Beyond the Target Customer: Social Effects of CRM Campaigns

Eva Ascarza, Peter Ebbes, Oded Netzer, Matthew Danielson

Link to article: http://journals.ama.org/doi/abs/10.1509/jmr.15.0442

WEB APPENDICES

In this web appendix, we present a detailed description of the analyses performed to obtain the results reported in the main document as well as multiple tests for the robustness of our results.

WEB APPENDIX – A1: RANDOMIZATION CHECK ON THE ORIGINAL VARIABLES

We replicate Table 3 of the main document to test the randomization of our experiment using the original (before log) variables. While Table 3 is for the log transformed activities (that were used as dependent variables for our main analyses in the diff-in-diff regression models), below we repeat the analyses for the original variables to corroborate the model free analyses in the main document that were done on the original variables. The randomization was implemented to select the egos, and as can be seen, none of the differences between treatment and control groups for the egos are significant. Similarly, we also find that the alters in the treatment group and the alters in the control group have, on average, the same usage levels before the intervention on all variables. For calls, the differences are marginally significant with p-values of 0.08 and 0.07. The treatment alters tended to call a little less, on average, before the intervention. Given that we do not find this on the log-transformed variables (Table 3 main document), we think this marginal difference is mostly driven by outliers in the call usage distribution for the alters (see e.g. Table 2 in the main document). Furthermore, because we use the log-transformed variables for our main analyses, we do not consider this a concern for our findings, and conclude that the randomization between the control and treatment groups was well executed.

	Control		Treatment		Difference		
	Mean	St. Error	Mean	St. Error	Difference	St. Err	p-value
Focal usage							
Inbound SMS	11.60	3.32	8.38	0.72	-3.22	2.70	0.23
Outbound SMS	39.42	4.35	32.66	1.92	-6.76	4.16	0.10
Inbound MIN	3.68	0.37	3.30	0.22	-0.38	0.40	0.35
Outbound MIN	24.07	1.79	21.53	0.99	-2.53	1.89	0.18
Inbound CALLS	35.77	4.99	35.98	3.90	0.21	6.38	0.97
Outbound CALLS	69.99	8.01	72.00	6.34	2.01	10.32	0.85
Ego usage							
Inbound SMS	38.38	3.59	40.08	3.34	-2.83	5.17	0.74
Outbound SMS	48.68	5.61	45.87	3.90	-8.84	6.67	0.67
Inbound MIN	16.50	1.55	18.90	1.56	1.38	2.37	0.31
Outbound MIN	30.76	2.89	30.36	2.00	-1.75	3.42	0.91
Inbound CALLS	116.66	11.56	106.74	8.08	-25.48	13.79	0.47
Outbound CALLS	88.74	10.66	78.80	7.38	-24.25	12.64	0.43

Table A - 1: Randomization check in all observed variables (before log) in the four weeks before the experiment

WEB APPENDIX – A2: TIME SERIES PLOTS FOR ACTIVITIES ON THE LOG SCALE

Here we present the same figures as Figure 2 (post-treatment ego usage), but for the log-transformed variables. With respect to alter usage, we present the other two activities (calls and sms) as well as the log-transformed variables for all three behaviors. The log-transformed variables were used in the diff-in-diffs regression models (tables 5, 6, 8, 9 main document). By eyeballing the time series plots in the following two figures, we can see that the treatment group generally exhibits higher consumption on most of the main activity variables (minutes, calls and SMS), for most of the time. Hence, these time-series plots are in support of the diff-in-diffs regression model results presented in the main document.

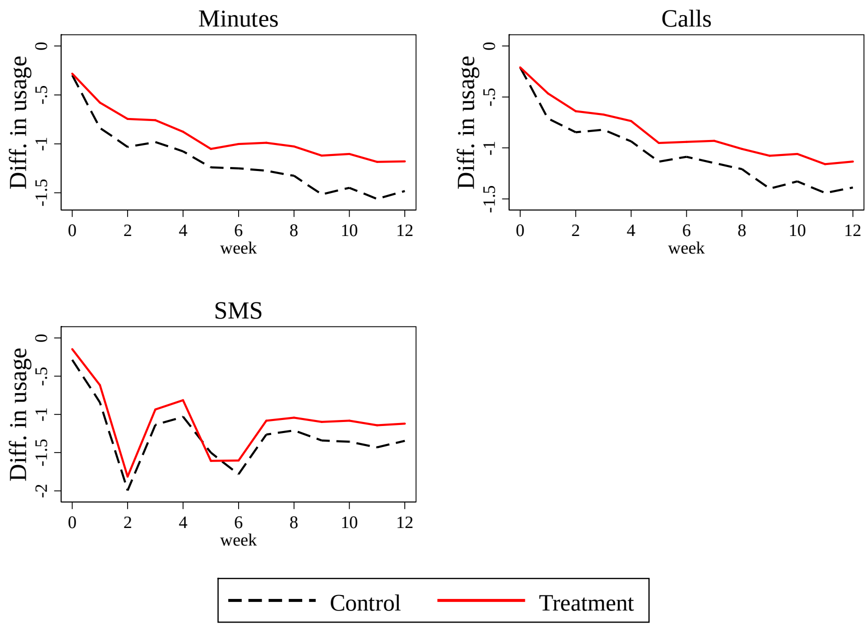

Figure A-1: Average difference between pre- and post-treatment ego usage, by treatment condition, on the log scale

