Call of duty: the effects of phone calls on blood donor motivation

Adrian Bruhin,[1] Lorenz Goette,[1] Adrian Roethlisberger,[2] Alexander Markovic,[2] Regula Buchli,[2] and Beat M. Frey[2]

BACKGROUND: Little is known about the long-term effects of interventions aimed at increasing turnout among voluntary blood donors.

STUDY DESIGN AND METHODS: We use a retrospective natural experiment with all 40,653 donors who were repeatedly invited to blood drives in Zurich, Switzerland, between 2010 and 2013. The intervention is a quasi-randomized phone call informing donors of a current shortage of their blood type. The panel structure of the data allows identification of different types of donors reacting to the phone call.

RESULTS: Our analysis reveals two types. Type 1 donors make up 27.1% of the population. They are highly motivated and exhibit a baseline donation rate of 59.4% ($p < 0.001$). The phone call raises their probability to donate by 9.9% at the upcoming blood drive ($p < 0.001$). However, the phone call reduces their donation rate by 2.3% ($p = 0.003$) at each future blood drive. In contrast, the 72.9% of Type 2 donors exhibit a low baseline donation rate of 5.8% ($p < 0.001$). The phone call raises their probability to donate by 5.8% at the upcoming blood drive ($p < 0.001$). Moreover, the phone call leads to habit formation in Type 2 donors and increases their donation rate by 2.1% at the next blood drive ($p = 0.03$).

CONCLUSION: Behavioral interventions are effective at increasing donation rates in the short run. However, they can crowd out the intrinsic motivation of the most motivated donors. Thus, blood donation services should avoid interventions for highly motivated donors and target them at irregular donors. Our results also sound a warning on using other interventions.

Most developed countries rely on voluntary blood donations to ensure a sufficient stock of whole blood transfusions. The system of voluntary blood donations works surprisingly well for much of the time, but blood donation services often struggle to meet a fluctuating demand. In response to these struggles, blood donation services have started using a variety of interventions to increase donation rates. They range from offering rewards such as gift vouchers, prepaid debit cards, or T-shirts and publicly thanking donors on Web pages or posters to changes in the message appealing to donors.[1-7]

Research from economics and psychology suggests that the effects of behavioral interventions on donation rates could vary both over time and across donors. While the immediate effect of behavioral interventions is often positive, their lagged and long-term effects are ambiguous: in the medium run, donors may form a habit of giving blood.[8] An initial increase in donation rates caused by an intervention may thus carry over to future periods, as has been observed for other forms of prosocial behavior.[9] However, if donors feel less obliged after having just donated, the initial increase may be followed by a

ABBREVIATION: BTSRC = Blood Transfusion Service of the Swiss Red Cross; FMM = finite mixture model.

From the [1]Faculty of Business and Economics (HEC Lausanne), University of Lausanne, Lausanne-Dorigny, Switzerland; and [2]Blood Transfusion Service, Swiss Red Cross, Schlieren, Switzerland.

Address reprint requests to: Lorenz Goette, University of Lausanne, Bâtiment Internef 536, 1015 Lausanne, Switzerland; e-mail: lorenz.goette@unil.ch.

This study was partly funded by the Swiss National Science Foundation (Grant 141767).

Received for publication January 20, 2015; revision received May 13, 2015; and accepted May 13, 2015.

doi:10.1111/trf.13236

TRANSFUSION 2015;55;2645–2652

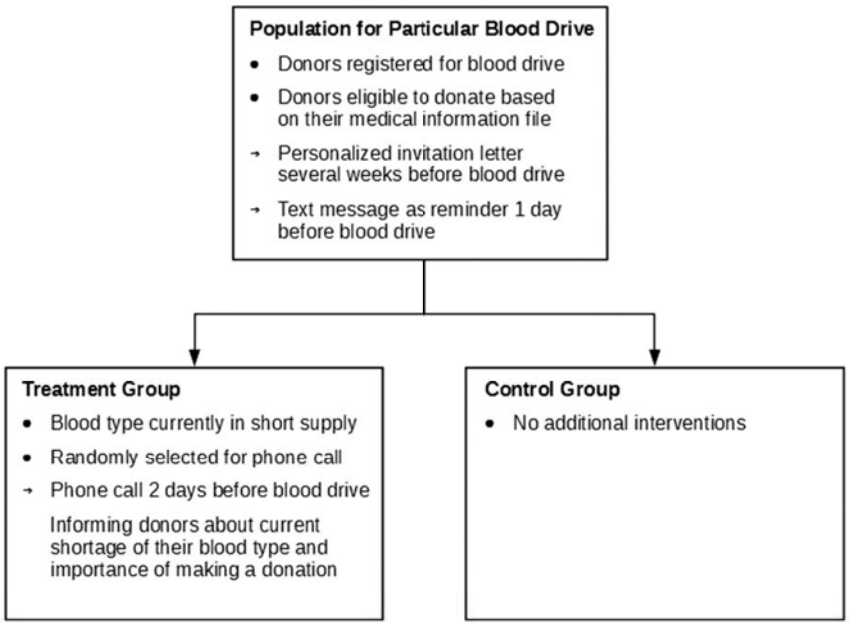

Fig. 1. Intervention procedure for a particular blood drive. The schema illustrates the intervention procedure for a particular blood drive. All registered and eligible donors receive a personalized invitation letter several weeks before the blood drive plus a text message reminder 1 day before the blood drive.

temporary decrease in future donation rates, as donors may target a certain level of prosocial activity per year.[10]

Psychological theories even raise the possibility that interventions could crowd out the donors' intrinsic motivation in the long term.[11-13] Prosociality is an important motivator for blood donors.[14,15] Psychological theories of intrinsic motivation raise the specter that any intervention by blood donation services may decrease these prosocial motivations.[12,16] Moreover, economic research has revealed substantial individual heterogeneity in prosocial motivations[17] and similar results have been found for blood donations.[14] This implies that the effects of interventions on donation rates could also differ for individuals with different strengths of prosocial motivations.

To investigate how behavioral interventions affect donation rates over time and across donors with differing motivation, we exploit data from a retrospective natural experiment comprising the universe of blood donors registered for blood drives in the canton of Zurich, Switzerland.

MATERIALS AND METHODS

Empirical setup

We study active blood donors who donated at least once before the onset of the study and were repeatedly invited to blood drives in the canton of Zurich. While the blood drives were coordinated by local organizations, such as local church chapters or sports clubs, the Blood Transfusion Service of the Swiss Red Cross (BTSRC) in Zurich, Switzerland, administered the invitation to donors and provided the equipment and personnel to take the blood. Our sample includes all 40,653 active blood donors, who received personalized invitations to a mean (±SD) of 3.10 (±1.26) blood drives. This covers overall 570 blood drives that took place over the sample period from January 2010 to January 2013 (for descriptive statistics of the data set see Table S1, available as supporting information in the online version of this paper). We only observe donation outcomes, age, and sex in the data. Medical data, in particular, data related to complications during transfusions, are not contained in our data set due to physician–patient privilege.

Behavioral intervention

The behavioral intervention in our study is a phone call to some donors made by the BTSRC 2 days before the blood drive. Figure 1 schematically illustrates the intervention procedure for a particular blood drive. Depending on the daily inventory in its blood stock, the BTSRC determines which blood group, A–, A+, O–, or O+, is in short supply, and puts a random subset of invited donors with matching blood type on a call list 2 days ahead of the blood drive. All blood donors, also those who were not on the call list, receive a text message 1 day ahead of the blood drive, reminding them of the event.

The BTSRC confines the phone calls to blood groups A–, A+, O–, or O+, as they are often in short supply. The phone call points out to the donors that their blood type is currently in short supply and that the benefits to society

of their donation would now be particularly high. The phone call, done in free form by administrative assistants, conveys the following generic message: "Your blood type *X* is in short supply. Thus, it is important that you come and donate at the upcoming blood drive you have been invited for." The intention of the phone call is to signal the scarcity of one's blood type on that particular day. Thus, it is meant to enhance a donor's sense of altruism at that time, but it is also possible that it affects a donor's feeling of self-determination negatively.[16] The BTSRC introduced the phone call and decided to analyze its impact prospectively after similar calls were used as a stopgap measure earlier on.

