distancing behavior etc. This prior for person i is denoted by $p_i = Prob_i\{h = 1\}$.

Expected utility from taking a COVID-19 test is given by:

$$EU_{i,1} = p_i \cdot u_{i,1,1} - (1 - p_i) \cdot u_{i,1,0} - e_i + m_i \tag{4}$$

where e_i denotes the private effort cost for taking the test (e.g., for walking or driving to the testing location) and m_i denotes a monetary incentive for taking the test. Similarly, the expected utility from not taking the test is given by:

$$EU_{i,0} = p_i \cdot u_{i,0,1} - (1 - p_i) \cdot u_{i,0,0} \tag{5}$$

To simplify, we assume the test result does not affect utility if a person is healthy, i.e. $u_{i,1,0} = u_{i,0,0}$, while testing positive for the disease helps choosing the right therapy and avoid further infection of other people. Therefore, a rational decision-maker without any social preferences would only take the a COVID-19 if:

$$p_i \cdot (u_{i,1,1} - u_{i,0,1}) - e_i + m_i \geq 0 \tag{6}$$

Private information about the risk of being infected will show up in p_i, while the higher utility for health-conscious people will show up in $u_{i,1,1} - u_{i,0,1}$. Therefore, (6) encompasses both types of potential selection biases discussed at the beginning of this section.

To understand how selection equation (6) leads to sample selection, consider the case where $u_{i,1,1} - u_{i,0,1} > 0$, but identical across people, and p_i varies across people. In this example, people with the highest beliefs that they are infected p_i, will be the first to take tests and share data, followed by people with lower beliefs of infection and so on. As a result, if p_i is higher for infected people than healthy people, shifts in either effort costs e_i or incentives to share data m_i should change the fraction of people with positive tests. A similar logic can be applied to heterogeneity in $u_{i,1,1} - u_{i,0,1}$.

However, our results in Tables 3 and 4 show that different levels of incentives or distance to the testing locations do not affect positivity rates for virus or antibody COVID-19 tests. This evidence suggests that sample selection in randomly sampled disease testing is less of an issue than previously believed.

7. Discussion

In this section, we draw together several themes, that have emerged. An enduring question in economics is how governments and companies can stimulate sharing of personal data. The standard approach to this question is ”mechanism design” and the analysis of direct mechanisms that abstract from institutional details, see Hurwicz (1960); Vickrey (1961); Clark (1971); Groves (1973); Myerson (1981); Maskin (1999) among others. In this standard mechanism design approach, incentives, also called ”information rents” are optimally chosen to induce people to share private information. However, our results, as others in behavioral economics, suggest that monetary incentives are highly ineffective tools for stimulating data sharing. Our field evidence suggests that a ”behavioral mechanism design” approach that harnesses social preferences to induce data sharing might can be much more powerful than the use of incentives. Furthermore, our results provide insight on what types of social preferences are especially promising for further theoretical and empirical study. Specifically, image motivation and social identity exhibit strong quantitative impacts on personal data sharing.

A unique feature of our study is our analysis of the interaction effects between incentives and social preferences, as well as among different types of social preferences. This analysis has highlighted a number of important complementarities, as well as crowding out effects. This suggests that firms and governments navigating the intersection of market and social domains need to focus their attention on questions of consistency among different tools to stimulate personal data sharing. This is an especially pertinent question for firms, which often invoke social preferences through corporate social responsibility and other image motivation initiatives, while also being party to market transactions with customers. This dual nature of how firms interact with customers can imply substantial risks. As our analysis of the interaction of incentives and different types of social preferences shows, incentives tend to strongly undermine the stimulative effect of some forms of social preferences, such as image motivation. On the other hand, our results also highlight opportunities for firms to exploit complementarities between incentives and moral engagement. This is highlighted in our results showing that recruitment letters with health framing and incentives significantly stimulated personal data sharing.

The substitutability of incentives and some types of social preferences, such as image motivation and identity, raises important questions for mechanism design and policy. As noted by Titmuss (1970), ineffectiveness of incentives might lead mechanism designers to overuse them. On the other hand, Bowles and Polanía-Reyes (2012) show that ineffectiveness of incentives can also imply that mechanism designers optimally use more of them, similar to a doctor who prescribes a higher dose of a weak medicine. A third alternative Bowles and Polanía-Reyes (2012) discuss is to abandon the treatment due to its ineffectiveness. We note that our results suggest important shortcomings of the ”use more weak medicine” approach. Our results indicate that image motivation through the use of canvassers is many times more effective in stimulating personal data sharing than incentives, even if the incentives are combined with moral engagement through health framing. As a result, continuing to increase incentives until an equivalent effect to the use of canvassers is reached can become incredibly wasteful.

Finally, our analysis emphasizes the importance of the social environment in matters of personal data sharing. We demonstrate the importance of social identity and partisanship for personal data sharing decisions. This point is most natural for public policy, which also always needs to take account of politics. But this point is also important for firms, which need to navigate the rising importance of corporate social responsibility, as well as the increase politicization of consumption decisions. A simple example for such politicization are calls for boycotts, that several premium brands, such as Amazon, Starbucks, BP and other companies had to confront in recent years[8]. Here, our analysis suggests that governments and firms need to deploy tools to stimulate personal data sharing in a highly targeted way. Use of social approval and moral engagement is effective for some

[8]For other examples, see https://www.ethicalconsumer.org/ethicalcampaigns/boycotts

groups, but can also backfire if used for the wrong groups. Here, a more extensive analysis of the interactions of different types of identity and social preference-based policies is urgently needed.

