Supporting Information

Hersh 10.1073/pnas.1315043110

SI Text

This section outlines: (*i*) how the victims were identified in public records before September 11, 2001 (hereafter 9/11) ; (*ii*) how the family members and neighbors of victims were identified in public records before 9/11 and again in 2013; and (*iii*) how a comparison group was generated to serve as a baseline against which to analyze changes in the political behaviors of families and neighbors between 2001 and 2013. The section also provides additional information about balance between treatment and control groups and on alternative estimation strategies for the figures in the report.

Identifying Victims in 2001-Era Records

I secured a database from the political data firm Labels & Lists that contained records of all New York state registered voters as of the summer of 2001. Before the implementation of the Help America Vote Act of 2002, which required states to generate statewide digital voter registration databases, political data vendors like Labels & Lists compiled local voter registration records so that their campaign clients could communicate with voters. Labels & Lists, like other data vendors, attempts to clean and standardize voter registration data coming from election offices. Labels & Lists has retained obsolete snapshots of voter registration records going back in time, and they sold a copy of their complete New York State voter registration database current as of summer 2001. Of the 9,995,513 records on the statewide New York voter file, the first step of the research plan was to identify the 9/11 victims who were residents of New York and registered to vote. I did not search for victims residing outside of New York because of data and cost limitations.

I compiled a list of 9/11 victims from public sources. I focused exclusively on victims who were residing in New York and who were killed in the attacks. (There is no public list of injured victims, and the Department of Justice will not release details about injured victims to protect these individuals' privacy.) The *New York Times* and *CNN*, among other outlets published short obituaries of victims. The Web site Legacy.com (www.legacy.com/Sept11/Home.aspx) lists obituaries from the *New York Times* as well as obituaries from other sources. All victims are publicly listed with their last and first name, often with middle name or middle initial, and with nicknames when applicable. Victims' ages at time of death are almost always available as well. [Of the 1,729 victims who resided in New York, only 16 (<1%) were not listed with an age.] All victims were also listed with their employer and their city of residence.

Identifying the 9/11 victims in the 2001 voter registration file required multiple sweeps at matching records. In the first sweep, I focused on victims who resided outside of New York City. For these victims, the first step was to match them based on the first name, last name, city, and birth year. If a victim matched on all of these variables, they were considered a match in the initial sweep. Of the victims residing outside of New York, 61% were found in the voter registration system through this initial match. These records, as with all records in the analysis, were inspected individually for evidence of false positive matches.

For the victims who lived outside New York City who did not match based on first name, last name, birth year, and city, I searched for them individually in the entire statewide voter file, not just in their city of residence. Some victims were registered in a different city; some were found in the same city but initially failed to match because of idiosyncrasies in naming conventions. For example, some O'Neills are listed with an apostrophe but others are not; some individuals with hyphenated surnames are listed with the hyphen, some without the hyphen, and some are listed just by a maiden or married name; some victims were listed with a nickname (e.g., Robert as Bob). The number of victims remaining after automated matching was sufficiently small that, although time-consuming, it was possible to search for victims case by case.

In some cases, during the manual look-up in which I looked for matches beyond the victims' listed city of residence, multiple registrants appeared as if they could be likely matches, based only on birth year and name. In these cases, I was able to determine the correct victim by looking for their family members who were also registered voters. Family members' names are typically listed in obituaries. If the family names in the obituaries matched the family names in the voter file, this was clearly a match. In some other cases, I used exact date-of-birth information, typically accessed from grave information on findagrave.com. (Findagrave.com allows the public to enter information about the deceased. Because the sources of the information on findagrave.com are not vetted in the same way as *New York Times* obituaries are likely vetted, I only used the birthdate information on findagrave.com to distinguish between multiple plausible matches, and not as part of the automated matches.)

For residents of New York City, the matching methodology was slightly different. The approach was different because New York City victims were not identified by borough or neighborhood, and thus their city of residence was not particularly informative in finding them on the voter file. I started with two matching sweeps, one based on exact first name, last name, and birth year, and a second based on first name, middle initial, last name, and birth year. The common problem with using middle names or initials in a matching procedure like this one is that when filling out forms—like a voter registration form—people sometimes use their middle name, sometimes their middle initial, and sometimes neither. Thus, although a middle initial can help us locate the correct victim, it might also prevent one from finding a victim who inconsistently reports his or her middle name.