The BTSRC has the capacity to make roughly 100 phone calls per day and uses a software tool that randomly selects donors from all invited donors with the desired blood type. Across days, the phone calls are driven by random fluctuations in the demand for a certain blood type; therefore, they should be random conditional on blood types. We formally verified the conditional randomization of the phone calls (for details see Table S2, available as supporting information in the online version of this paper).

In total, 15,962 donors were called. Of those, 43% were reached personally and the administrative assistants conferred the above message. If it was not possible to reach them despite multiple attempts, the BTSRC left a message on the mailbox (in 49% of the cases). For 8% of the donors, it was not possible to leave a message. In the empirical analysis below, we adopt a strict "intention-to-treat" approach and consider all individuals for whom call attempts were made as members of the treatment group.[18]

Identification of treatment effects

The BTSRC has applied this intervention for 3 years and fluctuations in the inventory of transfusions of a given blood type create temporal variation in the exposure to the phone call for a given donor. This temporal variation allows us to identify not only the immediate impact of a phone call on donation rates, but also how it affects subsequent decisions to donate, up to 3 years later. In particular, we obtain the following three treatment variables.

First, comparing the donation rates of the donors in the treatment group, that is, those who received a phone call for an upcoming blood drive, to those in the control group, that is, those who did not receive a phone call, identifies the immediate treatment effect. We hypothesize this immediate treatment effect to be positive. As the phone call mentions that the donor's blood type is currently in short supply, it makes the altruistic benefit of donating more salient.[19]

Second, differences in current donation rates between donors who were in the treatment group at their previous invitation compared to donors who were in the control group at their previous invitation identifies the lagged treatment effect. If the phone call is habit-forming, it may lead to a higher donation frequency even at the subsequent invitation. However, it is also possible that the previous donation leads to a reduced perceived obligation to donate and thus lower donation rates at the next invitation.[7,10]

Third, differences in current donation rates as a function of differences in the cumulative number of phone calls received, that is, the number of times a donor was in the treatment group, identifies the long-term treatment effect. Several psychological theories predict that if some intervention is perceived as controlling, it will crowd out the donors' intrinsic motivation in the long term because it may undermine a donor's sense of self-determination.[11-13] If exposure to phone calls lowers donation rates on all subsequent invitations, this would be strongly suggestive of such a crowding-out effect.

Statistical analysis

In our statistical analysis, we proceed in two steps. In a first step, we estimate the average treatment effects of the phone call by relating the decision to donate at each invitation to the three treatment variables mentioned. In particular, we apply a linear probability model in which each time a donor was invited to a blood drive, the dependent variable indicates whether the donor showed up at the that blood drive or not. We regress this indicator on the three treatment variables simultaneously to identify the different effects: 1) the variable whether the donor received a phone call for the upcoming blood drive measures the immediate treatment effect, 2) the variable whether the donor got a phone call for the previous blood drive captures the lagged treatment effect, and 3) the cumulative number of phone calls the donor received identifies the long-term treatment effect (for formal details see the supporting information, available in the online version of this paper). The estimated coefficients can directly be interpreted as changes in the probability to donate. The linear probability model also allows for donor-specific baseline donation rates by including individual fixed effects. Note that allowing for donor-specific baseline donation rates directly controls for any time-invariant donor-specific characteristics, such as unobservable differences in prosocial motivations, but also blood types, sex, and the number of donations made before receiving the first invitation. To increase statistical precision, we additionally control for the donors' age and blood drive fixed effects.

We also perform an analysis of heterogeneity in the treatment effects commonly used in earlier work.[6,20] In particular, we examine whether there are differences in treatment effects with respect to observable

characteristics, such as age, sex, and motivation of the blood donor as measured by the number of donations in the year before entering the study. To do so, we first add potential interactions between each of these observable characteristics and the three treatment variables as additional regressors to the linear probability model. Subsequently, we perform F tests of the null hypothesis that none of these interactions are significant.

In a second step of the statistical analysis, we turn to a more sophisticated, but realistic, way of modeling behavioral heterogeneity. We apply a finite mixture model (FMM) that exploits the panel structure of the data, that is, the fact that donors were invited multiple times (mean ± SD, 3.10 ± 1.26 times), to identify distinct types of donors that respond differently to the phone call.[21-23] The FMM allows us to pick up latent heterogeneity in the data that is not related to any observable characteristics, without having to prespecify which donors are of which type. It assumes the population of donors to be made up by a finite number of distinct types and estimates their type-specific responses to the phone call. This also allows us to classify each donor into the type he or she most likely stems from based on his or her individual probabilities of type membership. They correspond to the relative fit of the donor's behavior to the behavior implied by the estimated variables of the different types and thus enable us to see how clearly the donor's behavior fits to the predicted pattern (see Equation 14 in the supporting information). In contrast to the linear probability model, the FMM does not include individual fixed effects as it estimates type-specific instead of donor-specific baseline donation rates. Consequently, the FMM uses a wider set of control variables including blood types, sex, age, the number of donations made before receiving the first invitation, and blood drive fixed effects. For both models, the linear probability model and the FMM, we report individual cluster robust standard errors that take heteroskedasticity and potentially correlated decisions to donate into account.[24,25]

To determine the optimal number of types specified in the FMM, we perform a cross-validation[26] (for further details see the supporting information). We randomly partitioned the data 100 times into a training and a test sample. We estimated specifications of the FMM that differ in the number of types in the training sample and assessed their out-of-sample fit in the test sample. A FMM with too few types lacks the flexibility to capture the relevant behavioral heterogeneity and will fit poorly in the test sample because of that reason. By contrast, a FMM with too many types overfits the data and the parameter estimates for the nonexistent types are driven by random noise in the data of the training sample. For this reason, a FMM with too many types will also fit poorly in the test sample. The logic behind cross-validation is to pick the model specification with the best out-of-sample fit over a large number of randomly drawn test samples.

We evaluate the FMM's robustness and ability to capture heterogeneity by splitting the sample according to the donors' individual probabilities of type membership into type-specific subsamples. Subsequently, we estimate a linear probability model separately in each of these type-specific subsamples. If the FMM is robust, its estimates are close to the ones of the two linear probability models in the type-specific subsamples. Moreover, if the FMM captures the essential part of the individual heterogeneity, baseline donation rates do not vary across donors within the type-specific subsamples.

RESULTS

Mean treatment effects

Table 1 shows the estimated coefficients of the linear probability model. The mean baseline donation rate, that is, the donors' propensity to give blood without receiving any phone calls, amounts to 30.6% ($p < 0.001$). The immediate treatment effect of the phone call is positive and increases the mean donation rate by 7.8 percentage points ($p < 0.001$). Moreover, the coefficient on the lagged treatment effect indicates that a donor is on average 2.2 percentage points more likely to donate at the current blood drive if she was in the treatment group at the previous invitation ($p = 0.008$). We interpret this positive lagged treatment effect as evidence for habit formation. The coefficient of the long-term treatment effect is both small in magnitude and insignificant ($p = 0.74$). Thus, on average, there is no evidence that the phone call crowds out the donors' intrinsic motivation. However, an F test for individual fixed effects reveals that baseline donation rates vary substantially across donors ($p < 0.001$). This heterogeneity does not seem to be related to the treatment effects: in particular, the specification in Column 4 of Table 1 interacts the treatment effects with the number of donations in the year before the study. There is no evidence that the treatment effects vary across donors according to their baseline donation rates. The only evidence of heterogeneity we find is in the long-term treatment effect: it appears to be weakly negative for women ($p = 0.029$), and shows a significant interaction with men, for whom it appears mildly positive.