8. Conclusion

This paper establishes credible causal estimates of the effectiveness of incentives and social motivations for stimulating personal data sharing. We do so by running a sequence of RCTs among tens of thousands of households during the COVID-19 pandemic in Utah. Our main findings suggest that some types of social motivations, such as image motivation and identity, are much more effective than incentives to stimulate personal data sharing. Furthermore, incentives easily undermine the effectiveness of social motivations, while incentives themselves are only effective when paired with the right type of moral engagement.

Our main results suggest that a behavioral approach to mechanism design, which takes the role of social preferences in personal data sharing decisions seriously, is a promising area for further research. This type of analysis for personal data sharing decisions could be important for several applications.

First, a dozen startups have recently begun to establish the necessary technological infrastructure to enable individuals to sell their personal data to firms, thereby creating formal personal data sharing markets[9]. Such formal personal data sharing markets have been shown to have desirable welfare properties, see Jones and Tonetti (2020). But even beyond formalized spot markets for personal data, firms regularly seek out market research or consumer satisfaction data as basis for data-driven decisions. This type of data gathering has become increasingly difficult due to falling response rates. Our research suggests new ways to stimulate personal data sharing through social motivations.

Our results also have broader implications for several areas of public finance. For example, allocation of federal programs and apportioning of representation in Congress rely on population

[9]See for example: https://www.forbes.com/sites/lucysherriff/2019/03/29/this-app-enables-you-to-make-money-off-your-own-personal-data and https://www.bbc.com/news/business-47027072

estimates from the Census. Similarly, the Economic Census is used to measure economic activity in the US economy. Our results on personal data sharing are likely to be important to boost participation in population and economic Census programs in the face of the previously mentioned downward trend in response rates. Additionally, our results suggest that social preferences can play a crucial role in the design of mechanisms to induce the revelation of personal preferences for public good spending, a classic question of mechanism design, see Clark (1971); Groves (1973).

Finally, another natural application area is political economy, where public polling for elections faces much of the same measurement issues as proactive COVID-19 testing. Our results on the need to target incentives to some partisan groups but not others can improve the efficiency of data gathering efforts.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar, ''Testing, Voluntary Social Distancing and the Spread of an Infection,'' *Working Paper, Massachusetts Institute of Technology*, 2020.

Akerlof, G. and R. Kranton, ''Economics and Identity,'' *Quarterly Journal of Economics*, 2010.

Akerlof, G. and R. Kranton, ''Identity Economics,'' *Princeton University Press*, 2010.

Andreoni, J., ''Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence,'' *Journal of Political Economy*, 1989, *97* (6), 1447--1458.

Andreoni, J., ''Impure Altruism and Donations to Public Goods: Theory of Warm-Glow Giving,'' *Economic Journal*, 1990, *100* (401), 467--477.

Andreoni, J., J. Rao, and H. Trachtman, ''Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving,'' *Journal of Political Economy*, 2017, *125* (3), 625--653.

Ariely, D., A. Bracha, and S. Meier, ''Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially,'' *American Economic Review*, 2009.

Benabou, R. and J. Tirole, ''Incentives and Prosocial Behavior,'' *American Economic Review*, 2006.

Bowles, S. and S Polanía-Reyes, ''Economic Incentives and Social Preferences: Substitutes or Complements?,'' *Journal of Economic Literature*, 2012.

Burtch, G., Y. Hong, R. Bapna, and V. Griskevicius, ''Stimulating Online Reviews by Combining Financial Incentives and Social Norms,'' *Management Science*, 2018, *64* (5), 2065--2082.

Clark, E., ''Multipart Pricing of Public Goods,'' *Public Choice*, 1971.

DellaVigna, S., J. List, and U. Malmendier, ''Testing for altruism and social pressure in charitable giving,'' *Quarterly Journal of Economics*, 2012.

Gabisch, J. and G. Milne, ''The impact of compensation on information ownership and privacy control,'' *Journal of Consumer Marketing*, 2014.

Gallus, J., ''Fostering Public Good Contributions with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia,'' *Management Science*, 2016.

Gneezy, U. and A. Rustichini, ''Pay Enough or Dont Pay at All,'' *Quarterly Journal of Economics*, 2000.

Grant, R., ''Ethics and Incentives: A Political Approach,'' *American Political Science Review*, 2006.

Groves, T., ''Incentives in Teams,'' *Econometrica*, 1973.

Hayek, F.A., ''The Fatal Conceit,'' *University of Chicago Press.*, 1988.

Hui, K., H. Teo, and T. Lee, ''The value of privacy assurance: An exploratory field experiment,'' *MIS Quarterly*, 2007.

Hurwicz, L., ''Optimality and Informational Efficiency in Resource Allocation Processes,'' *Mathematical Methods in the Social Sciences*, 1960.

Jones, C. and C. Tonetti, ''Nonrivalry and the Economics of Data,'' *American Economic Review*, 2020.

Kranton, R. and S. Sanders, ''Groupy vs. Non Groupy Social Preferences: Personality, Region, and Political Party,'' *American Economic Review Papers and Proceedings*, 2017.

Kranton, R., M. Pease, S. Sanders, and S. Huettel, ''Deconstructing Bias in Social Preferences Reveals Groupy and Not Groupy Behavior,'' *Proceedings of the National Academy of Sciences*, 2020.

Lacertera, N., M. Macis, and R. Slonim, ''Economic Rewards to Motivate Blood Donations,'' *Science*, 2014.