After matching on first name, last name, and age, and first name, middle initial, last name, and age, the next step was to remove false-positives from the list. Unlike in the matching of non-New York City victims, which produced very few false-positives because residents were already situated in small cities and towns, the initial New York City match did produce a number of false-positives. For example, there may have been multiple John H. Smiths of the same age in the entire city of New York. For each case of multiple matches, I went case-by-case to determine which was the correct John Smith and which were false-positives. I did this by looking at the other family members listed in the household and by looking up exact date-of-birth data from findagrave.com. Similar to non-NYC victims, I then compiled a list of New York City victims who were not matched through the automated matching and I looked for them one at a time. I looked for them on the entire statewide file and identified idiosyncrasies in name patterns (like apostrophes, hyphens, and nicknames) that may have caused a nonmatch.

In total, 1,181 of 1,729 (68%) New York state victims were identified on the 2001 voter registration file. This is the proportion we would expect to find based on rates of participation in the voter registration system. According the Census, 68% of New York citizens were registered to vote in the 2000 election (see table A-5 in ref. 1). Although the matching procedure may have generated undiscovered false-positives and false-negatives, the proportion of victims identified on voter rolls compares favorably with estimated rate of registration in the state.

Identifying Family Members and Neighbors in 2001 and 2013 Records

For each victim, I found the associated family members and neighbors who were registered voters in 2001. To identify family members, I used a family identification number generated by Labels & Lists that groups individuals together who live in the same household and are thought to be members of the same family. The median victim had one other family member associated with them, but the number of family members ranged from zero (for those victims who lived alone) to six.

The goal in identifying the neighbors of 9/11 victims was to isolate a group of individuals who had a high probability of personally knowing a victim of the attacks. Of course, not everyone knows all of their residential neighbors, and some people know none of their neighbors, but it is uncontroversial that, compared with the typical New Yorker, an immediate neighbor of a 9/11 victim is more likely to have personally known a victim.

To isolate a group of neighbors for each victim, I used the following methodology. First, I composed a database of all registered voters who, in 2001, lived on the same Census block and on the same street as a victim. In urban areas where most of the downstate New York population lives, a Census block is typically the same as an actual block. Thus, the first step in identifying neighbors is to focus on neighbors who lived on the same block and on the same street. Some people may be more familiar with neighbors who live on the same street but on the other side of the street (and thus in a different block), but the data do not allow for easy identification of these same-street-different-block residents. This aspect does not affect the probability that the same-street-same-block residents are likely to know their neighbors.

The next step in identifying neighbors is to distinguish victims who live in apartment buildings versus those who live in houses. For those who live in houses (identified because no unit number is listed in the voter registration records), I further restricted the neighbors to the 10 whose house numbers were closest. For example, if the victim lived at 25 Main Street, and there were two registered voters each at numbers 21, 23, 27, and 29 Main Street, then these were the neighbors included in the analysis, even if there were additional registered voters on the block at 19 Main Street and 31 Main Street. If a victim lived in a house and the house was situated on a street that had both houses and apartment buildings, I further restricted his or her neighbors to just those who also lived in houses. The goal of these restrictions was to limit the individuals considered neighbors to those who could be identified based on available geographic information and who were most likely to have had regular contact with victims.

Because many of the victims of 9/11 were residents of high-density areas, many victims lived on blocks with multiple large apartment buildings. For those victims identified with an apartment unit, I used a different strategy to estimate their nearest neighbors from the strategy used for victims living in houses. First, for victims living in an apartment, I restricted their neighbors to those who lived in the same building (i.e., the same number address). Second, for victims living in large buildings (here defined as buildings with more than 25 registered voters living in the building), I restrict neighbors to those who live in the same building and on the same floor as the victim. To accomplish this, I examined each victim's apartment building address to identify the labeling convention of units. If, for example, the victim lived in apartment 45, I restricted the neighbors to those who lived in units labeled 40 through 49. If a building had fewer than 25 registered voters, I considered all registrants in the building to be neighbors. There was a small number of victims who lived in large apartment buildings (i.e., more than 25 people registered at the address), but who were not listed with an apartment unit number on the voter file. Using the same threshold of 25 units, I only consider them in the neighbor analysis if fewer than 25 people lived at the same address. In the few cases where a victim was not identified with a unit number but resided in a very large apartment building, I excluded them from the analysis of neighbors because their nearest neighbors were not identifiable.