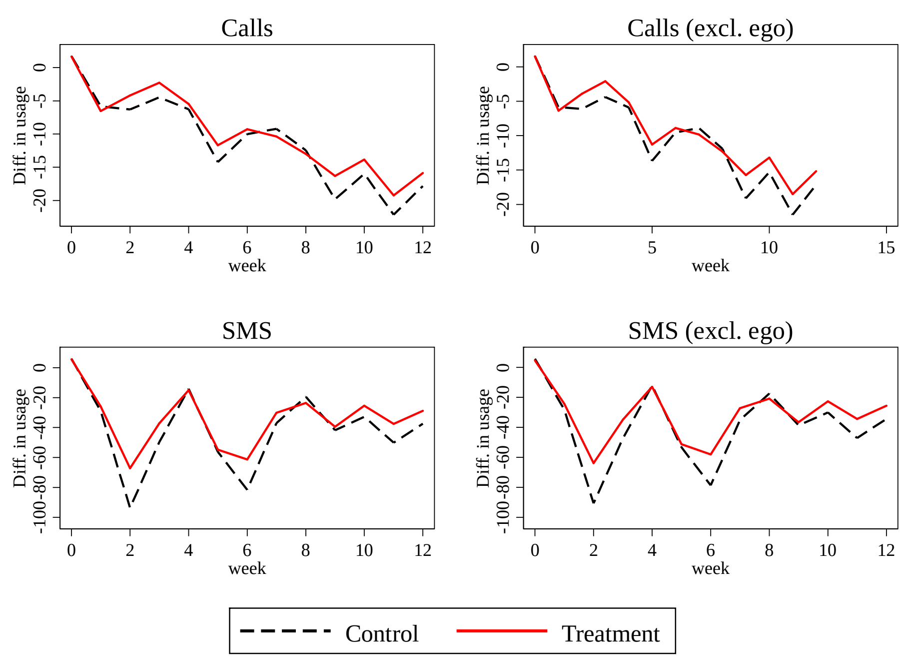

Figure A-2: Average difference between pre- and post-treatment alter usage (calls and SMS), by treatment condition