Lacertera, N., M. Macis, and R. Slonim, ''Rewarding Volunteers: A Field Experiment,'' *Management Science*, 2014.

List, J., J. Murphy, M. Price, and A. James, ''Do Appeals to Donor Benefits Raise More Money than Appeals to Recipient Benefits? Evidence from a Natural Field Experiment with Pick.Click.Give.,'' *NBER Working Papers*, 2019.

Maskin, E., ''Nash Equilibrium and Welfare Optimality,'' *Review of Economic Studies*, 1999.

Mellstrom, C. and M. Johannesson, ''Crowding Out in Blood Donation: Was Titmuss Right?,'' *Journal of the European Economic Association*, 2008.

Myerson, R., ''Optimal Auction Design,'' *Mathematics of Operations Research*, 1981.

Premazzi, K., S. Castaldo, M. Grosso, P. Raman, S. Brudvig, and C. Hofacker, ''Customer information sharing with evendors: The roles of incentives and trust,'' *International Journal of Electronic Commerce*, 2010.

Samore, Matthew, Adam Looney, Brian Orleans, Nathan Seegert Tom Greene, Julio C Delgado, Angela Presson, Chong Zhang, Jian Ying, Yue Zhang, Jincheng Shen, Patricia Slev, Maclean Gaulin, Mu-Jeung Yang, Andrew T. Pavia, and Stephen C. Alder, ''SARS-CoV-2 seroprevalence and detection fraction in Utah urban populations from a probability-based sample,'' 2020.

Seegert, Nathan, Maclean Gaulin, Mu-Jeung Yang, and Francisco Navarro-Sanchez, ''Information Revelation of Decentralized Crisis Management: Evidence from Natural Experiments on Mask Mandates,'' *Working Paper Eccles Institute*, 2020.

Smith, A., ''An Inquiry into the Nature and Cause of the Wealth of Nations,'' *https://www.ibiblio.org/ml/libri/s/SmithA*$_{W}ealthNations_{p}.pdf$, 1776.

Smith, Adam, *The Theory of Moral Sentiments*, London: A. Miller, 1759.

Smith, V., ‘‘Constructivist and Ecological Rationality in Economics,’’ *Nobel Prize Lecture*, 2002.

Titmuss, R., ‘‘The Gift Relationship,’’ *London: Allen and Unwin*, 1970.

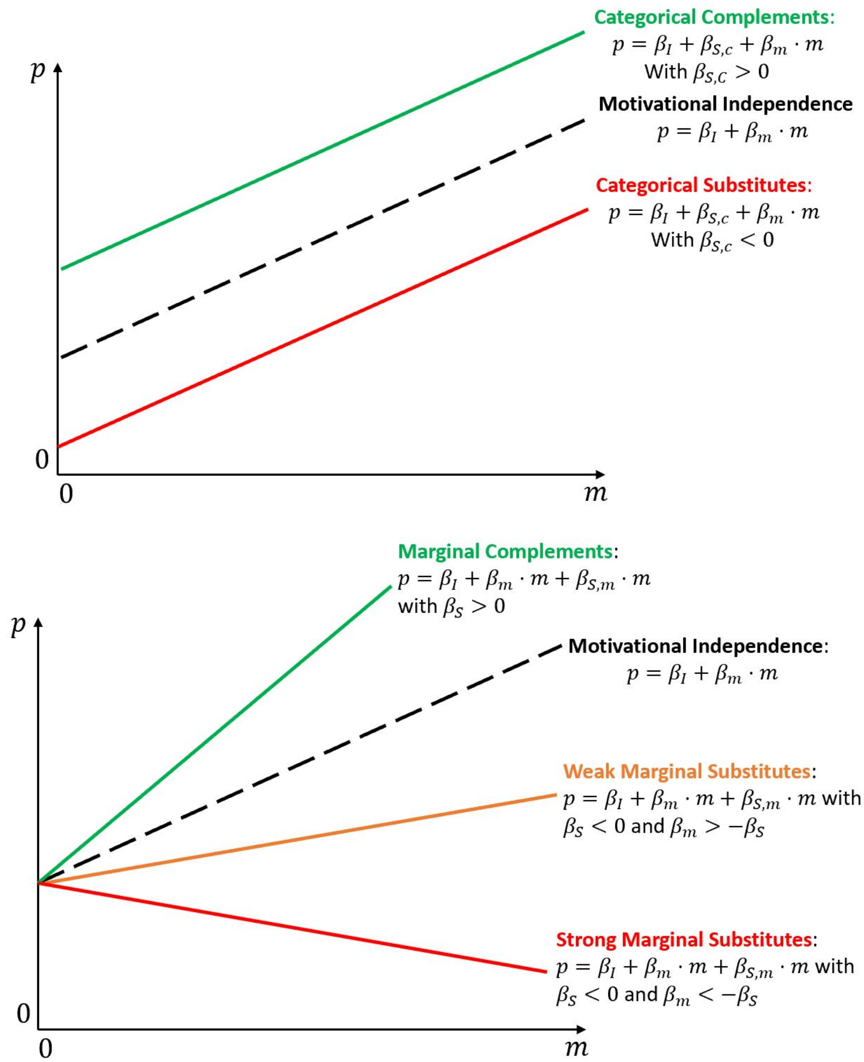
Vickrey, W., ‘‘Counterspeculation, Auctions and Competitive Sealed Tenders,’’ *Journal of Finance*, 1961.