Type-specific treatment effects

We next report the results of the FMM that identifies type-specific treatment effects that may not be related to observable characteristics. The cross-validation results in Table 2 show that the FMM with two distinct types of donors is optimal because it exhibits the best out-of-sample fit as measured by the cross-validated log likelihood. Figure 2 visually confirms that this specification

TABLE 1. Mean treatment effects (TE) of the phone call on donation rates*

Dependent variable: donation at upcoming blood drive	Treatment effect interacted with			
	Column 1: no interaction	Column 2: age	Column 3: male	Column 4: number of prior donations
Baseline donation rate	0.306† (0.001)	0.306† (0.001)	0.306† (0.001)	0.306† (0.001)
Immediate TE	0.078† (0.008)	0.078† (0.008)	0.069† (0.011)	0.074† (0.011)
Lagged TE	0.022† (0.008)	0.021† (0.008)	0.016 (0.011)	0.016 (0.011)
Long-term TE	−0.002 (0.005)	−0.001 (0.005)	−0.015‡ (0.007)	−0.004 (0.007)
Immediate TE × interaction		−0.001 (0.001)	0.016 (0.015)	0.003 (0.009)
Lagged TE × interaction		0.000 (0.001)	0.011 (0.015)	0.006 (0.009)
Long-term TE × interaction		−0.000 (0.000)	0.024† (0.010)	0.003 (0.005)
Age	−0.011† (0.002)	−0.010† (0.002)	−0.011† (0.002)	−0.010† (0.002)
F test H0 p value: no interaction of TEs with characteristics		0.69	0.003	0.69
F test H0 p value: no individual-specific variation in baseline donation rates	<0.001	<0.001	<0.001	<0.001
Number of observations	126,123	126,123	126,123	126,123
Number of donors	40,653	40,653	40,653	40,653

* The TEs of the phone call on donation rates are estimated using a linear probability model (see Equations 3 and 4 in the supporting information). In all four specifications of the model, the dependent variable is a binary indicator whether a donation was made at the upcoming blood drive or not. The control variable age is centered on the sample mean of 43.4 years. The specifications in Columns 2 to 4 perform commonly used subgroup analyses by interacting the TEs with observable characteristics (age, sex, and donations prior to the study). All specifications additionally control for blood drive and individual fixed effects. Individual cluster robust standard errors are reported in parentheses. Level of significance (t test with H0: coefficient is zero).
† $p < 0.01$.
‡ $p < 0.05$.

TABLE 2. Out-of-sample fit for determining the optimal number of distinct types of donors*

Relative size of training sample	Number of types		
	1	2	3
1/4	−133,430.33	−63,091.52	−63,720.89
1/3	−87,590.25	−48,398.05	−48,504.55
1/2	−46,344.35	−33,539.30	−33,226.99

* FMM's out-of-sample fit, measured by the cross-validated log likelihood that must be maximized to determine the optimal number of types (see supporting information for further details). The approximation of the cross-validated log likelihood relies on the following procedure: First, the data set is randomly split into a test and a training sample. Second, the coefficients of the FMM are estimated in the training sample and, third, used to obtain the log likelihood of the model in the test sample. This procedure was repeated 100 times for each relative size of the training sample. The reported values represent the mean log likelihood obtained in the test sample. The optimal number of types is 2 as the corresponding FMM consistently achieves the highest cross-validated log likelihood for each relative size of the training sample.

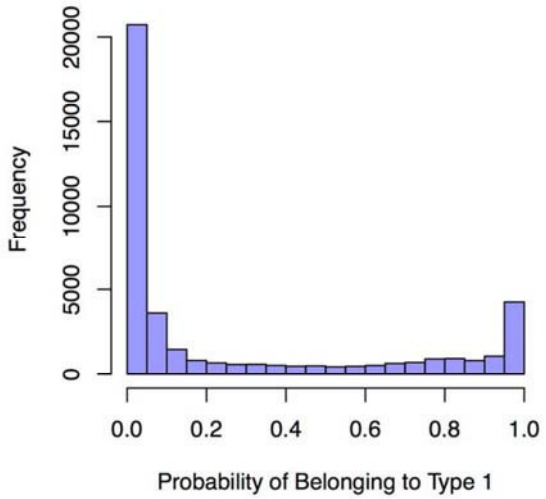

Fig. 2. Unambiguous classification of donors into types. The histogram displays the distribution of the individual probabilities of belonging to the first type of donors based on the FMM (see Equation 14 in the supporting information). The FMM discriminates well between the two types of donors, as almost all donors are unambiguously classified into either the first or the second type.

discriminates well between the two types, as it classifies almost all donors unambiguously into either the first or the second type.

Table 3 shows how the two types of donors differ in their reactions to the phone call by presenting the estimated coefficients of the FMM. Type 1 donors make up 27.1% of the population (95% confidence interval [CI], 26.5%-27.8%]. They exhibit a high baseline donation rate of 59.4% ($p < 0.001$). The phone call has a positive immediate effect of 9.9 percentage points on their probability to donate in the upcoming blood drive ($p < 0.001$). There is no evidence for habit formation for this type of donor as the coefficient of the lagged treatment effect is small and insignificant ($p = 0.71$). However, the phone call has a long-term treatment effect: each additional phone call

TABLE 3. Type-specific treatment effects (TE) of the phone call on donation rates*

Dependent variable: donation at upcoming blood drive	Type 1 donors	Type 2 donors
Type-specific coefficients		
Share among the population	0.271† (0.003)	0.729† (0.003)
Baseline donation rate	0.594† (0.007)	0.058† (0.004)
Immediate TE	0.099† (0.014)	0.058† (0.008)
Lagged TE	0.008 (0.021)	0.021‡ (0.010)
Long-term TE	−0.023† (0.008)	−0.002 (0.003)
Common coefficients		
Blood group O–	−0.003 (0.008)	
Blood group A+	−0.002 (0.003)	
Blood group A–	0.004 (0.006)	
Male	0.028† (0.002)	
Age	0.003† (0.000)	
Number of donations in the year before the study		
One	0.128† (0.003)	
Two	0.251† (0.005)	
Three	0.325† (0.010)	
Four	0.208† (0.035)	
Five	0.016 (0.066)	
SD of error term	0.366† (0.001)	
Number of observations	126,123	
Number of donors	40,653	

* Type-specific TEs of the phone call on donation rates are estimated using a FMM with two types (see Equation 5 in the supporting information). The dependent variable is a binary indicator whether a donation was made at the upcoming blood drive or not. The control variable age is centered on the sample mean of 43.4 years. The model additionally controls for blood drive fixed effects. Individual cluster robust standard errors are reported in parentheses. Level of significance (t test with H0: coefficient is zero).
† $p < 0.01$.
‡ $p < 0.05$.

reduces the donation rate of Type 1 donors by 2.3 percentage points ($p = 0.003$). Consequently, after receiving a phone call the predicted probability that an average Type 1 donor shows up is 59.4% + 9.9% − 2.3% = 67.0% at the upcoming blood drive, but only 59.4% − 2.3% = 57.1% at every future blood drive.

Type 2 donors account for 72.9% of the population (95% CI, 72.2%-73.5%). Their baseline donation rate is low and amounts to just 5.8% ($p < 0.001$). The phone call also exerts a positive immediate effect of 5.8 percentage points for this type of donor ($p < 0.001$). However, in contrast to the Type 1 donors, Type 2 donors are habit formers: The coefficient of the lagged treatment effect of the phone call indicates a 2.1-percentage-point increase in the donation rate if a Type 2 donor received a phone call for the previous blood drive ($p = 0.03$). Moreover, there is no evidence of a long-term treatment effect since the corresponding coefficient is small and insignificant ($p = 0.514$).