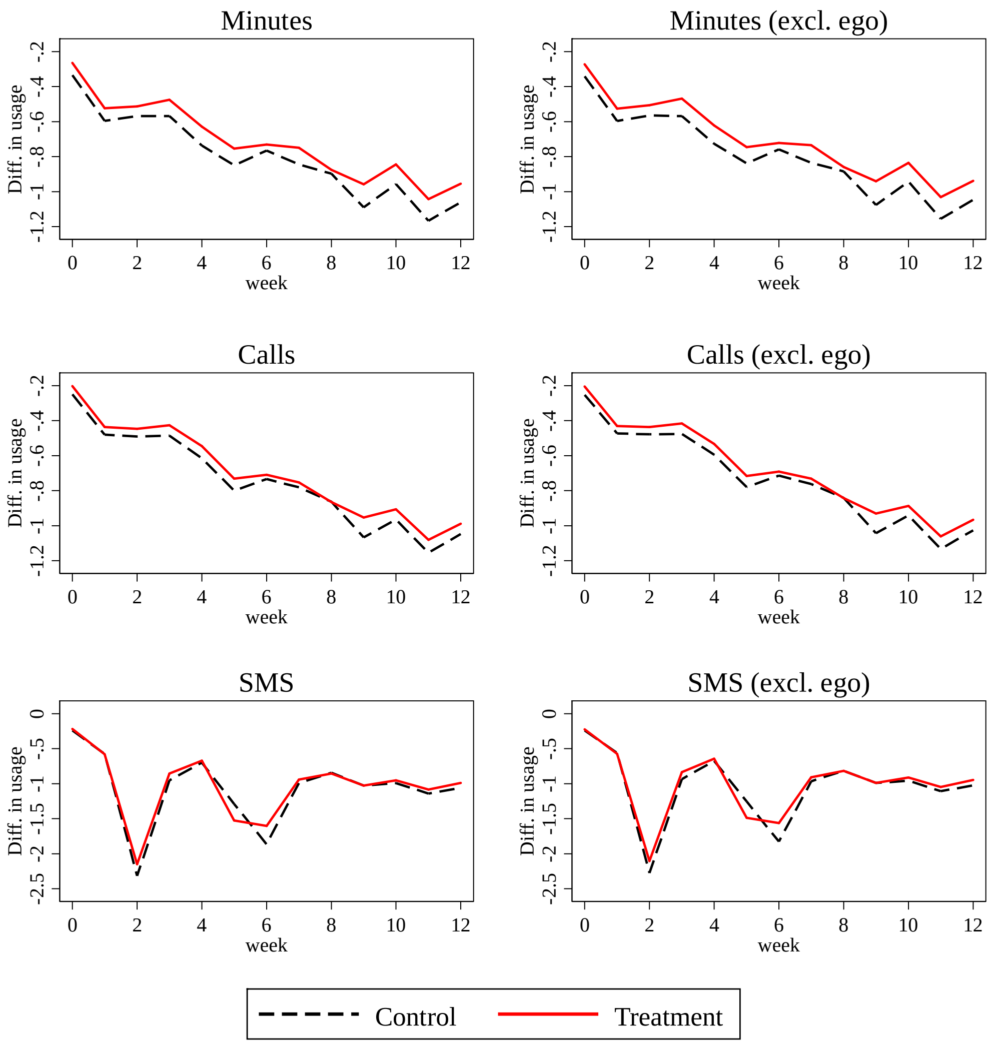

Figure A-3: Average difference between pre- and post-treatment alter usage (minutes, calls and SMS), by treatment condition, on log scale

WEB APPENDIX – A3: MODELING APPROACH

Here we derive the main model equations to estimate the effect of the treatment dummy on ego and alter usage. We exploit the panel nature of our data by using a difference in differences approach (diff-in-diffs). The diff-in-diffs model allows us to control for unobserved heterogeneity in behavior by comparing the pre-treatment behavior to the post-treatment behavior. More specifically, we consider the following baseline model for the effect of treatment:

$$y_{it}^{ego} = \alpha_i + \beta T_i + \lambda_t + \epsilon_{it} \;, \qquad \text{(A-1)}$$

where y_{it}^{ego} represents the usage (e.g., number of minutes called) of ego $i = 1, \ldots, I$ in week $t = 1, \ldots, T$. The term λ_t is a time-specific (week) effect, α_i is an ego user-specific intercept (capturing unobserved heterogeneity in usage), T_i is the treatment dummy that equals 1 if ego i received the treatment and 0 otherwise, and ϵ_{it} is an error term.

We consider the following pre-treatment model (say time period $t = 0$), which has the same structure as the baseline model in (A-1):

$$y_{i0}^{ego} = \alpha_i + \beta \times 0 + \lambda_0 + \epsilon_{i0} \;, \qquad \text{(A-2)}$$

where all symbols are defined similarly, and we use the fact that $T_i = 0$ for all i before the treatment. Subtracting the two equations, the term α_i drops, resulting in the following (diff-in-diffs) regression model:

$$\Delta y_{it}^{ego} = y_{it}^{ego} - y_{i0}^{ego} = \beta T_i + (\lambda_t - \lambda_0) + (\epsilon_{it} - \epsilon_{i0}) = \beta T_i + \tilde{\lambda}_t + \tilde{\epsilon}_{it} \;. \qquad \text{(A-3)}$$

In our study, we operationalize y_{i0}^{ego} as the log of the average usage of ego i in the four weeks prior to the treatment, i.e., $y_{i0}^{ego} = \log(\frac{1}{4}\sum_{t=-3}^{0} y'^{ego}_{it} + 1)$, where y'^{ego}_{it} is the observed usage of ego i in week t. Furthermore, y_{it}^{ego} is the log of the observed activity plus 1, i.e. $y_{it}^{ego} = \log(y'^{ego}_{it} + 1)$. Because we have a limited number of observations (time periods) per ego, estimating the diff-in-diffs regression model in (A-3) is preferred to estimating the baseline model (A-1) with a random intercept α_i.[1] We use robust (panel corrected) standard errors to account for potential serial correlation in the model error terms (Xtpcse command in STATA, e.g., Hoechle 2007) at the ego level.