In total, there were 1,220 family members and 9,632 neighbors of 9/11 victims listed in the 2001 voter file. For each of these registered voters, I generated a file incorporating their first, middle, and last names, their birthdates, sex, and their addresses as of 2001. In July 2013, I transmitted this file to a political data firm, Catalist. Catalist maintains a national database of all registered voters, updated continually. I contracted with Catalist to match the 2001 voters to their current registration profiles. Catalist searched for families and neighbors not only in New York, but also in every other state, to capture individuals who moved across state lines between 2001 and 2013. I contracted with Catalist for this task because their matching algorithm has been validated for its accuracy (2) and because by using commercial sources and National Change of Address data, Catalist is equipped to identify individuals who have moved residences.

Of the 1,220 family members whose names were sent to Catalist, Catalist identified 1,095 (90%) of them in their national database. Of these family members, 87% were registered to vote in 2013, 10% were registered at some point since Catalist began collecting records (the company began collecting data in 2006) but were unregistered in 2013, and the remaining 3% were never registered since Catalist began collecting voter registration data. Of the victims' 9,632 neighbors sent to Catalist, Catalist identified 86% of them in their national database, 84% of whom were registered to vote in 2013, 13% were registered at some point between 2006 and 2013 but not in the summer of 2013, and the remainder of whom were not registered since the firm began collecting data.

Generating a Comparison Group

To estimate the change in political behavior among victims' family and neighbors, it is necessary to assess their change relative to a baseline. The baseline is an estimate of the population of citizens from which victims' families and neighbors are drawn. To generate a baseline, or control group, I used the following methodology. For each victim, I use a combination of exact matching and distance matching to find up to five "control victims" whose observable traits are as close as possible to the observable traits of the real victims. All variables that are used in matching come from the 2001 New York voter file and are thus pretreatment variables. The variables used in exact matching are party affiliation (Democratic, Republican, or Other), vote history in the 2000 and 1998 federal general elections (voted, abstained, ineligible), an indicator variable representing whether the registrant has ever voted in any party primary between 1992 and 2000, sex (male, female), and state legislative district, of which there are 150 in New York and are of approximately equal population.

Within the stratum of voters who fit a victim exactly on party, 2000 vote history, 1998 vote history, primary voting participation, sex, and district, I then used Mahalanobis distance matching to search for up to the five best matches on a number of cardinal and continuous variables. [I implemented the selection of control units through the STATA package mahapick (3).] The variables used in distance matching are Census block-group median household income, Census block-group percentage non-Hispanic white, year of birth, number of family members registered in the household of the victim, and total number of elections in which the victim participated since 1992.

For 1,171 of the 1,181 victims, the matching procedure generated five control victims. In the remaining 10 cases, there were one to four matches in two cases and zero matches in eight cases. The procedure would fail to generate five matches if the exact matching criteria left an insufficient number of plausible matches. In a few cases, the matching procedure resulted in a victim being matched to a member of the victim's own family or neighbor. A typical instance of this would be if a victim who died in 9/11 had a sibling in the same household who had similar attributes as the victim. I excluded from the control group the 66 control victims and 132 control family members (2% of all control matches) who were themselves true family members or true neighbors. Using the same procedures as used for the true victims, for each control victim I identified the person's family members and nearest neighbors.

In total, there were 5,851 control family members and 49,095 control neighbors of 9/11 control victims listed in the 2001 voter file. Of the control family members whose names were sent to Catalist, Catalist matched 89% of them to their database, 88% of whom were registered to vote as of July 2013. Of the control neighbors sent to Catalist, 86% were found in Catalist's database, 85% of whom were registered to vote as of July 2013. These numbers parallel the matching statistics for the true victims' families and neighbors, as would be expected given a successful formation of a control group.

Balance Between Treatment and Control

In Table S1, I report the means and SEs of traits for true victims versus control victims, true family members versus control family members, and true neighbors versus control neighbors. As is evident, given the large number of observations available for generating a control group, the treatment and control groups are well-balanced on pretreatment observables.