Besides the impact of the phone call on donation rates, it is also interesting to examine the impact of donor characteristics: The common coefficients in Table 3 indicate that there is no significant relationship between blood types and donation rates. However, male donors are on average 2.8 percentage points more likely to donate upon an invitation than female donors ($p < 0.001$). Each additional year of age is on average associated with a 0.3-percentage-point higher donation rate ($p < 0.001$). The relationship between donation history in the year before the study and donation rates during the study is positive for up to four previous donations ($p < 0.001$ for all coefficients) and inconclusive for donors with more than four donations per year ($p = 0.81$).

Finally, we evaluated the FMM's robustness and ability to capture heterogeneity. Table 4 shows the estimated treatment effects in two separate subsamples containing only Type 1 and Type 2 donors, respectively. The estimates in Table 4 confirm the robustness of the FMM, as they by and large coincide with the estimates in Table 3. Furthermore, within the two type-specific subsamples, baseline donation rates do not vary across donors (F tests for individual fixed effects: $p = 0.56$ for Type 1 donors and $p = 0.92$ for Type 2 donors). Thus, the FMM captures the essential part of the individual heterogeneity, both with respect to the response to the phone call and with respect to differences in the baseline donation rate.

DISCUSSION

Our results show that for both types of donors, a simple phone call can raise the propensity to give blood on impact. However, there is important heterogeneity in how the phone call affects donor motivation over time.

For Type 1 donors we find the hallmark signs of crowding out of intrinsic motivation: it is the combination of a high baseline motivation and a long-term reduction in motivation after the behavioral intervention that is characteristic for motivational crowding.[11,12] This finding is important, since it shows that even a "mild" intervention like a phone call can have a sizable crowding-out effect on intrinsic motivation. We have also explored in more detail the specific functional form of the crowding-out effect. Since we observe individuals who receive multiple phone calls, we can distinguish whether there is just a one-time reduction in the motivation after the first phone call or whether the effect scales with the incidence of phone calls. Our evidence favors the model in which the crowding-out effect scales with the number of phone calls (for details see Table S3, available as supporting information in the online version of this paper). Our results also leave little doubt that the reduction in the motivation is long term and extends over at least 1 year, as the lagged treatment effect of a phone call (6 months ago) is not significant for Type 1 donors. Overall, this finding sounds a cautionary note with regard to using other behavioral interventions in untargeted ways on blood donors, as they

TABLE 4. Treatment effects (TEs) of the phone call on donation rates in separate type-specific subsamples*

Dependent variable: donation at upcoming blood drive	Type 1 donors	Type 2 donors
Baseline donation rate	0.630† (0.007)	0.045† (0.004)
Immediate TE	0.096† (0.011)	0.059† (0.007)
Lagged TE	0.005 (0.014)	0.030† (0.008)
Long-term TE	−0.026† (0.005)	−0.002 (0.003)
Blood group O−	0.001 (0.011)	−0.013‡ (0.007)
Blood group A+	−0.001 (0.004)	−0.001 (0.002)
Blood group A−	0.009 (0.007)	0.004 (0.005)
Male	0.025† (0.003)	0.025† (0.002)
Age	0.002† (0.000)	0.002† (0.000)
Number of donations in the year prior to the study		
One	0.129† (0.005)	0.106† (0.002)
Two	0.220† (0.005)	0.234† (0.004)
Three	0.275† (0.007)	0.309† (0.009)
Four	0.248† (0.030)	0.144† (0.031)
Five	−0.055† (0.010)	0.089 (0.107)
F test H0 p value: no individual-specific variation in baseline donation rates	0.54	0.89
Number of observations	32,579	93,544
Number of donors	10,699	29,954

* The two columns show the TEs of the phone call on donation rates in two separate type-specific subsamples, estimated with linear probability models. The type-specific subsamples are obtained by classifying the donors according to their individual probabilities of type-membership from the FMM (see Equation 14 in the supporting information). The dependent variable is a binary indicator whether a donation was made at the upcoming blood drive or not. The control variable age is centered on the sample mean of 43.4 years. The linear probability models additionally control for blood drive fixed effects. Individual cluster robust standard errors are reported in parentheses. Level of significance (t test with H0: coefficient is zero).
† $p < 0.01$.
‡ $p < 0.05$.

may permanently damage the motivation of the (previously) most reliable donors. While aggregate studies reveal no such effects, our results show how important it is to take behavioral heterogeneity into account.[5]

One might argue that the negative response to previous multiple phone calls by Type 1 donors may be due to the induced past donations. Notice, however, that the phone call has approximately the same immediate effect on donations for Type 1 and Type 2 donors. Therefore, multiple phone calls induce approximately the same change in the donation history of both types of donors. If donors were simply reacting to different past induced patterns, the response to a given number of previous phone calls should then be the same for Type 1 and Type 2 donors. However, we observe a sharply different pattern for the two. This rules out that argument.

By contrast, Type 2 donors display a substantial habit-forming effect: more than one-third of the immediate treatment effect carries over into a higher donation probability on the subsequent invitation, even if they receive no additional phone call. This result is consistent with a growing literature showing the importance of habit formation in prosocial behavior and other domains.[9,27,28]

Blood donation services should take this heterogeneity into account. They should exclusively target only irregular Type 2 donors with interventions and leave the highly motivated Type 1 donors alone. For Type 1 donors the phone call increases the immediate donation rate at the expense of a lower future donation rate. Calculated over 2 years, a phone call increases the overall donation rate by just 1.5 percentage points for a Type 1 donor. By contrast, for Type 2 donors, it not only increases the immediate donation rate but also leads to a higher future donation rate. Given our estimates, the total impact of a one-time call is 7 percentage points for Type 2 donors over the same 2-year period. Consequently, exclusively targeting Type 2 donors is much more cost-effective.

While a full identification of the donors' types would require the donation service to estimate a FMM, our calculations show that a simple heuristic can help identifying Type 2 donors: given the much lower baseline donation rate of Type 2 donors, the probability that an individual who has not donated in response to the last two invitations is a Type 2 donor is 0.935, thus virtually as good as the full identification based on the FMM.

More broadly, our results also have important implications for theories of prosocial motivations and public policy. They are consonant with the heterogeneity in blood donor motivations found in survey studies,[14,15] even though we use a completely different method. Our results reveal that even an intervention as subtle as a phone call can undermine the motivation in the most reliable group of donors. Moreover, they show that in contrast to what is assumed in most domains of behavioral science, interventions can change deeply rooted prosocial motivations.[29,30] These results raise the specter that other interventions, such as monetary or nonmonetary incentives, may have a negative long-term effect on the intrinsic motivation of the most dependable prosocial contributors. This does, of course, not imply that all interventions aimed at highly motivated donors may backfire. Previous research has been unable to identify interventions that work better for highly motivated donors.[5] Future research, using a statistical framework more suitable to detecting heterogeneity, such as ours, should address this question.

ACKNOWLEDGMENTS

AB and LG had full access to all of the data and take responsibility for the accuracy of the data analysis. We are grateful to three anonymous referees who provided various comments that greatly helped us to improve the manuscript.

CONFLICT OF INTEREST

The authors have disclosed no conflicts of interest.