Similarly, we estimate the effect of the treatment on the alter usage with a diff-in-diffs regression approach. The effect of the treatment on alter usage may be estimated from the following baseline model:

[1] The diff-in-diffs models presented in the main manuscript include the intercept and *T*-1 week dummies. Note that such models are equivalent to equation (A-4) that includes λ_t for all $t = 1, \ldots, T$.

$$y_{ijt}^{alter} = \alpha_{ij} + \beta T_i + \lambda_t + \epsilon_{ijt}\,, \qquad \text{(A-4)}$$

where $i = 1, \ldots, I$ indicates ego, $j = 1, \ldots, J_i$ represents the j-th alter of ego i, and t indicates week. All other symbols are defined as before. Here, the intercept α_{ij} captures (unobserved) heterogeneity that is specific to the relationship between ego i and her alter j. The pre-experiment model is

$$y_{ij0}^{alter} = \alpha_{ij} + \beta \times 0 + \lambda_0 + \epsilon_{ij0}\,, \qquad \text{(A-5)}$$

and subtracting Equation (A-5) from Equation (A-4) gives the diff-in-diffs model for the effect of treatment on alter usage:

$$\Delta y_{ijt}^{alter} = y_{ijt}^{alter} - y_{ij0}^{alter} = \beta T_i + (\lambda_t - \lambda_0) + \left(\epsilon_{ijt} - \epsilon_{ij0}\right) = \beta T_i + \tilde{\lambda}_t + \tilde{\epsilon}_{ijt}\,. \quad \text{(A-6)}$$

The dependent variable is operationalized in a similar way to that of the ego usage, where y_{ijt}^{alter} is the log of alter j (of ego i)'s observed usage (plus 1), and y_{ij0}^{alter} is the log of the average observed usage (plus 1) of alter j of ego i across the four weeks before the treatment. We use robust (panel corrected) standard errors to account for potential serial correlation in the model error terms at the alter level.

For alters, we also estimate the effect of treatment on suspension and churn. Given the binary nature of these two variables, we do not employ a diff-in-diffs approach but rather use a binary probit model (last two columns of Tables 8 and 9 in the main document). To account for the panel nature of our data and unobserved heterogeneity in the probit model, we estimate a model that clusters the data at the alter level to appropriately estimate the standard errors of the estimated regression effects.

WEB APPENDIX – A4: ALTERNATIVE METRICS FOR ALTER USAGE

We replicate the results shown in Section 3.2 in the main document by using different metrics for alter usage, namely calls and SMS. The following results compare with Tables 8 and 9 in the main document which shows the effect of the treatment on alters for outbound minutes. We find similar results when using outbound calls and SMS instead of the outbound minutes activity reported in the main document. The effect of treatment on alter outbound calls and SMS is significant and substantial.

	Outbound Calls		
	Total	Total (excl. Ego)	To Ego
Treatment	0.0518***	0.0488***	0.0281***
	(0.017)	(0.017)	(0.007)
Constant	−0.484***	−0.477***	−0.175***
	(0.023)	(0.022)	(0.01)
Week dummies	Yes	Yes	Yes
Observations	27,987	27,987	27,987

Short-term effects of treatment on alter usage. Linear (diff-in-diffs) regression for usage. *** p<0.01.Robust standard errors in parentheses. The number of observations is 6 (weeks) x 4,700 (alters), excluding alters that are cancelled in a particular week.

Table A - 2: Short-term effect of treatment on alter calls (weeks 1-6 after the treatment)

	Outbound Calls		
	Total	Total (excl. Ego)	To Ego
Treatment	0.0550***	0.0550***	0.0253***
	(0.02)	(0.02)	(0.007)
Constant	−0.797***	−0.776***	−0.271***
	(0.027)	(0.026)	(0.01)
Week dummies	Yes	Yes	Yes
Observations	27,598	27,598	27,598

Long-term effects of treatment on alter usage. Linear (diff-in-diffs) regression for usage. *** p<0.01. Robust standard errors in parentheses. The number of observations is 6 (weeks) x 4,700 (alters), excluding alters that are cancelled in a particular week.

Table A - 3: Long-term effect of treatment on alter calls (weeks 7-12 after the treatment)

	Outbound SMS		
	Total	Total (excl. Ego)	To Ego
Treatment	0.0527**	0.0531**	0.0289**
	(0.022)	(0.022)	(0.011)
Constant	−0.608***	−0.600***	−0.328***
	(0.03)	(0.029)	(0.015)
Week dummies	Yes	Yes	Yes
Observations	27,987	27,987	27,987

Short-term effects of treatment on alter usage. *** p<0.01, ** p<0.05. Robust standard errors in parentheses. The number of observations is 6 (weeks) x 4,700 (alters), excluding alters that are cancelled in a particular week.

Table A - 4: Short-term effect of treatment on alter SMS (weeks 1-6 after the treatment)

	Outbound SMS		
	Total	Total (excl. Ego)	To Ego
Treatment	0.0343	0.0389*	0.0397***
	(0.023)	(0.022)	(0.012)
Constant	−0.983***	−0.954***	−0.495***
	(0.03)	(0.03)	(0.016)
Week dummies	Yes	Yes	Yes
Observations	27,598	27,598	27,598

Long-term effects of treatment on alter usage. *** p<0.01, ** p<0.05. Robust standard errors in parentheses. The number of observations is 6 (weeks) x 4,700 (alters), excluding alters that are cancelled in a particular week.