In Table S2, I show balance on an additional three variables, including Census block-group measures of percent African-American, percent of residents on welfare, and percent of homes valued at more than $500,000. What distinguishes these variables from those in Table S1 is that these were not incorporated into the matching algorithm. The balance achieved on these variables lends credibility to the assumption that the matching method generated a control group similar to the treatment group. Remaining differences between treatment and control are dealt with by controlling for pretreatment covariates in a regression framework.

Three other pieces of evidence presented in the balance tables and in the body of the text provide additional assurance of balance between treatment and control groups. First, Table S1 shows that in two primary elections, there is balance between treatment and control. Unlike the federal general elections, the primary elections were incorporated into the distance matching part of the process, not exact matching. Even so, there is balance on these primary elections. Second, note that all of the matching was based on the attributes of victims, not based on families and neighbors. Among the variables used in matching is the count of family members in the victims' household and neighborhood characteristics, but measures like party affiliation, age, and sex were matched victim to control victim. The balance on these variables for families and neighbors, despite families and neighbors not explicitly being matched to one another, suggests the groups are comparable.

Most importantly, the other piece of evidence indicative of balance is Fig. 2, which measures donations. Unlike the other analyses, which measure vote history and party affiliation, nothing about donor behavior was used in the matching algorithm. Yet, before 9/11 victims' families were no more likely to make political contributions than control families. Only after 9/11 did the families' behavior change. Because the matching method did not constrain the pool of donors at all, the pre-9/11 donation behavior of treatment and control groups provides, arguably, the cleanest test that balance was achieved.

Note on Campaign Contribution Data

The 2001-era data on victims' families and neighbors was matched to campaign contribution data by the researcher, not through a data vendor. Political Scientist Adam Bonica has previously cleaned and organized campaign contribution data from the Federal Election Commission (FEC) going back many years (4). Bonica generously shared his data files for New York donors. Because old records of donations are not purged from FEC donor archives, as registered voters are sometimes purged from voter registration records, the data reported in Fig. 2 go back all of the way to 1992 rather than just to 1998.

Much less information is reported about donors in FEC records than is reported about registered voters in voter files. In particular, donors' birthdates and exact addresses are not listed. As a result, I matched the families and neighbors of true victims and control victims to the FEC data based only on first name, last name, and zip code. Because political donations are only made by a small fraction of American citizens (typically, only contributions over $250 are reported in the public record) and because the sample of control and treatment groups from the 9/11 data are relatively small, the incidence of false-positive matches is likely low. The Catalist-matched data revealed that true families moved residences since 2001 at only slightly higher rates than control families (31% are listed with a different zip code in 2013 than 2001 compared with 25% for control families). This finding likely only biases the result downward. To be more precise, one could simply interpret the result in Fig. 2 as a treatment effect conditional on research subjects (control and treatment) living in the same zip code in 2013 as in 2001.

Regression-Based Estimates

The figures in the text show simple difference of means, with 95% confidence intervals about those means, between treatment groups and control groups. As discussed above, the control groups were designed to mimic the treatment group on pre-9/11 demographics, and as Tables S1 and S2 show, the control families and neighbors are nearly identical to the true families and neighbors on an array of observable variables. As such, difference-of-means tests may be sufficient.

Nevertheless, to account for remaining imbalance between treatment and control, and to account for clustering of errors, I now replicate the entire analysis in a regression framework. As control variables, I used pre-9/11 records of sex; Census block-group median household income; Census block-group percentage white; the count of registered voters associated with the household; binary variables for age groups under 30 (the baseline group), 30–40, 40–60, and older than 60; and binary variables for Democrats, Republicans, and Independents (the baseline group). I also constructed a variable that measures the percentage of all general, primary, special, and local elections a person voted in before 9/11, conditional on the person being old enough to be eligible. Note that this variable will be excluded in estimates of pre-9/11 voting in Figs. S1 and S3, because the dependent variables in those estimates are measures of pre-9/11 turnout.

With these control variables, I show estimates for the treatment effect using ordinary least-squares (OLS) regression. I clustered the SEs around the true victims. Recall that each family member and neighbor, and each control family member and neighbor, can be associated with a true victim of 9/11. Rather than clustering control families and neighbors around their associated control victims and clustering true families and neighbors around their associated victim, I used a more conservative estimate of SEs by clustering all control and treatment groups based on the true victim with whom they are associated. Note that in the regression-based estimates, I also used weights to account for the fact that for 10 victims of 9/11, I was able to find one to four control victims rather than five control victims. The weights bring the families and members associated with these 10 victims in line with the other victims. However, because only 10 victims had fewer than five matches, this correction does not substantively affect the results.