REFERENCES

1. Ferrari JR, Barone RC, Jason LA, et al. The use of incentives to increase blood donations. J Soc Psychol 1985; 125:791-3.
2. Godin G, Vézina-Im LA, Bélanger-Gravel A, et al. Efficacy of interventions promoting blood donation: a systematic review. Transfus Med Rev 2012;26:224-37.
3. Reich P, Roberts P, Laabs N, et al. A randomized trial of blood donor recruitment strategies. Transfusion 2006;46:1090-6.
4. Lacetera N, Macis M. Do all material incentives for pro-social activities backfire? The response to cash and non-cash incentives for blood donations. J Econ Psychol 2010;31:738-48.
5. Lacetera N, Macis M, Slonim R. Economic rewards to motivate blood donations. Science 2013;340:927-8.
6. Lacetera N, Macis M, Slonim R. Rewarding volunteers: a field experiment. Manage Sci 2014;60:1107-29.
7. Goette L, Stutzer A. Blood donations and incentives: evidence from a field experiment. In: EconPapers. Discussion Paper No. 3580, IZA Discussion Papers. Bonn: Institute for the Study of Labor (IZA); 2008.
8. Aarts H, Dijksterhuis AP. Habits as knowledge structures: automaticity in goal-directed behavior. J Pers Soc Psychol 2000;78:53-63.
9. Meer J. The habit of giving. Econ Inq 2013;51:2002-17.
10. Meier S. Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. J Eur Econ Assoc 2007;5:1203-22.
11. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull 1999;125:627.
12. Frey B. Not just for the money. An economic theory of personal motivation. Cheltenham (UK): Edward Elgar Publishing; 1997.
13. Frey B, Oberholzer-Gee F. The cost of price incentives: an empirical analysis of motivation crowding-out. Am Econ Rev 1997;87:746-55.
14. Ferguson E, Atsma F, de Kort W, et al. Exploring the pattern of blood donor beliefs in first-time, novice, and experienced donors: differentiating reluctant altruism, pure altruism, impure altruism, and warm glow. Transfusion 2012;52: 343-55.
15. Glynn SA, Williams AE, Nass CC, et al. Attitudes toward blood donation incentives in the United States: implications for donor recruitment. Transfusion 2003;43:7-16.
16. Deci EL, Ryan RM. Intrinsic motivation and self-determination in human behavior. Berlin: Springer Science & Business Media; 1985.
17. Fehr E, Fischbacher U. The nature of human altruism. Nature 2003;425:785-91.
18. Angrist JD, Pischke JS. Mostly harmless econometrics: an empiricist's companion. Princeton (NJ): Princeton University Press; 2012.
19. Bordalo P, Gennaioli N, Shleifer A. Salience theory of choice under risk. Q J Econ 2012;27:1243-85.
20. Goette L, Stutzer A, Yavuzcan G, et al. Free cholesterol testing as a motivation device in blood donations: evidence from field experiments. Transfusion 2009;49: 524-31.
21. McLachlan G, Peel D. Finite mixture models. Wiley Series in Probabilities and Statistics. Hoboken (NJ): John Wiley & Sons; 2000.
22. Houser D, Keane M, McCabe K. Behavior in a dynamic decision problem: an analysis of experimental evidence using a Bayesian type classification algorithm. Econometrica 2004; 72:781-822.
23. Bruhin A, Fehr-Duda H, Epper T. Risk and rationality: uncovering heterogeneity in probability distortion. Econometrica 2010;78:1375-412.
24. Huber PJ. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Berkeley (CA): University of California Press; 1967.
25. Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge: MIT Press; 2002.
26. Smyth P. Model selection for probabilistic clustering using cross-validated likelihood. Stat Comput 2000;10: 63-72.
27. Charness G, Gneezy U. Incentives to exercise. Econometrica 2009;77:909-31.
28. Maréchal K. Not irrational but habitual: the importance of "behavioural lock-in" in energy consumption. Ecol Econ 2010;69:1104-14.
29. Bowles S. Endogenous preferences: the cultural consequences of markets and other economic institutions. J Econ Lit 1998;36:75-111.
30. Fehr E, Hoff K. Introduction: tastes, castes and culture: the influence of society on preferences. Econ J 2011;121:F396-F412.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1.

Supporting Information

Call of Duty: The Effects of Phone Calls on Blood Donor Motivation

Adrian Bruhin, PhD[1]; Lorenz Goette, PhD[1]; Adrian Roethlisberger[2]; Alexander Markovic, MD[2]; Regula Buchli[2]; Beat M. Frey, MD[2]

[1]University of Lausanne, Faculty of Business and Economics, Department of Economics, 1015 Lausanne-Dorigny, Switzerland

[2]Blood Transfusion Service, Swiss Red Cross, Rütistrasse 19, 8952 Schlieren, Switzerland

Contents

1 Data

This section provides additional information on the data and verifies the conditional randomization of the phone call.

1.1 Additional information on the data set

The data set consists of the universe of all blood donors who had donated at a local blood drive with the BTSRC and have blood type A or O. The BTSRC did not include blood types B and AB in the design because of their limited usefulness. Blood donors are regularly invited to donate at the blood drive at which they have donated before. Only donors who, according to the medical information on file, are cleared to donate blood are invited to a blood drive. Each invitation to a blood drive constitutes one observation in our data set, as it represents an opportunity to donate blood. Our measure of donations indicates whether a donor showed up at a blood drive, irrespective of whether a blood transfusion was successfuly performed or not.

Table S1 provides a basic picture of the characteristics of the donors in the sample. Their average age is 43 years, and there are somewhat more men in the sample than women. In the 12 months before the donors were added to the intervention, they donated on average 0.9 times. The average propensity to donate upon receiving an invitation was 31 percent. The bottom half of table S1 shows the breakdown of the different blood types in the study. Not surprisingly, negative blood types are much less frequent than positive blood types. The phone calls are mostly concentrated on negative blood types. We include the positive blood types nevertheless, as they help us identify the coefficients of the control variables and thus add to the precision of the model. Furthermore, positive blood types are sometimes in short supply as well and treated with the phone call.

Individual characteristics	Mean	Std. Dev.
Age	43.38	13.26
Fraction male	0.57	0.49
Fraction donated on invitation	0.31	0.46
Number of donations in 12 months prior to entering the study design	0.90	0.81
Frequency of blood types	whole sample	treatment group
O+	0.40	0.08
O-	0.09	0.65
A+	0.43	0.03
A-	0.08	0.24

Table S1: Descriptive statistics. The table summarizes the individual characteristics of the 40,653 donors in the study. Only donors with blood types A and O were used in the study. Overall, 126,123 invitations were sent out.

1.2 Checks of the conditional randomization of the phone call

In this section, we check the conditional randomization of the phone call by testing whether the current phone call is correlated with *previous* donations. The software tool used by the BTSRC for administering the phone call ensures that no particular donor, holding constant her blood type, was targeted by the phone call. However, it is nevertheless useful to empirically verify its conditional randomization in the data, since it is always possible that the conditional randomization simply failed by chance.

To empirically test the conditional randomization, we estimate the following linear regression using ordinary least squares (OLS):

$$d_{i,t-1} = \beta_0 + \beta_1 T_{it} + \gamma' F_i + u_{it}\,, \tag{1}$$

where the dependent variable $d_{i,t-1}$ is a binary indicator whether donor i gave blood following the invitation to the previous blood drive at time $t-1$. The regression includes the treatment variable T_{it}, i.e. whether donor i received a phone call at for the upcoming blood drive at time t, and a vector of control variables F_i. In the first specification, the vector of control variables only includes blood types. We cluster the standard errors on individual donors to take any potential serial correlation within donors into account (*1, 2*).

In a second specification, we estimate the equation

$$d_{i,t-1} = \beta_0 + \beta_1 T_{it} + \gamma' F_{it} + \delta_{j(i)t} + u_{it}\,, \tag{2}$$

where F_{it} now includes the donor's gender, her age and fixed effects for the number of donations made in the 12 months prior to entering the study. We also add a fixed effect $\delta_{j(i)t}$ for blood drive j to which donor i was invited at time t, capturing differences in donations that affect all donors invited to a particular blood drive j alike. These control variables coincide with the ones in our main analysis using the FMM as specified in equation 5.

Table S2 shows the results of the randomization checks. As can be seen in column (1), the phone call is in no way correlated with previous donations, conditional on the donors' blood types. The point estimate is very small and far from being significant. The same is true for the specification shown in column (2) that includes all control variables of our main analysis. The point estimate is now even closer to zero. Thus, we conclude that the intervention passes the randomization check.