Yang, M., A. Looney, M. Gaulin, and N. Seegert, ‘‘What Drives the Effectiveness of Social Distancing in Combatting COVID-19 across U.S. States?,’’ *Working Paper, University of Utah*, 2020.

Yang, MJ, N. Seegert, M. Gaulin, A. Looney, B. Orleans, A. Pavia, K. Stratford, M. Samore, and S. Alder, ‘‘What is the Active Prevalence of COVID-19?,’’ *Working Paper, University of Utah*, 2020.

9. Tables and Figures

Figure 1: Complementarity or substitutability between incentives and social preferences



Note: Figures display value of cash incentive (”money”) on the x-axis and intensity of activity (e.g. participating in personal data sharing) in the y-axis. Top panel shows categorical complementarity/substitutability of incentives and social preferences. Bottom panel shows marginal complementarity/substitutability of incentives and social preferences.

Figure 2: Canvassers and Testing Sites

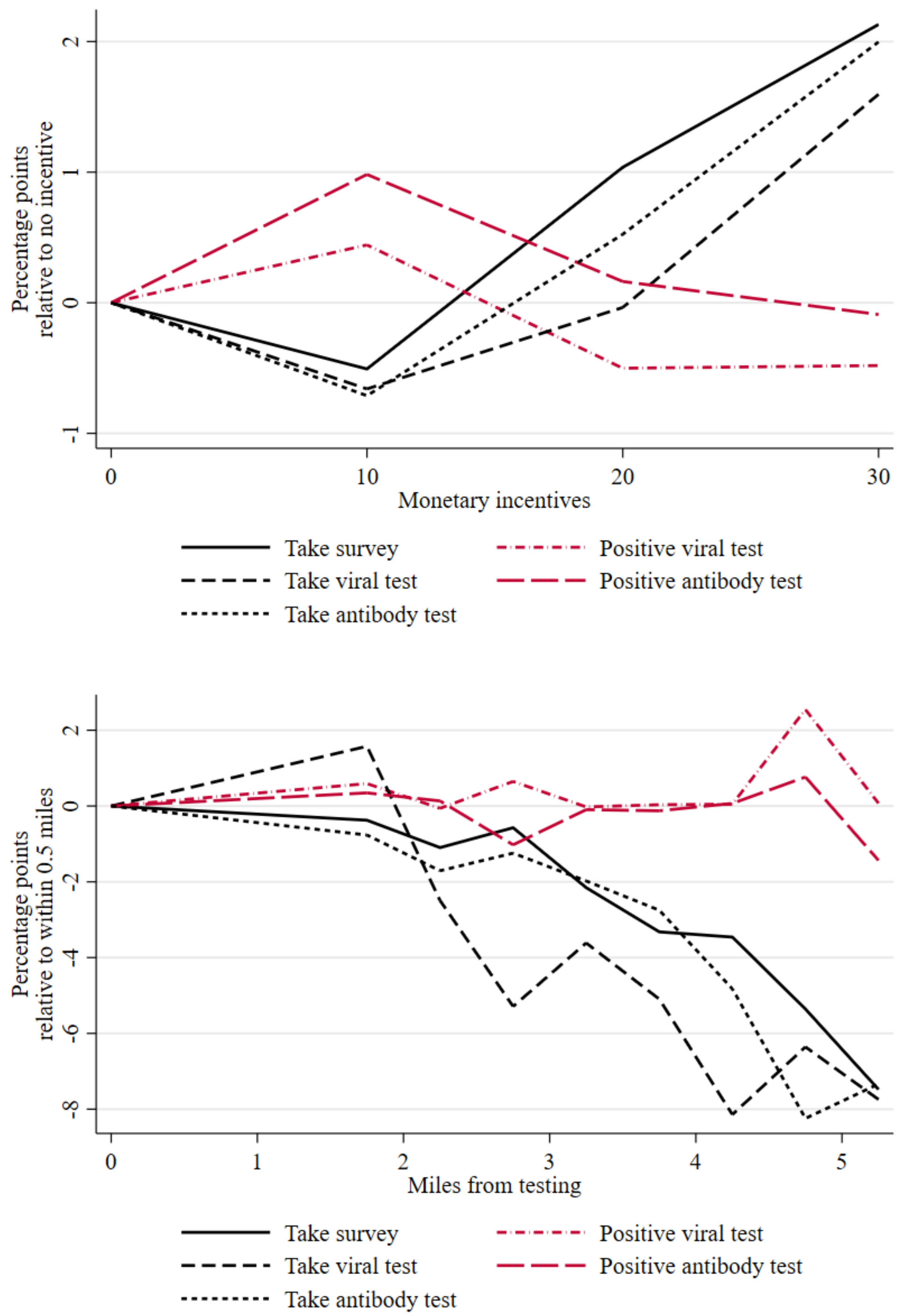
DRIVE-THROUGH TESTING SITE

MOBILE TESTING SITE

Canvassers from Hope Corps

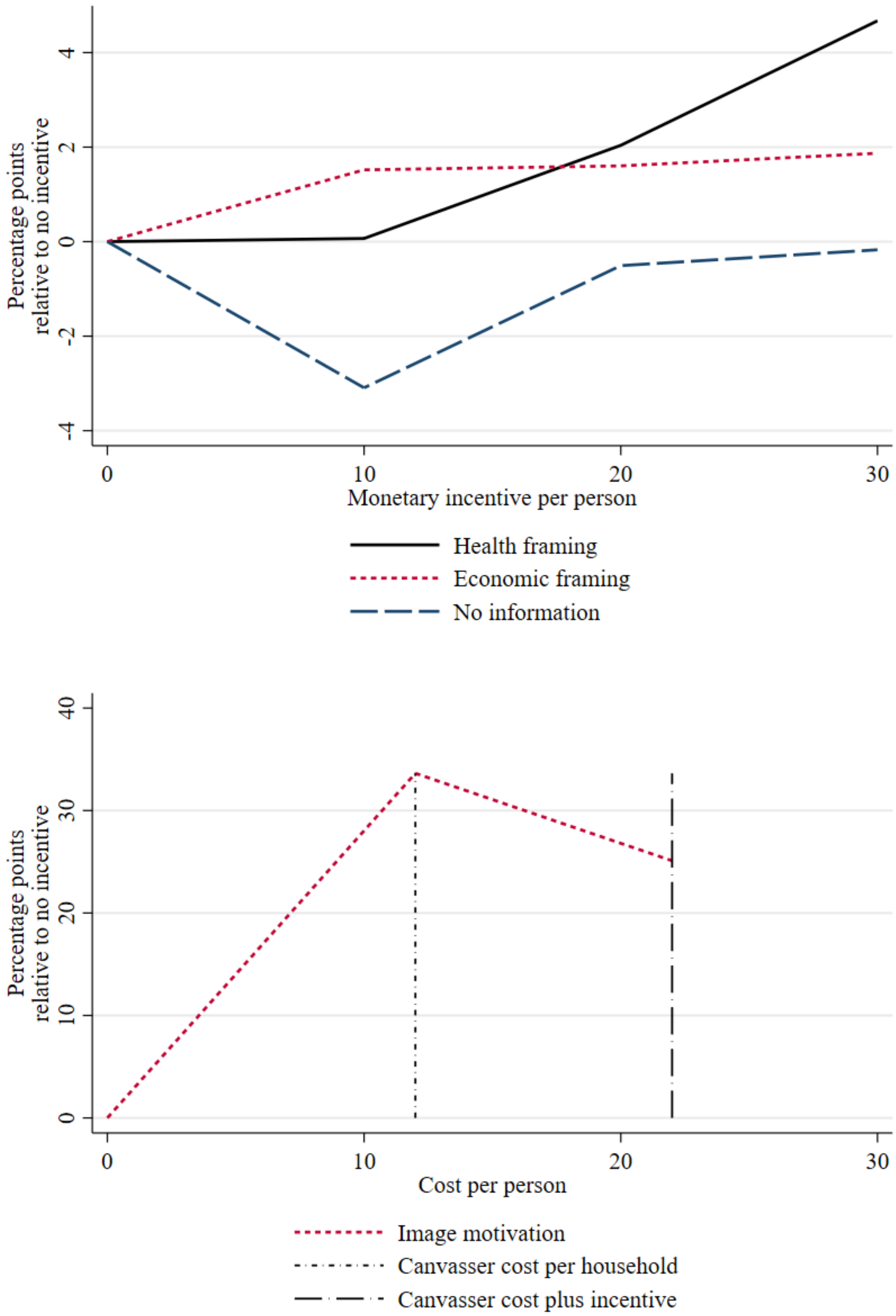
Note: Canvassers were college students, who were hired in cooperation with Utah Community Builders, The David Eccles School of Business at the University of Utah.

Figure 3: Effect of incentives and effort on personal data sharing



Note: Top panel shows the effect of different incentive levels on personal data sharing. Bottom panel shows the effect of distance to closest testing location (which can be either a drive-through site or a mobile testing location) on personal data sharing.

Figure 4: Interactions of incentives and framing or image motivation in personal data sharing



Note: Figures the the impact of moral engagement and image motivation on personal data sharing. Personal data sharing is measured by participation in health survey. Top panel shows different levels of incentive treatments for different types of framings in recruitment letters. Bottom panel shows the interaction of image motivation (canvassers) and a $10 incentive. Canvasser cost per person was approximately $12.50

Table 3: **Effect of monetary incentives on personal data sharing**

This table shows the results from Washington and Glendale county, where participants were offered one of {$0, $10, $20, $30}. Standard errors are given in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
$10 incentive	-0.005	-0.007	-0.007	-0.007	0.004	0.010
	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)
$20 incentive	0.010*	-0.000	0.005	0.005	-0.005	0.002
	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.006)
$30 incentive	0.021***	0.016***	0.020***	0.020***	-0.005	-0.001
	(0.006)	(0.006)	(0.006)	(0.006)	(0.004)	(0.006)
Constant	0.131***	0.111***	0.116***	0.116***	0.007**	0.012***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
F-statistic	6.874	5.173	7.172	7.037	2.333	1.116
R-Squared	0.001	0.001	0.001	0.001	0.003	0.001
Observations	26,487	26,487	26,487	26,487	2,584	2,823

