

Unexpected Event during Surveys Design: Promise and Pitfalls for Causal Inference *

Jordi Muñoz Albert Falcó-Gimeno Enrique Hernández
University of Barcelona University of Barcelona Universitat Autònoma de Barcelona

Accepted for publication at *Political Analysis*

Abstract

An increasing number of studies exploit the occurrence of unexpected events during the fieldwork of public opinion surveys to estimate causal effects. In this paper we discuss the use of this identification strategy based on unforeseen and salient events that split the sample of respondents into treatment and control groups: the Unexpected Event during Surveys Design (UESD). In particular we focus on the assumptions under which unexpected events can be exploited to estimate causal effects and we discuss potential threats to identification, paying especial attention to the observable and testable implications of these assumptions. We propose a series of best practices in the form of various estimation strategies and robustness checks that can be used to lend credibility to the causal estimates. Drawing on data from the European Social Survey we illustrate the discussion of this method with an original study of the impact of the Charlie Hebdo terrorist attacks (Paris, 01/07/2015) on French citizens' satisfaction with their national government.

*This research has received financial support from the Spanish Ministry of Science, Innovation, and Universities through the research grant CSO2017-89085-P. We are grateful to Lucas Leeman, Silja Häusermann, Macarena Ares, Guillem Rico, Salomo Hirvonen, the editor and anonymous reviewers of *Political Analysis*, and participants at the University of Zurich IPZ publication seminar, the Autonomous University of Barcelona DEC Seminar, and the 2018 EPISA conference for helpful comments and suggestions. We are especially grateful to Erik Gahner Larsen for his detailed and insightful comments, and for generously sharing with us his collection of UESD references. Enrique Hernández also thanks the University of Zurich Political Science Department for hospitality during the spring of 2018. Replication materials are available at the *Political Analysis Dataverse*: doi.org/10.7910/DVN/RDIIVL (Hernández et al., 2019).

1 Introduction

Natural disasters, violence outbreaks, political scandals or sports results are examples of unexpected events that can occur during the fieldwork of public opinion surveys. Survey researchers traditionally considered that a changing context could threaten the validity of their measurement instruments. Other scholars, though, have identified these occurrences as an opportunity to identify the causal effect of these events on public opinion.

Already in 1969 *The Public Opinion Quarterly* published a series of articles aimed at studying the effect of unexpected events such as Martin Luther King’s assassination (Hofstetter, 1969) or an explosion in Johannesburg (Lever, 1969) on multiple outcomes. More recently, the influence of the so-called *identification revolution* has led to a significant increase in the number of publications that exploit these unexpected events as opportunities to conduct design-based studies that can yield valid causal estimates of theoretically relevant shocks.

In this paper we discuss the strengths and limitations of this identification strategy, which we call the “Unexpected Event during Surveys Design” (UESD). In particular, we focus on the assumptions under which unexpected events can yield valid causal estimates. We discuss these assumptions and their observable and testable implications in detail. Next, we present several strategies to address potential violations of the assumptions, as well as a series of robustness checks that can be used to assess and establish their plausibility. Our review of over 40 published studies that rely on this identification strategy shows wide variation in the estimation strategies, tests and robustness checks that are performed to substantiate causal claims. While most researchers acknowledge some of the key identifying assumptions of this design, there is still no apparent consensus on the empirical strategy and the robustness checks that should be performed. Our aim is that this paper will help set a standard of good practices that should increase the credibility of causal claims based on the UESD.

We illustrate the strengths and weaknesses of the UESD through an original study of the impact of the Charlie Hebdo terrorist attacks (Paris, 01/07/2015) on individuals’ satisfaction with government in France, relying on the 7th round of the European Social Survey (ESS), that was being fielded at the time of the attacks.

2 The Unexpected Event during Surveys Design

We define the UESD as the research design that exploits the occurrence of an unexpected event during the fieldwork of a public opinion survey to estimate its causal effect on a relevant outcome by comparing responses of the individuals interviewed before the event $t_i < t_e$ (control group) to those of respondents interviewed after the event $t_i > t_e$ (treatment group).

This identification strategy relies on the assumption that the moment at which each respondent is interviewed during the fieldwork of a given survey is independent from the time when the unexpected event occurs. The occurrence of the event should, therefore, assign survey respondents into treatment and control groups as good as randomly. Hence, a UESD study relies on the use of cross-sectional survey data in which time provides exogenous variation in a substantively relevant variable, such as exposure to some experience or information. Formally, the timing of the survey t is exploited as an instrument of exposure to the event T the effects of which the researchers are interested in.