	(1)	(2)
Treatment Group T_{it}	-0.012	-0.003
	(0.010)	(0.010)
Blood type O-	0.101***	0.065***
	(0.011)	(0.010)
Blood type A+	-0.011***	-0.010***
	(0.004)	(0.004)
Blood type A-	0.051***	0.033***
	(0.008)	(0.007)
Male (=1)		0.041***
		(0.004)
Age		0.004***
		(0.000)
Number of donations before entering the study		
1 donation in the year prior to the study		0.178***
		(0.004)
2 donations in the year prior to the study		0.345***
		(0.006)
3 donations in the year prior to the study		0.457***
		(0.011)
4 donations in the year prior to the study		0.339***
		(0.048)
5 donations in the year prior to the study		0.366**
		(0.144)
No. of observations	85,529	85,529
No. of donors	35,764	35,764

Table S2: Checks of conditional randomization of the phone call (see equations 1 and 2). The dependent variable of the estimated equations is the decision to donate at the previous invitation ($d_{i,t-1}$). The number of observations is lower than in the full sample, because the most recent donation has to be dropped. This also reduces the number of donors to 35,764. Individual cluster robust standard errors are reported in parentheses. *Level of significance (t-test with H0: coefficient is zero)*: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

2 Statistical models

This section explains the statistical models we apply for analyzing the treatment effects of the phone call over time and across donors. It first presents the linear probability model for estimating the average treatment effects. Subsequently, it discusses the finite mixture model (FMM) that takes different types of donors into account and estimates the type-specific treatment effects.

2.1 Linear probability model for estimating average treatment effects

The linear probability model we apply for estimating the average treatment effects of the phone call on donation rates is based on the following specification:

$$d_{it} = \beta_{0i} + \beta_1 T_{it} + \beta_2 T_{i,t-1} + \beta_3 P_{it} + \gamma' F_{it} + \delta_{j(i)t} + u_{it}\,, \tag{3}$$

where the dependent variable d_{it} indicates whether donor i followed the invitation and showed up at the upcoming blood drive j at time t. The linear probability model is consistent and allows us to directly interpret the estimated coefficients as average changes of the probability to donate (*3*). Hence, the parameter β_{0i} measures donor i's baseline donation rate, i.e. her propensity to donate without experiencing any phone calls, that stems from any fixed attribute of the donor. Note that including individual fixed effects automatically controls for blood types (which is a fixed attribute of a donor), thus ensuring that the phone call is randomized conditional on the controls included in this regression.

The parameter β_1 captures the immediate treatment effect of a phone call, T_{it}, at time t. β_2 corresponds to the lagged treatment effect of a past phone call, $T_{i,t-1}$, and allows us to distinguish between habit formation ($\beta_2 > 0$) and guilt relief ($\beta_2 < 0$). The parameter β_3 measures the long-term treatment effect of the cumulative number of phone calls, $P_{it} = \sum_{s=1}^{t} T_{is}$, the donor received up to the current invitation at t. Additionally, the parameter vector γ captures

the effects of time-variant individual characteristics in F_{it}, which only consists of the donor's age in this case. In addition, the specification includes a fixed effect $\delta_{j(i)t}$ for blood drive j to which donor i was invited in period t, capturing differences in donations rates that affect all donors invited to a particular blood drive j alike. We remove the blood-drive fixed effect using the within-groups transformation (*2*). We report robust standard errors clustered at the individual level (*1, 2*).

We also perform an analysis of heterogeneity in the treatment effects commonly used in earlier work (*4, 5*). It tests whether there are differences in treatment effects along dimensions that are observable. In particular, we examine whether the treatment effect differs by age, gender and motivation of the blood donor as measured by the number of donations in the year prior to entering the study design. Specifically, we estimate

$$\begin{aligned} d_{it} = \beta_{0i} + \beta_1 T_{it} + \beta_2 T_{i,t-1} + \beta_3 P_{it} + \beta_4 (T_{it} \times z_i) + \beta_5 (T_{i,t-1} \times z_i) + \beta_6 (P_{it} \times z_i) \\ + \gamma' F_{it} + \delta_{j(i)t} + u_{it}\,, \end{aligned} \tag{4}$$

where the coefficients $\beta_4, \beta_5, \beta_6$ measure whether the individual characteristic z_i of donor i affects her immediate, lagged and long-term treatment effect. We then also perform an F-test of the joint hypothesis $\beta_4 = \beta_5 = \beta_6 = 0$, i.e. that there is no interaction between the characteristic z_i and any of the treatment effects.

2.2 Finite mixture model for estimating type-specific treatment effects

This section describes the FMM to estimate the type-specific treatment effects of the phone call on donation rates. Even though FMMs have applications in various fields (*6*), they are still relatively new to analyzing data from behavioral experiments (*7–12*).

2.2.1 Main specification

The FMM assumes the population to be made up by K distinct types of donors. Consequently, the baseline donation rate and the reactions to phone call are type-specific as indicated by the subscript k:

$$d_{it} = \beta_{0k} + \beta_{1k} T_{it} + \beta_{2k} T_{i,t-1} + \beta_{3k} P_{it} + \gamma' F_{it} + \delta_{j(i)t} + u_{it} \,. \tag{5}$$

The error term u_{it} is normally distributed with mean zero and variance σ^2. The control variables in F_{it} include a full set of dummies for blood types, age, gender and a full set of dummies for the number of donations in the 12 months prior to entering the study design. Again, the fixed effect $\delta_{j(i)t}$ for blood drive j to which donor i was invited in period t, captures differences in donations rates that affect all donors invited to a particular blood drive j alike. As above, we remove the blood-drive effect using the within-groups transformation (*2*).

Thus, the residual, $d_{it} - \hat{d}_{it}$, is normally distributed as well, leading to donor i's type-specific individual density,

$$f(\beta_k, \gamma, \sigma; X_i) = \prod_{t=1}^{T} \frac{1}{\sigma} \phi \left(\frac{d_{it} - \hat{d}_{it}}{\sigma} \right) \,, \tag{6}$$

where ϕ denotes the density function of the standard normal distribution and X_i corresponds to donor i's behavior and characteristics. Donor i's individual likelihood contribution,

$$\ell(\beta_k, \gamma, \sigma; X_i) = \sum_{k=1}^{K} \pi_k \, f(\beta_k, \gamma, \sigma; X_i) \,, \tag{7}$$

equals the sum over all K types of her type-specific densities, $f(\beta_k, \gamma, \sigma; X_i)$, weighted by the relative sizes of the corresponding types π_k. Since we do not know a priori to which type donor i belongs, the types' relative sizes π_k may be interpreted as the ex-ante probabilities of type-membership. Hence, the FMM's log likelihood is given by

$$\ln L(\Psi; X) = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k \, f(\beta_k, \gamma, \sigma; X_i) \,, \tag{8}$$

where the vector $\Psi = (\pi_1, \ldots, \pi_{K-1}, \beta_1, \ldots, \beta_K, \gamma, \sigma)$ contains all the parameters of the model. As in the linear probability model for the average treatment effects of the phone call, we also report individual cluster robust standard errors to take potential serial correlation in the donors' decisions into account (*1, 2*).

2.2.2 Estimation using the EM algorithm

As it is generally the case with FMMs, direct maximization of the log likelihood function is difficult and may encounter several problems. These problems are mainly caused by the particular non-linear form of the log likelihood function which involves a product between the types' relative sizes, π_k, and the corresponding individual densities, $f(\beta_k, \gamma, \sigma; X_i)$ (See (*6*) for a more extensive discussion).

However, the maximization problem would be much simpler if individual type-membership were observable a priori and indicated by $t_{ik} \in \{0, 1\}$. In that case, the individual contribution to the likelihood function would be given by

$$\tilde{\ell}(\beta_k, \gamma, \sigma; X_i, t_i) = \prod_{k=1}^{K} \left[\pi_k \, f(\beta_k, \gamma, \sigma; X_i)\right]^{t_{ik}} \ , \tag{9}$$

which would directly yield the complete-data log likelihood function

$$\ln \tilde{L}(\Psi; X, t) = \sum_{i=1}^{N} \sum_{k=1}^{K} t_{ik} \left[\ln \pi_k + \ln f(\beta_k, \gamma, \sigma; X_i)\right] \ . \tag{10}$$

Note that the types' relative sizes, π_k, and the individual densities, $f(\beta_k, \gamma, \sigma; X_i)$, enter the complete-data log likelihood as summands which could be maximized separately. Furthermore, the maximum likelihood estimates of the types' relative sizes, $\hat{\pi}_k = 1/N \sum_{i=1}^{N} t_{ik}$, would be given analytically.