Table 4: **Effect of effort costs on personal data sharing**

This table shows how participation changes with distance in primary and secondary frames. Distance is in miles. Sample is truncated at 5.5 miles away. Ninety percent of observations are within 5.5 miles and the maximum distance is 16 miles. Standard errors are given in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
$1.5 \leq x < 2$ miles	-0.004	0.016**	-0.008	-0.007	0.006***	0.003
	(0.009)	(0.008)	(0.008)	(0.008)	(0.002)	(0.004)
$2 \leq x < 2.5$ miles	-0.011	-0.025**	-0.017	-0.018	-0.001	0.001
	(0.011)	(0.010)	(0.011)	(0.011)	(0.003)	(0.006)
$2.5 \leq x < 3$ miles	-0.006	-0.053***	-0.012	-0.013	0.006	-0.010
	(0.012)	(0.011)	(0.012)	(0.012)	(0.005)	(0.007)
$3 \leq x < 3.5$ miles	-0.022	-0.036***	-0.020	-0.018	-0.000	-0.001
	(0.013)	(0.012)	(0.013)	(0.013)	(0.005)	(0.008)
$3.5 \leq x < 4$ miles	-0.033**	-0.051***	-0.027*	-0.028*	0.000	-0.001
	(0.015)	(0.014)	(0.015)	(0.015)	(0.006)	(0.009)
$4 \leq x < 4.5$ miles	-0.035***	-0.081***	-0.048***	-0.049***	0.000	0.001
	(0.013)	(0.012)	(0.013)	(0.013)	(0.006)	(0.008)
$4.5 \leq x < 5$ miles	-0.054**	-0.064***	-0.082***	-0.083***	0.025***	0.008
	(0.021)	(0.019)	(0.020)	(0.020)	(0.009)	(0.016)
$5 \leq x < 5.5$ miles	-0.075***	-0.077***	-0.073***	-0.074***	0.001	-0.014
	(0.018)	(0.016)	(0.018)	(0.018)	(0.008)	(0.013)
canvassing	0.301***	0.168***	0.211***	0.211***	0.004*	-0.004
	(0.007)	(0.006)	(0.007)	(0.007)	(0.002)	(0.004)
Constant	0.204***	0.161***	0.191***	0.191***	-0.001	0.015***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.002)	(0.004)
F-statistic	343.532	169.462	193.770	193.810	1.983	0.672
R-Squared	0.112	0.058	0.066	0.066	0.004	0.001
Observations	24,603	24,603	24,603	24,603	4,777	6,109

Table 5: **Framing and Moral (Dis-)Engagement**

This table shows the results from Washington and Glendale county, limited to those houses offered $30. Standard errors are given in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Health and economic motivations						
Health	0.020**	0.024***	0.021***	0.022***	0.002	0.008
	(0.008)	(0.007)	(0.008)	(0.008)	(0.002)	(0.006)
Economic	-0.020**	-0.012*	-0.018**	-0.017**	0.002	-0.001
	(0.008)	(0.007)	(0.008)	(0.008)	(0.002)	(0.006)
Constant	0.174***	0.128***	0.156***	0.156***	-0.000	0.008*
	(0.006)	(0.005)	(0.006)	(0.006)	(0.002)	(0.004)
F-statistic	11.762	13.003	12.582	12.725	0.511	1.315
R-Squared	0.002	0.002	0.002	0.002	0.001	0.001
Observations	12,923	12,923	12,923	12,923	1,464	1,794
Panel B: Types of health information						
Positive tests	-0.014	-0.019	-0.012	-0.012	0.001	-0.003
	(0.017)	(0.014)	(0.014)	(0.014)	(0.007)	(0.014)
Total tests	-0.026	-0.034**	-0.036**	-0.036**	0.001	0.000
	(0.017)	(0.014)	(0.014)	(0.014)	(0.008)	(0.015)
Constant	0.278***	0.172***	0.180***	0.180***	0.005	0.023**
	(0.012)	(0.010)	(0.010)	(0.010)	(0.005)	(0.010)
F-statistic	1.171	2.958	3.261	3.261	0.013	0.030
R-Squared	0.001	0.001	0.002	0.002	0.000	0.000
Observations	4,101	4,101	4,101	4,101	557	598

Table 6: **Image motivation (canvassers)**

This table shows all participants (canvassers and letters) in phase 1, glendale, cache and weber counties. Standard errors are given in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
Canvassing	0.266***	0.150***	0.179***	0.179***	0.001	-0.001
	(0.004)	(0.003)	(0.003)	(0.003)	(0.001)	(0.002)
Constant	0.159***	0.122***	0.141***	0.142***	0.003***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
F-statistic	5222.664	2057.713	2616.442	2625.886	0.879	0.076
R-Squared	0.084	0.035	0.044	0.044	0.000	0.000
Observations	57,138	57,138	57,138	57,138	8,432	10,141

Table 7: **Social identity**

This table shows voter registration with controls for design

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
Registered conservative	-0.739***	-0.553***	-0.631***	-0.632***	0.000	0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.001)	(0.002)
Constant	0.887***	0.658***	0.768***	0.770***	0.002	0.013***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.003)
F-statistic	5683.418	3104.955	4258.825	4276.046	2.209	2.405
R-Squared	0.311	0.197	0.252	0.253	0.001	0.001
Observations	88,328	88,328	88,328	88,328	12,033	14,248

Table 8: **Effect of monetary incentives and framing on personal data sharing**

This table shows the results from Washington and Glendale county, where participants were offered one of {\$0, \$10, \$20, \$30} and different information prompts. Standard errors are given in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Survey			Antibody test		
	Health	Economic	No information	Health	Economic	No information
	(1)	(2)	(3)	(4)	(5)	(6)
\$10 incentive	0.001	0.015	-0.031***	0.002	0.001	-0.024***
	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
\$20 incentive	0.020**	0.016	-0.005	0.010	0.008	-0.002
	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
\$30 incentive	0.047***	0.019*	-0.002	0.049***	0.013	-0.002
	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
Constant	0.124***	0.121***	0.147***	0.111***	0.110***	0.126***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
F-statistic	8.000	1.503	3.988	9.137	0.696	2.915
R-Squared	0.003	0.001	0.001	0.003	0.000	0.001
Observations	8,853	8,795	8,839	8,853	8,795	8,839