Table A - 5: Long-term effect of treatment on alter SMS (weeks 7-12 after the treatment)

WEB APPENDIX – A5: DETAILS AND ROBUSTNESS OF THE INSTRUMENTAL VARIABLE (IV) ANALYSES

This web appendix includes two sub-sections. In Sub-section A5.1 we present the details for the IV regression presented in Section 3.3.1 of the main document. Furthermore, because the IV approach has been shown to be sensitive to underlying model assumptions (e.g., Rossi 2014; Germann, Ebbes and Grewal 2015), we also present several robustness checks in Sub-section A5.2.

A5.1 Main IV regression in section 3.3.1

Our goal is to estimate the dashed-arrow in Figure 5b in the main document. As argued in the main document in Section 3.3.1, a simple regression model that regresses the alter usage on ego usage would likely suffer from endogeneity bias due to the presence of omitted variables that could affect the usage of both egos and alters. It should be noted that we can consistently estimate the (causal) effect of the marketing campaign on the alters' usage and churn (results reported in Section 3.3.1 and represented by arrow B in Figure 5a in the main document) because the treatment variable is exogenous by design and is therefore uncorrelated with any unobservable. The endogeneity problem only emerges when one tries to establish a causal link between ego usage and alter usage or churn (i.e. the dashed line in Figure 5b).

We choose weeks 1—6 (short-term) to measure egos' usage and weeks 7—12 (long-term) to measure alters' usage or churn for two main reasons. First, we want to allow some time for the alters to notice the change in the network activity. Second, we want to ensure that there is no simultaneity in the consumption decisions of egos and alters. More formally, we would expect that the cause (ego usage) precedes the effect (alter usage or churn).

The results presented in the main document were obtained using a control function approach[2] (Petrin and Train 2010; Germann, Ebbes and Grewal 2015). Specifically, we estimate the following regression equation using OLS:

$$\Delta y_{ijt}^{alter} = \emptyset_0 + \emptyset_1 \Delta\, y_{ij,short}^{ego} + \sum_{\tau=8}^{12} \emptyset_{\tau-4} D_{\tau t} + \emptyset_7 \widehat{\omega}_{ij} + \zeta_{ijt} \quad \text{for } t = 7,8,\dots,12, \qquad \text{(A1)}$$

where Δy_{ijt}^{alter} is defined as in Equation (4) in the main document, $\emptyset_k$ are regression parameters, $D_{\tau t}$ are time dummies as in Equations (3)-(6) in the main document, and ζ_{ijt} is an error term with 0 mean variance σ_ζ^2. The term $\Delta y_{ij,short}^{ego}$ represents the endogenous variable capturing short-term activity of ego i directed to its j-th alter (e.g., average minutes that ego i called to alter j in weeks 1 to 6). This variable is computed as $\Delta y_{ij,short}^{ego} = \frac{1}{6}\sum_{\tau=1}^{6} \Delta y_{ij\tau}^{ego}$, where $\Delta y_{ijt}^{ego} = y_{ijt}^{ego} - y_{ij0}^{ego}$ is the communication between ego i and alter j (in logs plus 1) in week t less the average pre-experiment communication between ego i to alter j (in logs plus 1). Lastly, the term $\widehat{\omega}_{ij}$ is the "control function" component, which is computed as the estimated residual of the first stage regression:

$$\Delta y_{ij,short}^{ego} = \chi_0 + \chi_1\, T_i + \omega_{ij}. \qquad \text{(A2)}$$

Here, χ_0 and χ_1 are the (first-stage) regression parameters, T_i is defined as in Equation (1)-(4) in the main document, and ω_{ij} is the error term that is (potentially) correlated with the error term ζ_{ijt} in Equation (A1). Thus, the treatment dummy T_i acts as the instrument in a standard IV regression approach.