Figs. S1–S5 replicate the results of Figs. 1–5 in the main analysis. Instead of showing differences of means, these figures show the regression-based treatment effect. Accounting for pre-9/11 variables that mitigate any imbalances between treatment and control group, and accounting for the clustering of errors, these figures show the difference between treatment and control groups for each of the phenomena studied. Because the treatment and control groups are well-balanced, there are no sizeable differences in point estimates between Figs. 1–5 and Figs. S1–S5. In some cases (Figs. S3 and S4), a couple of key estimates have 95% confidence intervals that barely overlap with zero. This result is because I used SEs clustered around the true victims, a conservative method that enlarges the SEs. Nevertheless, the evidence from the regression-based estimates in Figs. S1–S5 is consistent with the difference-in-means displayed in Figs. 1–5.

1. US Census Bureau (2012) *Population Report, November 2012 and Earlier Years* (US Census Bureau). Available at www.census.gov/hhes/www/socdemo/voting/publications/historical/index.html. Accessed July 15, 2013.
2. Ansolabehere S, Hersh E (2012) Validation: What survey misreporting reveal about survey misreporting and the real electorate. *Polit Anal* 20(4):437–459.
3. Kantor D (2012) *Mahapick* (STATA Module). Available at http://ideas.repec.org/c/boc/bocode/s456703.html. Accessed July 15, 2013.
4. Bonica A (2013) Ideology and interests in the political marketplace. *Am J Pol Sci* 57(2):294–311.

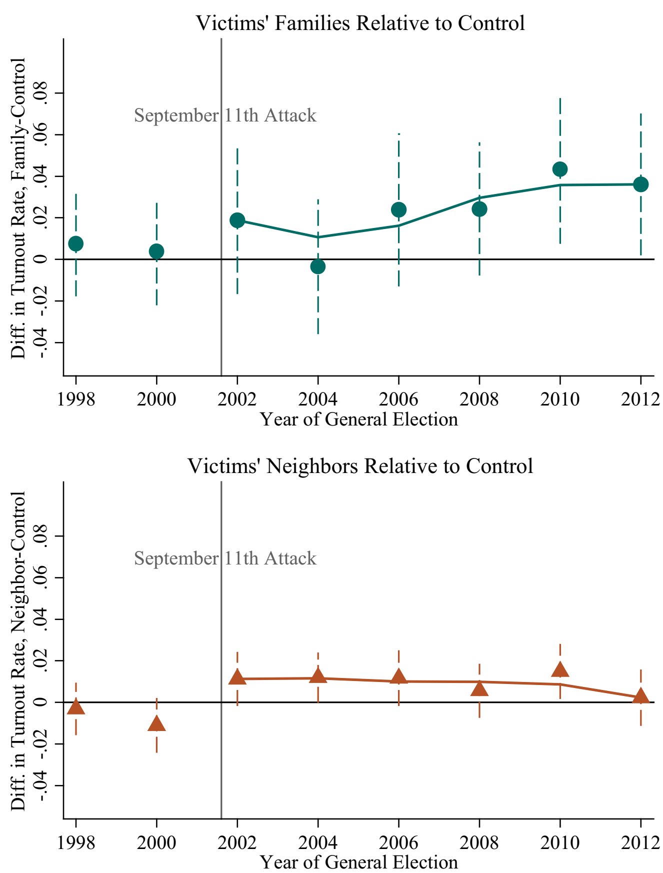

Fig. S1. Regression-based estimate for Fig. 1. Each dot represents a treatment effect, with a 95% confidence interval, from a multivariate OLS regression estimated with clustered SEs, of the same analysis shown in Fig. 1.