Table 9: **Interaction of Image Motivation (canvassers) and Incentives**

This table shows the results Standard errors are given in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
$10 incentive	-0.013***	-0.009**	-0.016***	-0.016***	-0.001	-0.005
	(0.005)	(0.004)	(0.004)	(0.004)	(0.002)	(0.004)
$20 incentive	-0.015*	-0.006	-0.016**	-0.016**	-0.001	-0.004
	(0.008)	(0.007)	(0.008)	(0.008)	(0.004)	(0.006)
$30 incentive	-0.003	0.009	-0.001	-0.001	-0.001	-0.002
	(0.009)	(0.008)	(0.008)	(0.008)	(0.004)	(0.007)
Canvassers	0.336***	0.202***	0.254***	0.255***	-0.002	-0.006*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.002)	(0.004)
$10 incentive $\times$ Canvasser	-0.085***	-0.062***	-0.089***	-0.089***	0.004	0.009*
	(0.008)	(0.007)	(0.008)	(0.008)	(0.003)	(0.005)
Constant	0.167***	0.127***	0.151***	0.152***	0.004***	0.016***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
F-statistic	1136.524	470.575	645.384	647.569	0.550	0.775
R-Squared	0.094	0.041	0.056	0.056	0.001	0.001
Observations	54,767	54,767	54,767	54,767	8,031	9,721

Table 10: **Interaction of effort costs and image motivation (canvassers)**

This table shows

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
$1.5 \leq x < 2$ miles	0.048***	0.043***	0.033***	0.035***	0.007*	0.016**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.004)	(0.008)
$2 \leq x < 2.5$ miles	0.033**	0.028**	0.035**	0.034**	-0.000	0.011
	(0.014)	(0.013)	(0.014)	(0.014)	(0.005)	(0.009)
$2.5 \leq x < 3$ miles	0.018	-0.025**	0.014	0.013	-0.000	-0.012
	(0.013)	(0.012)	(0.013)	(0.013)	(0.005)	(0.009)
$3 \leq x < 3.5$ miles	0.005	-0.028**	0.006	0.006	-0.000	-0.001
	(0.014)	(0.013)	(0.014)	(0.014)	(0.006)	(0.009)
$3.5 \leq x < 4$ miles	-0.013	-0.046***	-0.013	-0.014	-0.000	-0.004
	(0.015)	(0.014)	(0.015)	(0.015)	(0.007)	(0.010)
$4 \leq x < 4.5$ miles	-0.003	-0.064***	-0.020	-0.021	-0.000	0.005
	(0.014)	(0.013)	(0.013)	(0.013)	(0.006)	(0.009)
$4.5 \leq x < 5$ miles	-0.027	-0.040**	-0.051**	-0.052**	-0.000	-0.012
	(0.021)	(0.019)	(0.020)	(0.020)	(0.009)	(0.017)
$5 \leq x < 5.5$ miles	-0.053***	-0.064***	-0.052***	-0.053***	-0.000	-0.012
	(0.018)	(0.017)	(0.018)	(0.018)	(0.009)	(0.014)
canvassing	0.061***	0.021**	0.030***	0.029***	-0.001	0.004
	(0.010)	(0.009)	(0.010)	(0.010)	(0.004)	(0.006)
$1.5 \leq x < 2$ miles and canvassing	-0.087***	-0.043***	-0.066***	-0.069***	-0.002	-0.018**
	(0.017)	(0.016)	(0.017)	(0.017)	(0.005)	(0.009)
$2 \leq x < 2.5$ miles and canvassing	-0.099***	-0.131***	-0.122***	-0.122***	-0.003	-0.016
	(0.022)	(0.021)	(0.022)	(0.022)	(0.007)	(0.011)
$2.5 \leq x < 3$ miles and canvassing	-0.027	-0.124***	-0.059*	-0.059*	0.036***	0.018
	(0.035)	(0.032)	(0.034)	(0.034)	(0.012)	(0.016)
$3 \leq x < 3.5$ miles and canvassing	-0.003	0.073**	-0.018	-0.002	-0.003	0.008
	(0.034)	(0.032)	(0.034)	(0.034)	(0.009)	(0.017)
$3.5 \leq x < 4$ miles and canvassing	0.029	0.279***	0.166**	0.166**	-0.003	0.042*
	(0.070)	(0.065)	(0.069)	(0.069)	(0.014)	(0.025)
$4 \leq x < 4.5$ miles and canvassing	-0.081	0.024	-0.056	-0.056	-0.003	-0.015
	(0.050)	(0.046)	(0.049)	(0.049)	(0.014)	(0.026)
$4.5 \leq x < 5$ miles and canvassing	0.161	-0.063	-0.084	-0.083	1.001***	0.996***
	(0.108)	(0.101)	(0.107)	(0.107)	(0.055)	(0.108)
$5 \leq x < 5.5$ miles and canvassing	0.180	0.196*	0.126	0.126	-0.003	0.001
	(0.111)	(0.104)	(0.110)	(0.110)	(0.026)	(0.050)
Constant	0.180***	0.144***	0.168***	0.168***	0.000	0.012**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.003)	(0.005)
F-statistic	330.616	155.201	192.086	192.791	20.224	5.935
R-Squared	0.195	0.102	0.123	0.124	0.071	0.017
Observations	24,601	24,601	24,601	24,601	4,775	6,107

Table 11: **Incentives and social identity**

This table shows the interaction social identity based on the majority of registered voters in a household being affiliated with a conservative party.