The EM algorithm maximizes the complete-data log likelihood while treating the unobservable indicator t_{ik} as missing data (*13*). It iteratively proceeds in two steps, E and M, until it converges:

- In the E-step of the $(r+1)$th iteration, the EM algorithm augments the missing data given the model's actual fit, $\Psi^{(r)}$. By applying Bayes' rule, it computes each donor's individual probabilities of type-membership,

$$\tau_{ik} = \frac{\pi_k^{(r)} f(\beta_k^{(r)}, \gamma^{(r)}, \sigma^{(r)}; X_i)}{\sum_{m=1}^{K} \pi_m^{(r)} f(\beta_m^{(r)}, \gamma^{(r)}, \sigma^{(r)}; X_i)} \,. \tag{11}$$

- In the subsequent M-Step, the EM algorithm updates the model's fit. Hence, it maximizes the complete-data log likelihood by using the τ_{ik} from the E-step to replace the missing indicators t_{ik}:

$$\pi_k^{(r+1)} = 1/N \sum_{i=i}^{N} \tau_{ik} \,, \tag{12}$$

and

$$(\beta_1^{(r+1)}, \ldots, \beta_K^{(r+1)}, \gamma^{(r+1)}, \sigma^{(r+1)}) = \\ \underset{\beta_1,\ldots,\beta_K,\gamma,\sigma}{\arg\max} \sum_{i=1}^{N} \sum_{k=1}^{K} \tau_{ik} \, \ln f(\beta_k^{(r)}, \gamma^{(r)}, \sigma^{(r)}; X_i) \,. \tag{13}$$

The EM algorithm monotonically converges to the maximum likelihood estimate, $\hat{\Psi}$, since the likelihood never decreases from one iteration to the next. Convergence is achieved once the improvement in the log likelihood function between two iterations falls below an arbitrary small threshold. We applied several randomly generated start values to rule out solutions in which the EM algorithm converged to a local maximum.

2.2.3 Classification of donors into types

Once we obtained the FMM's parameter estimates, $\hat{\Psi}$, we can classify each donor into the type who's parameter estimates best fit the donor's behavior. In particular, we obtain the donor's individual probabilities of type-membership,

$$\tau_{ik} = \frac{\hat{\pi}_k f(\hat{\beta}_k, \hat{\gamma}, \hat{\sigma}; X_i)}{\sum_{m=1}^{K} \hat{\pi}_m f(\hat{\beta}_m, \hat{\gamma}, \hat{\sigma}; X_i)} \,, \tag{14}$$

based on the FMM's fitted parameters and the donor's behavior and characteristics X_i. According to Bayes' rule, these individual probabilities of type-membership correspond to the donor's type-specific likelihood contribution normalized by her overall likelihood contribution.

2.2.4 Cross-validation to determine the optimal number of distinct types of donors

An important aspect of estimating a FMM is to find the optimal number of types, K^*, the model controls for. If K is too low, the FMM disregards minority types as it lacks the flexibility to cope with the full extent of behavioral heterogeneity in the data. If K is too high, on the other hand, the FMM overfits the data as it models random noise instead of systematic behavioral differences between the types.

Unfortunately, standard statistical tests, such as likelihood ratio tests, are not applicable for determining K^*, since the distribution of the corresponding test stastistics is unknown (*14*). Moreover, classical model selection criteria, such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), are well known to perform badly when being applied to determine the optimal number of types (*15–18*).

We use the cross-validated log likelihood to find the optimal number of types K^* (*19*). To approximate the cross-validated log likelihood, we apply the following procedure: First, we randomly split the sample into a training and a test sample. Second, we estimate three FMMs with $K = 1, 2, 3$ types in the training sample. Third, we use the estimates obtained in the training sample to evaluate the log likelihood of the model in the test sample. We repeat this procedure 100 times and average over the log likelihood of the model in the test sample. The negative of the resulting cross-validated log likelihood is an unbiased estimate of the Kullback-Leibler distance between the true model with K^* types and the actual model with K types (see (*19*) for further details).

Intuitively, if the actual model has too few types, i.e. $K < K^*$, raising K will improve the

out-of-sample fit of the model in the test sample. Hence, the cross-validated log likelihood in the test sample will increase with a higher K. If the actual model has too many types, i.e. $K > K^*$, the model overfits the data as randomness in the training sample drives the parameter estimates. Thus, the model achieves only a poor out-of-sample fit in the test sample. Consequently, we choose the FMM that achieves the highest cross-validated log likelihood for determining the optimal number of types.

Table 2 in the article shows that the FMM with $K^* = 2$ types of donors represents the best compromise between parsimony and flexibility, as it consistently achieves the highest cross-validated log likelihood. The FMM with $K^* = 2$ types yields a substantially higher cross-validated log likelihood than the linear probability model with just $K = 1$ representative type, and a slightly higher cross-validated log likelihood than a FMM with $K = 3$ types.

Moreover, the FMM with $K^* = 2$ types provides a clean classification of donors into types. The histogram in figure 2 in the article illustrates this clean classification by showing the distribution of the donors' individual probabilities of type-membership, τ_{ik} (see equation 14). It reveals that almost all donors are cleanly classified either into the first or the second type, since nearly all of them exhibit a probability of belonging to the first type that is either very close to zero or very close to one. In contrast, the histograms in figure S1 show that a FMM with $K = 3$ types yields an ambiguous classification with substantial overlap between the types. This substantial overlap, which is especially pronounced between the second and third type, indicates that a FMM with $K = 3$ types overfits the data and tries to identify more types than exist. Consequently, we can identify two cleanly separated types of donors that differ in their behavior.

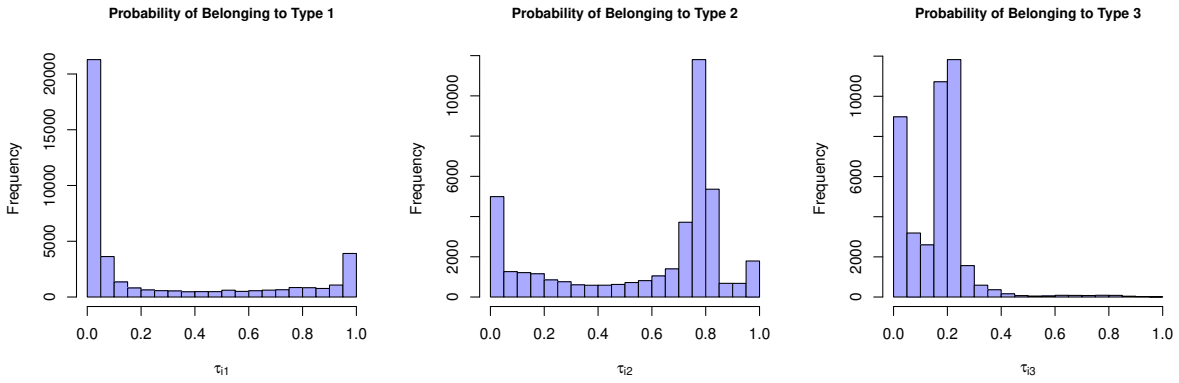

Figure S1: A finite mixture model with $\mathbf{K = 3}$ types yields an ambiguous classification of donors into types. The histograms show the distribution of the individual probabilities of type-membership, τ_{ik}, based on the finite mixture model with $K = 3$ types (see equation 14). The finite mixture model with $K = 3$ types overfits the data, since the resulting classification of donors into types is ambiguous and the types 2 and 3 overlap substantially.