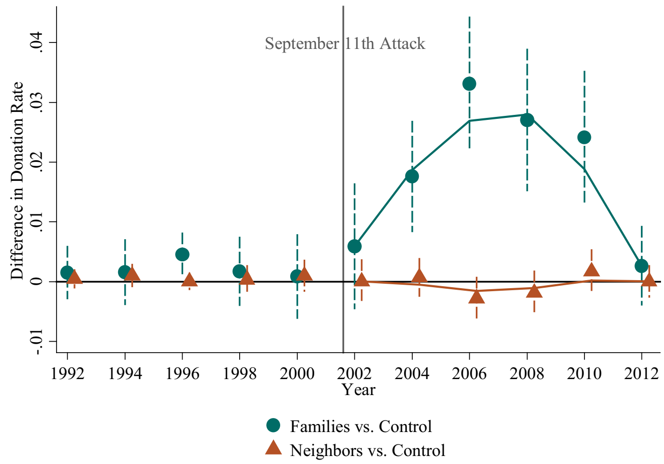

Fig. S2. Regression-based estimate for Fig. 2. Each dot represents a treatment effect, with a 95% confidence interval, from a multivariate OLS regression estimated with clustered SEs, of the same analysis shown in Fig. 2.

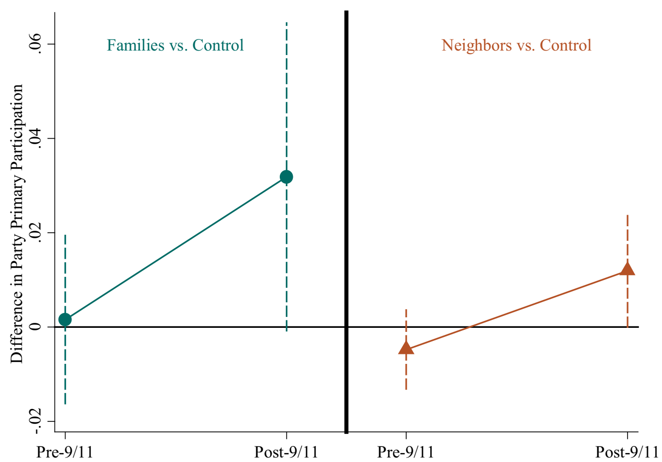

Fig. S3. Regression-based estimate for Fig. 3. Each dot represents a treatment effect, with a 95% confidence interval, from a multivariate OLS regression estimated with clustered SEs, of the same analysis shown in Fig. 3.

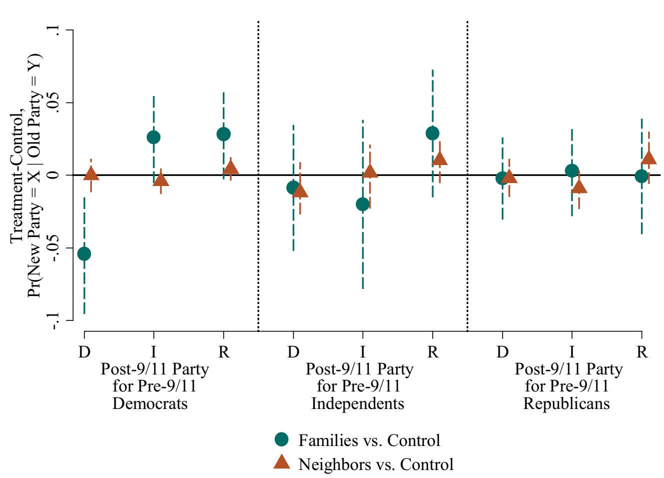

Fig. S4. Regression-based estimate for Fig. 4. Each dot represents a treatment effect, with a 95% confidence interval, from a multivariate OLS regression estimated with clustered SEs, of the same analysis shown in Fig. 4.