We estimate a separate model for each type of activity (minutes, calls, and SMS). More specifically, when analyzing minutes we compute the short-term ego usage variable ($\Delta y_{ij,short}^{ego}$) using the number of minutes ego i called alter j, and the dependent variable (Δy_{ijt}^{alter}) using the number of minutes alter j called other individuals, excluding the calls she made to ego i. We conduct similar analysis for SMS and calls. We also estimate the effect of ego usage on alter churn. We use a similar

[2] In the next section A3.2 of this web appendix, we compare the estimates from the control function approach with those of 2SLS for the linear case and the two-step estimator (Newey 1987) for the probit case. We find that the insights from the IV approach presented in the main text in Table 10 do not change.

regression specification as in Equations (A1) and (A2) where the dependent variable now is the indicator variable y_{ijt}^{alter}, that equals 1 if alter j of ego i churns in week t, and 0 otherwise. Formally, we estimate:

$$Prob(y_{ijt}^{alter}) = \rho_0 + \rho_1 \, \Delta y_{ij,short}^{ego} + \sum_{\tau=8}^{12} \rho_{\tau-6} D_{\tau t} + \rho_7 \widehat{\omega}_{ij} + \varsigma_{ijt} \text{ for } t = 7,8,\ldots,12, \quad \text{(A3)}$$

where all regressors are defined as in Equations (A1) and (A2). The term ς_{ijt} is normally distributed with 0 mean and variance σ_{ς}^2, resulting in a standard probit model estimated including a control function component. The results for these IV regressions are presented in Table 10 in the main document.

A5.2 Alternative estimation approach for the IV regression

In estimating the effect of ego usage on alter usage in Section 3.3 in the main document, we use the control function approach as opposed to a two stage least squares (2SLS) approach to avoid including the week dummies in the first stage, as the endogenous variable $\Delta y_{ij,short}^{ego}$ is time invariant. The 2SLS procedure generally includes all exogenous independent variables of the main regression equation (including all dummy variables) in the first stage regression (e.g., Wooldridge 2002 p.91). While this does not create any estimation issues per se, the time dummies are redundant in the first stage in this particular case, and may therefore lead to a less efficient instrumental variable estimator. Nevertheless, in this appendix we replicate the results presented in Section 3.3 in the main document by estimating Equation (A1) with two stage least squares (2SLS), excluding the control $\widehat{\omega}_{ij}$. The instrumental variable results of estimating Equation (A1) with 2SLS are given in Table A-5. We can see that the standard errors are indeed higher for 2SLS relative to control function approach. The point estimate of the effect of ego usage on alter usage are practically the same for the two estimation approaches.[3] But more importantly, the significance of the model parameters are not affected.

For the probit model, we examine the robustness of our IV probit regressions (for suspension and churn) using the two-step estimator (Newey 1987). We confirm that the effect of ego usage on alter churn is robust because the ratio between the estimated regressor effect and the intercept are similar for the two approaches, 1.67 (=−7.689/−4.601) for the control function approach (Table 10, main document) and 1.67 (=−1.527/−0.915) for the two-step approach (Table A-6).

[3] Note that the point estimates would have been identical if all regressors were included in the first-stage regression (A2).

Thus, while we use the control function approach for convenience, estimating the IV regression using 2SLS and or a two-step estimator approach leads to nearly identical results.

	Alter usage					
	Minutes	Churn	Calls	Churn	SMS	Churn
Ego to Alter (regressor)						
Minutes	3.204***	−1.528***				
	(1.054)	(0.031)				
Calls			1.765***	−2.248***		
			(0.679)	(0.074)		
SMS					0.891*	−1.129***
					(0.536)	(0.022)
Intercept	−0.0298	−0.921***	−0.458***	−1.201***	−0.520**	−1.105***
	(0.248)	(0.224)	(0.113)	(0.28)	(0.249)	(0.224)
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,598	27,598	27,598	27,598	27,598	27,598

Effect of short-term ego ego-to-alter usage on long-term alter usage using 2SLS and on churn using the two-step estimator (Newey 1987). The regressor ego usage is operationalized in the same way as in the main text. *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parentheses.

Table A - 6: Effect of short-term ego-to-alter usage $\Delta y_{ij,short}^{ego}$ on long-term alter usage (Instrumental variable regressions using 2SLS and two-step estimator instead of a control function approach)

Alternative metrics for short-term ego usage: We replicate the results presented in Section 3.3 in the main document using alternative metrics for the regressor that captures short-term ego activity in Equations (A1) and (A3). The results in Table 10 in the main document use the average differenced ego usage (Δy_{ijt}^{ego}) in the first 6 weeks after the campaign, i.e. $\Delta y_{ij,short}^{ego} = \frac{1}{6}\sum_{t=1}^{6} \Delta y_{ijt}^{ego}$, as the endogenous regressor. Here we consider two alternative specifications to the averaged differences for short term ego usage: (1) the six-week lag of the differenced usage, i.e. $\Delta y_{ijt,lag}^{ego} = y_{ijt-6}^{ego} - y_{ij0}^{ego}$, and (2) the differenced usage averaged up to the week prior to week t, for $t = 7,8,\dots,12$. That is, $\Delta y_{ijt,upto}^{ego} = \frac{1}{t-1}\sum_{\tau=1}^{t-1}(y_{ij\tau}^{ego} - y_{ij0}^{ego})$, for $t = 7,8,\dots,12$.