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
$10 incentive	0.038***	0.038***	-0.003	-0.004	-0.003	0.001
	(0.013)	(0.012)	(0.012)	(0.012)	(0.003)	(0.005)
$20 incentive	0.214***	0.250***	0.210***	0.208***	-0.006	-0.003
	(0.039)	(0.038)	(0.039)	(0.039)	(0.008)	(0.014)
$30 incentive	0.200***	0.279***	0.205***	0.203***	-0.006	-0.003
	(0.044)	(0.042)	(0.043)	(0.043)	(0.008)	(0.015)
Registered conservative	-0.671***	-0.481***	-0.606***	-0.608***	-0.004	0.005
	(0.010)	(0.009)	(0.010)	(0.010)	(0.002)	(0.004)
$10 incentive and registered conservative	-0.035***	-0.036***	0.005	0.005	0.005*	-0.002
	(0.013)	(0.013)	(0.013)	(0.013)	(0.003)	(0.005)
$20 incentive and registered conservative	-0.200***	-0.248***	-0.200***	-0.199***	0.007	0.011
	(0.040)	(0.038)	(0.039)	(0.039)	(0.009)	(0.015)
$30 incentive and registered conservative	-0.174***	-0.263***	-0.180***	-0.179***	0.007	0.013
	(0.044)	(0.042)	(0.043)	(0.043)	(0.009)	(0.016)
canvassing	0.136***	0.048***	0.052***	0.052***	0.001	0.009
	(0.007)	(0.007)	(0.007)	(0.007)	(0.004)	(0.006)
Constant	0.820***	0.588***	0.744***	0.746***	0.004	0.011***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.002)	(0.004)
F-statistic	1730.163	866.112	1223.648	1229.102	0.963	1.800
R-Squared	0.307	0.181	0.238	0.239	0.002	0.003
Observations	54,765	54,765	54,765	54,765	8,029	9,719

Table 12: **Interaction of social identity and framing**

This table shows interaction social identity based on the majority of registered voters in a household being affiliated with a conservative party

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
Economic	-0.052***	-0.072***	-0.016	-0.017	0.006	0.006
	(0.015)	(0.015)	(0.015)	(0.015)	(0.003)	(0.006)
Health	-0.064***	-0.112***	-0.028*	-0.032**	-0.002	0.005
	(0.015)	(0.015)	(0.015)	(0.015)	(0.003)	(0.006)
Positive tests	0.034	-0.069*	-0.072*	-0.074*	-0.005	-0.002
	(0.039)	(0.037)	(0.038)	(0.039)	(0.010)	(0.017)
Total tests	0.050	-0.033	-0.100**	-0.102**	0.016*	0.002
	(0.042)	(0.040)	(0.041)	(0.041)	(0.010)	(0.018)
Registered conservative	-0.726***	-0.564***	-0.628***	-0.631***	0.000	0.005
	(0.010)	(0.010)	(0.010)	(0.010)	(0.002)	(0.004)
Economic and registered conservative	0.046***	0.068***	0.009	0.010	-0.006	-0.008
	(0.016)	(0.015)	(0.015)	(0.015)	(0.004)	(0.006)
Health and registered conservative	0.067***	0.118***	0.032**	0.035**	0.006	0.005
	(0.016)	(0.015)	(0.015)	(0.015)	(0.004)	(0.006)
Positive tests and registered conservative	-0.050	0.055	0.066*	0.068*	0.007	-0.002
	(0.040)	(0.038)	(0.039)	(0.039)	(0.010)	(0.018)
Total tests and registered conservative d	-0.072*	0.006	0.076*	0.079*	-0.021**	-0.002
	(0.042)	(0.040)	(0.041)	(0.041)	(0.011)	(0.019)
canvassing	0.145***	0.061***	0.064***	0.064***	0.001	0.010
	(0.011)	(0.010)	(0.010)	(0.010)	(0.005)	(0.009)
Constant	0.877***	0.671***	0.767***	0.770***	0.000	0.008*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.003)	(0.004)
F-statistic	1512.028	756.801	1067.265	1072.124	1.478	2.277
R-Squared	0.306	0.181	0.238	0.239	0.003	0.004
Observations	54,765	54,765	54,765	54,765	8,029	9,719

Table 13: **Interaction of social identity and image motivation**

This table shows interaction of social identity and image motivation

	Survey	Test			Positive	
		Viral	Antibody	Any	Viral	Antibody
	(1)	(2)	(3)	(4)	(5)	(6)
Canvassing	-0.142***	-0.219***	-0.264***	-0.262***	-0.001	0.001
	(0.013)	(0.013)	(0.013)	(0.013)	(0.004)	(0.007)
Registered conservative	-0.863***	-0.668***	-0.789***	-0.790***	-0.000	0.005
	(0.009)	(0.009)	(0.009)	(0.009)	(0.002)	(0.004)
Canvassing and registered conservative	0.294***	0.282***	0.332***	0.330***	0.002	-0.001
	(0.012)	(0.012)	(0.012)	(0.012)	(0.003)	(0.005)
Constant	1.004***	0.767***	0.918***	0.919***	0.002	0.011***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.002)	(0.004)
F-statistic	2868.779	1473.944	2059.128	2070.006	1.770	1.844
R-Squared	0.312	0.189	0.245	0.246	0.002	0.002
Observations	57,057	57,057	57,057	57,057	8,430	10,139