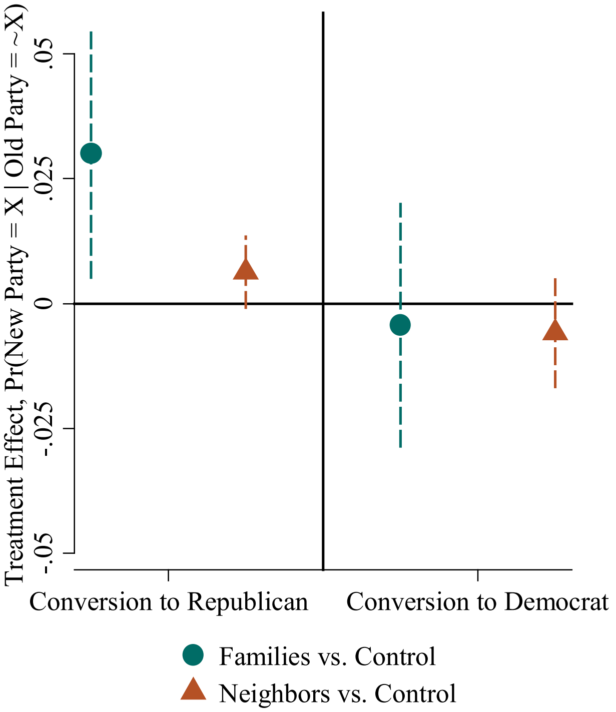

Fig. S5. Regression-based estimate for Fig. 5. Each dot represents a treatment effect, with a 95% confidence interval, from a multivariate OLS regression estimated with clustered SEs, of the same analysis shown in Fig. 5.

Table S1. Covariate balance, September 11, 2001 victims, families, and neighbors versus comparison groups

	Victims		Family		Neighbors	
Variables	Treatment	Control	Treatment	Control	Treatment	Control
P(Democrat)	0.36 (0.014)	0.36 (0.006)	0.37 (0.014)	0.38 (0.006)	0.46 (0.005)	0.47 (0.002)
P(Republican)	0.39 (0.014)	0.39 (0.006)	0.36 (0.014)	0.37 (0.006)	0.29 (0.005)	0.28 (0.002)
P(vote G00)	0.76 (0.012)	0.76 (0.006)	0.74 (0.013)	0.74 (0.006)	0.67 (0.005)	0.68 (0.002)
P(vote G98)	0.38 (0.014)	0.38 (0.006)	0.36 (0.014)	0.36 (0.006)	0.36 (0.005)	0.36 (0.002)
P(vote P00)	0.05 (0.007)	0.06 (0.003)	0.06 (0.007)	0.06 (0.003)	0.07 (0.003)	0.08 (0.001)
P(vote P98)	0.04 (0.006)	0.05 (0.003)	0.06 (0.007)	0.06 (0.003)	0.08 (0.003)	0.08 (0.001)
Age	40.62 (0.289)	40.87 (0.136)	43.34 (0.431)	43.45 (0.200)	47.81 (0.175)	47.45 (0.077)
P(female)	0.22 (0.012)	0.22 (0.005)	0.67 (0.013)	0.64 (0.006)	0.54 (0.005)	0.54 (0.002)
P(White in BG)	0.70 (0.008)	0.70 (0.004)	0.75 (0.007)	0.75 (0.003)	0.69 (0.003)	0.70 (0.001)
BG median household income ($)	65,851 (920)	65,414 (411)	70,762 (933)	70,314 (421)	65,769 (330)	65,163 (142)
Count of family members	2.04 (0.031)	2.11 (0.017)	3.15 (0.036)	3.27 (0.022)	1.98 (0.010)	2.04 (0.005)
Observations	1,181	5,809	1,220	5,851	9,632	49,095

SEs are in parentheses. All statistics are based on the summer 2001 voter registration file. The observation counts are slightly lower for vote history and block-group statistics as a result of some individuals being ineligible for all elections (i.e., they registered after the election took place) in the first case and because their address was not matchable to a Census block group in the second case. BG stands for Census Block Group. Block Group median income is represented in dollars. On vote history variables, G represents federal general elections and P represents federal primary elections, followed by a designation of the year.

Table S2. Additional covariate balance, September 11, 2001 victims, families, and neighbors versus comparison groups

	Victims		Family		Neighbors	
Variables	Treatment	Control	Treatment	Control	Treatment	Control
P(Black in BG)	0.09 (0.006)	0.09 (0.003)	0.07 (0.005)	0.07 (0.002)	0.10 (0.002)	0.10 (0.001)
P(on welfare in BG)	0.04 (0.001)	0.04 (0.001)	0.03 (0.001)	0.03 (0.001)	0.04 (0.001)	0.04 (0.000)
P(house worth > 500K)	0.11 (0.007)	0.11 (0.003)	0.11 (0.007)	0.11 (0.003)	0.11 (0.003)	0.11 (0.001)
Observations	1,117	5,809	1,217	5,851	9,610	49,095

SEs are in parentheses. All statistics are based on the summer 2001 voter registration file. BG stands for Census Block Group.