The IV results with $\Delta y_{ijt,lag}^{ego}$ and $\Delta y_{ijt,upto}^{ego}$ specification for the endogenous variables are given in Tables A-7 and A-8, respectively. We note that the underlying activity data to operationalize

$\Delta y_{ijt,lag}^{ego}$ and $\Delta y_{ijt,upto}^{ego}$ is only the (directed) activity of the ego to the alter, as in the main document in Section 3.3. It follows from Tables A-7 and A-8 that the results for $\Delta y_{ijt,lag}^{ego}$ and $\Delta y_{ijt,upto}^{ego}$ are very similar to the results reported in the main document (Table 10) and to each other. Importantly, in both specifications the exogenous instrumental variable (treatment dummy) is strongly significant for the three activities in the first-stage regression. This result reinforces our conclusion that when the ego uses more in the short term, the alters tend to use more and churn less in the long term. The effect sizes are slightly larger for $\Delta y_{ijt,upto}^{ego}$ relative to the results reported in the paper and relative to $\Delta y_{ijt,lag}^{ego}$, which may be expected as this specification of ego usage includes ego activity up to a week before the alter activity takes place, resulting in a shorter effective time lag between the ego and alter activities.

In sum, we investigate three different specifications to represent short-term ego usage, for three different types of activities (minutes, calls, and SMS), as well as two estimation approaches (control function and 2SLS or two-step estimator). We find that our results are robust to the operationalization of the ego usage variable. Furthermore, the different model specifications and estimation methods produce similar results and insights: higher activity of the egos in the short term leads to lower likelihood of churn and higher activity of the alters in the long term.

	Alter usage					
	Minutes	Churn	Calls	Churn	SMS	Churn
Ego to Alter (regressor)						
Minutes	3.230***	−7.580***				
	(0.732)	(2.55)				
Calls			1.783***	−7.585***		
			(0.644)	(2.722)		
SMS					0.894*	−5.436***
					(0.529)	(1.907)
Intercept	−0.0283	−4.545***	−0.462***	−4.012***	−0.539**	−5.294***
	(0.172)	(0.606)	(0.107)	(0.471)	(0.246)	(0.902)
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,598	27,598	27,598	27,598	27,598	27,598
1st stage t-stat	3.375	3.375	4.935	4.935	3.447	3.447

Effect of short-term ego ego-to-alter usage on long-term alter usage and churn. The regressor ego usage is operationalized as the ego usage 6 weeks earlier ($\Delta y_{ijt,lag}^{ego}$). Bootstrapping is used to estimate the standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

Table A - 7: Effect of short-term ego-to-alter usage $\Delta y_{ijt,lag}^{ego}$ on long-term alter usage (Instrumental variable regressions using control function approach)

	Alter usage					
	Minutes	Churn	Calls	Churn	SMS	Churn
Ego to Alter (regressor)						
Minutes	3.106***	−7.297***				
	(0.708)	(2.539)				
Calls			1.816***	−7.717***		
			(0.653)	(2.588)		
SMS					0.904*	−5.476***
					(0.528)	(1.845)
Intercept	−0.001	−4.610***	−0.427***	−4.155***	−0.490*	−5.506***
	(0.179)	(0.659)	(0.119)	(0.481)	(0.262)	(0.94)
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,598	27,598	27,598	27,598	27,598	27,598
1st stage t-stat	4.054	4.054	5.748	5.748	3.958	3.958

Effect of short-term ego usage (ego-to-alter) on long-term alter usage (total usage) and churn. The regressor ego usage is operationalized as the average of ego usage across all earlier weeks up to the current week ($\Delta y_{ijt,upto}^{ego}$). Bootstrapping is used to estimate the standard errors (in parentheses). *** p<0.01, ** p<0.05, * p<0.1.

Table A - 8: Effect of short-term ego-to-alter usage $\boldsymbol{\Delta y_{ijt,upto}^{ego}}$ on long-term alter usage (Instrumental variable regressions using control function approach)

WEB APPENDIX – A6: ALTERNATIVE MEASURE OF STRENGTH OF TIES

In this web appendix we replicate the results presented in Section 3.3.2 in which we investigate the moderating effect of strength of ties on the social effect. Recall from Section 3.3.2 that we operationalized strength of ties as the average number of minutes an alter called to her ego during the 4 weeks prior to the experiment. Alternatively, we could define strength as the number of minutes the ego called the alter (i.e. the other way around). Table A-9 shows the results of this analysis. We observe that the findings are consistent with those presented in Table 11 of the main document; the treatment effect is stronger for those connections with stronger ties.