3 Additional analyses

This section contains additional analyses, mainly to assess the robustness of the results reported in the article.

3.1 Testing whether crowding-out scales with the incidence of phone calls

In this section, we test whether the crowding-out effect in the group of type 1 donors scales with the incidence of phone calls, or whether it just represents a one time reduction in donation rates following the first phone call. We estimate the following alternative specification of the FMM,

$$d_{it} = \beta_{0k} + \beta_{1k} T_{it} + \beta_{2k} T_{i,t-1} + \beta_{3k} P_{it} + \beta_{4k} I(P_{it} \geq 1) + \gamma' F_{it} + \delta_{j(i)t} + u_{it}\,, \qquad (15)$$

where the additional indicator variable $I(P_{it} \geq 1) = 1$ following the first phone call, and $I(P_{it} \geq 1) = 0$ otherwise.

Table S3 shows the estimated coefficients of the alternatively specified FMM. It provides strong evidence that the crowding-out effect in the group of type 1 donors scales with the incidence of phone calls, since $\hat{\beta}_{31}$ remains negative and highly significant. In contrast, there is no evidence that having received the first phone call has any additional effects, as $\hat{\beta}_{4k}$ is insignificant for both types of donors. Moreover, the remaining coefficients are robust, and the alternative specification does not outperform the original one (Likelihood ratio test: P=0.073).

Dependent variable:	Type 1	Type 2
Donation at upcoming blood drive	Donors	Donors
Type-specific coefficients		
Share among the population	0.271	0.729
	(0.003)	(0.003)
Baseline donation rate	0.463***	-0.072***
	(0.007)	(0.004)
Immediate treatment effect	0.090***	0.054***
	(0.015)	(0.009)
Lagged treatment effect	0.005	0.018*
	(0.021)	(0.010)
Long-term treatment effect	-0.029***	-0.007
	(0.011)	(0.005)
Has been called	0.025	0.018
	(0.021)	(0.011)
Common coefficients		
Blood type 0-	-0.005	
	(0.008)	
Blood type A+	-0.002	
	(0.003)	
Blood type A-	0.000	
	(0.006)	
Male	0.028***	
	(0.002)	
Age	0.003***	
	(0.000)	
1 donation in the year prior to the study	0.128***	
	(0.003)	
2 donations in the year prior to the study	0.250***	
	(0.005)	
3 donations in the year prior to the study	0.324***	
	(0.010)	
4 donations in the year prior to the study	0.202***	
	(0.035)	
5 donations in the year prior to the study	-0.002	
	(0.057)	
Error term's standard deviation	0.366	
	(0.001)	
No. of observations	126,123	
No. of donors	40,653	

Table S3: Alternative specification of the finite mixture model with $\mathbf{K^* = 2}$ types for testing whether crowding-out scales with the incidence of phone calls (see equation 15). The control variable age is centered on the sample average of 43.4 years. The model controls for blood drive fixed effects. Individual cluster robust standard errors are reported in parentheses. *Level of significance (t-test with H0: coefficient is zero)*: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

3.2 Testing whether the history of phone calls influences their immediate treatment effect

Here, we analyze whether the history of phone calls influences their immediate treatment effect. We estimate the following alternative specification of the FMM,

$$d_{it} = \beta_{0k} + \beta_{1k} T_{it} + \beta_{2k} T_{i,t-1} + \beta_{3k} P_{it} + \beta_{4k} \left(T_{it} \times P_{it}\right) + \gamma' F_{it} + \delta_{j(i)t} + u_{it}\,, \tag{16}$$

where $(T_{it} \times P_{it})$ corresponds to the interaction between the phone call at t, T_{it}, and the cumulative number of phone calls received up to t, P_{it}.

Table S4 shows the estimated coefficients of the alternatively specified FMM. It reveals that the history of phone calls has no influence on the immediate treatment effect, since the corresponding coefficient, $\hat{\beta}_{4k}$, is insignificant for both types of donors. Furthermore, the remaining coefficients are robust, and the alternative specification does not outperform the original one (Likelihood ratio test: P=0.147).

Dependent variable:	Type 1	Type 2
Donation at upcoming blood drive	Donors	Donors
Type-specific coefficients		
Share among the Population	0.271	0.729
	(0.003)	(0.003)
Baseline donation rate	0.463***	-0.072***
	(0.007)	(0.004)
Immediate treatment effect	0.113***	0.069***
	(0.017)	(0.011)
Lagged treatment effect	0.008	0.022**
	(0.021)	(0.010)
Long-term treatment effect	-0.018**	0.001
	(0.008)	(0.004)
$T_{it} \times P_{it}$	-0.011	-0.007
	(0.010)	(0.006)
Common coefficients		
Blood type 0-	-0.006	
	(0.008)	
Blood type A+	-0.002	
	(0.003)	
Blood type A-	0.000	
	(0.006)	
Male	0.028***	
	(0.002)	
Age	0.003***	
	(0.000)	
1 donation in the year prior to the study	0.128***	
	(0.003)	
2 donations in the year prior to the study	0.250***	
	(0.005)	
3 donations in the year prior to the study	0.324***	
	(0.010)	
4 donations in the year prior to the study	0.205***	
	(0.035)	
5 donations in the year prior to the study	0.005	
	(0.060)	
Error term's standard deviation	0.366	
	(0.001)	
No. of observations	126,123	
No. of donors	40,653	

Table S4: Alternative specification of the finite mixture model with $\mathbf{K^* = 2}$ types for testing whether the history of phone calls influences the immediate treatment effect of the phone call (see equation 16). The control variable age is centered on the sample average of 43.4 years. The model controls for blood drive fixed effects. Individual cluster robust standard errors are reported in parentheses. *Level of significance (t-test with H0: coefficient is zero)*: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

4 References

1. Huber PJ (1967) in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press).

2. Wooldridge JM (2002) *Econometric Analysis of Cross Section and Panel Data* (MIT Press).

3. Moffitt RA (1999) New developments in econometric methods for labor market analysis. *Handbook of labor economics* 3:1367–1397.

4. Lacetera N, Macis M, Slonim R (2014) Rewarding volunteers: A field experiment. *Management Science* 60:1107–1129.

5. Goette L, Stutzer A, Yavuzcan G, Frey BM (2009) Free cholesterol testing as a motivation device in blood donations: Evidence from field experiments. *Transfusion* 49:524–531.

6. McLachlan G, Peel D (2000) *Finite Mixture Models* (Wiley Series in Probabilities and Statistics).

7. El-Gamal M, Grether D (1995) Are people bayesian? uncovering behavioral strategies. *Journal of the American Statistical Association* 90:1137–1145.

8. Stahl D, Wilson P (1995) On players a models of other players: Theory and experimental evidence. *Games and Economic Behavior* 10:218–254.

9. Houser D, Keane M, McCabe K (2004) Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica* 72:781–822.

10. Houser D, Winter J (2004) How do behavioral assumptions affect structural inference? *Journal of Business and Economic Statistics* 22:64–70.

11. Bruhin A, Fehr-Duda H, Epper T (2010) Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica* 78:1375–1412.

12. Conte A, Hey JD, Moffat PG (2011) Mixture models of choice under risk. *Journal of Econometrics* 162:79–88.

13. Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B* 39:1–38.

14. Cameron CA, Trivedi PK (2005) *Microeconometrics: Methods and Applications* (Cambridge University Press).

15. Atkinson A (1981) Likelihood ratios, posterior odds and information criteria. *Journal of Econometrics* 16:15–20.

16. Geweke J, Meese R (1981) Estimating regression models of finite but unknown order. *International Economic Review* 22:55–70.

17. Celeux G, Soromenho G (1996) An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* 13:195–212.

18. Biernacki C, Celeux G, Govaert G (2000) An improvement of the nec criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters* 20:267–272.

19. Smyth P (2000) Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing* 9:63–72.