	Outbound Minutes	
	Total	Total (exc. Ego)
Treatment	0.0984***	0.101***
	(0.0227)	(0.0226)
Tie strength	−0.0305*	−0.027
	(0.0178)	(0.0176)
Tie strength * Treatment	0.0843***	0.0902***
	(0.0231)	(0.0228)
Constant	−0.843***	−0.832***
	(0.0304)	(0.0303)
Week dummies	Yes	Yes
Observations	27,598	27,598

Long-term effects on alter usage. *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses. Tie strength is operationalized as the number of minutes the ego called the alter before the intervention.

Table A - 9: Long-term effect of treatment on usage (weeks 7-12 after the treatment) moderated by tie strength.

WEB APPENDIX – A7: CALCULATING THE FINANCIAL INCREMENTAL VALUE

In this web appendix we provide the details behind the profit calculations presented in Section 4.2 of the main document. Given that we did not observe profitability figures for the customers in our sample, we made certain assumptions to transform the usage metrics into profitability. We assume that all customers make phone calls at the average level of consumption in absence of the marketing campaign. Furthermore, based on information provided by the firm, we assume that the average consumption levels pre-campaign corresponds to a weekly average revenue of $5 (~$20 a month). Based on these assumptions and the model estimates, we estimate the revenue generated by an average alter for a period of 12 weeks, consistent with our data window.

In order to transform revenue into profitability and to aggregate 12 periods into a single metric, we further made assumptions about operating margins of our data provider and a reasonable discount factor. We assume a weekly discount factor of 0.27% (~15% annually) and an operating margin of 15%, which is approximately the average rate for telecom services.[4]

Table A - 10 shows the calculations used to estimate the financial incremental value of the CRM campaign (Section 4.2 of the main document). The revenue for the treatment condition is proportional to that of the control condition, but amplified by the estimated effect of the treatment both in the short (Table 8) and in the long term (Table 9). The retention figures are computed weekly, by transforming the odds ratios from the churn models presented in Tables 8 and 9, and then accumulated over time. Based on our model estimates, and the assumptions discussed above, the incremental value of each alter due to the targeted campaign is $0.85 for the 12 weeks following the intervention.

			Control alter			Treatment alter		
	Week	Margin	Revenue	Retention	Discounted Profit*	Revenue	Retention	Discounted Profit
Short-tem	1	0.15	$5.00	99.7%	$0.75	$5.42	99.7%	$0.81
	2	0.15	$5.00	99.3%	$0.74	$5.42	99.3%	$0.80
	3	0.15	$5.00	99.0%	$0.74	$5.42	99.0%	$0.80
	4	0.15	$5.00	98.6%	$0.73	$5.42	98.6%	$0.79
	5	0.15	$5.00	98.3%	$0.73	$5.42	98.3%	$0.79
	6	0.15	$5.00	98.0%	$0.72	$5.42	98.0%	$0.78
Long-term	7	0.15	$5.00	97.6%	$0.72	$5.54	97.8%	$0.80
	8	0.15	$5.00	97.3%	$0.71	$5.54	97.6%	$0.79
	9	0.15	$5.00	97.0%	$0.71	$5.54	97.5%	$0.79
	10	0.15	$5.00	96.6%	$0.71	$5.54	97.3%	$0.79
	11	0.15	$5.00	96.3%	$0.70	$5.54	97.2%	$0.78
	12	0.15	$5.00	96.0%	$0.70	$5.54	97.0%	$0.78
				Sum	$8.65		Sum	$9.50
							Incremental value	**$0.85**

*Discounted profit $= \frac{m \times rev \times ret}{d^{\text{week}}}$, where *m* denotes margin, *rev* denotes revenue, *ret* denotes retention, and *d* denotes discounted rate, assumed to be 0.27% which corresponds to a 15% annual discount rate.

Table A - 10: Calculations for the financial incremental value of the CRM campaign

[4] See http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/margin.html (last accessed: March 2016) for margins across various industries.

References

Germann, Frank, Peter Ebbes and Rajdeep Grewal (2015), "The Chief Marketing Officer matters!," *Journal of Marketing*, 79(3), 1-22.

Hoechle, Daniel (2007), "Robust standard errors for panel regressions with cross-sectional dependence." *Stata Journal*, 7(3), 281.

Newey, Whitney K. (1987), "Efficient estimation of limited dependent variable models with endogenous explanatory variables," *Journal of Econometrics*, *36*(3), 231-250.

Rossi, Peter E. (2014), "Even the rich can make themselves poor: A critical examination of IV methods in marketing applications," *Marketing Science*, 33(5), 655-672.

Wooldridge, Jeffrey M. (2002), Econometric analysis of cross section and panel data, MIT Press, Cambridge Massachusetts