A crucial feature of the UESD is, therefore, that the event occurs unexpectedly, since predictable events might lead to violations of the key assumptions of this design (excludability and ignorability). Foreseeable events (e.g. an election result) might be anticipated by respondents, who might change their attitudes in response to the event even before it has occurred. This could be especially problematic if certain respondents are more/less likely to anticipate the event and they are, at the same time, more/less likely to be interviewed after the event. Moreover, if an event is foreseeable some respondents might postpone their participation in the survey and self-select into the treatment group.¹

This identification strategy has been frequently exploited to study the effects of important phenomena that cannot be directly manipulated through controlled experiments. This is for example the case of violent events such as the assassination of political leaders (Boomgaarden and de Vreese, 2007), or terrorist attacks (Boydston et al., 2018; Silva, 2018; Balcells and Torrats-Espinosa, 2018; Coupe, 2017; Dinesen and Jæger, 2013; Finseraas and Listhaug, 2013; Geys and Qari, 2017; Jakobsson and Blom, 2014; Legewie, 2013; Metcalfe et al., 2011; Perrin and Smolek, 2009) Researchers have also analyzed the effects of other theoretically relevant

¹For example, in online surveys respondents might wait until the event has occurred to complete to questionnaire, and in face-to-face or telephone surveys respondents might delay their interview (by refusing or making an appointment for a later date)

events such as political scandals (Ares and Hernández, 2017; De Vries, 2018; Solaz et al., 2018), protests and demonstrations (Branton et al., 2015; Motta, 2018; Silber Mohamed, 2013; Zepeda-Millán and Wallace, 2013), presidential candidates speeches (Flores, 2018), Nobel prize awards (De Vries, 2018), the announcement of election results and election fraud (Pierce et al., 2016), policy reforms (Larsen, 2018), or flu epidemics (Jensen and Naumann, 2016).²

Some of these events commonly analyzed through the UESD such as terrorist attacks should *a priori* be unexpected for most respondents. However, there are others that might be more foreseeable like for example election results or a scheduled legal demonstration. In fact, the same type of event might be unexpected in some contexts but not in others. For example, while in some cases the victory of a particular party or candidate in an election might be expected (e.g. CDU victory in the 2017 German election), in some others the election results might come as a surprise by being more unexpected (e.g. the victory of Donald Trump in 2016 analyzed by Minkus et al. (2018)). Researchers should, therefore, make a compelling argument about the plausibility of unexpectedness on a case-by-case basis.

The increasing popularity of the UESD can be attributed to its capacity to provide credible estimates of the causal effects of unexpected events on theoretically relevant outcomes. This capacity is based on three crucial features of this design. First, in contrast to standard observational survey studies, exogenous assignment to *treatment* and *control* groups increases the internal validity of the estimates, as it can shield them from biases related to unobserved confounders or reverse causality. Second, the fact that the UESD relies on *naturally* occurring events (not manufactured by researchers) provides a level of external validity that goes well beyond what controlled experiments can offer. Third, in contrast to most analyses of the impact of events that use longitudinal data or different surveys stacked around the date of an event, UESD studies employ a single survey to focus on the immediate period of time surrounding the event, which ensures that the impact of other macro-level confounders is less of a concern. However, the absence of true random assignment and the lack of control by the researchers also pose challenges to identification that need to be addressed upfront if this strategy is to be used to substantiate causal claims.

²See Section B in Supporting Information for an overview of 44 published UESD studies

2.1 The UESD in context

The UESD shares some properties with other research designs commonly used in social sciences, such as the Regression Discontinuity Design (RDD), the Interrupted Time Series (ITS) design or event studies, which are common in economics and finance. While these research designs can provide relevant cues for the use of the UESD, this design has a range of unique features and particularities that warrant some specific methodological considerations.

Under some specific conditions, the UESD might be analogous to the RDD in which an observation is treated if the forcing or running variable exceeds some predetermined cutoff. In the case of an unexpected event occurred during the fieldwork of a survey, the running variable would be the fieldwork days when interviews took place, and the cutoff the moment when the event occurred. However, there are several important differences. First, unlike the RDD, in the UESD the receipt of the treatment is not linked to a given score of the running variable in a predetermined/known way. Moreover, in most applications of the RDD the running variable will be substantively related to the potential outcomes and, therefore, inference will be restricted to the cases just below and above the cutoff. While in some UESD applications this might also be the case, in most of its empirical applications the specific day when a respondent is interviewed is not related to the potential outcomes, other than through the unforeseen event. As a consequence, the need to rely on narrow bandwidths (and the estimation of local effects) is generally less stringent.

The UESD also bears some resemblance to the ITS design (St. Clair et al., 2014). This design, commonly used in epidemiology and health policy, is based on repeated measurements of an outcome (e.g. hospital admissions) for the same units before and after a given intervention (e.g. the approval of a policy). The measurements before the intervention are used to build expectations on the outcome trend in the absence of the intervention. An effect is identified if the post-intervention observations deviate significantly from this counterfactual trend.

The core challenges to causal identification in the ITS design are related to the aggregated and times-series nature of the data: the presence of seasonality, temporal autocorrelation, and time-varying confounders. While in some cases ITS designs have been applied to the analysis of survey data, these data were recorded or used at the aggregate level (see e.g. Tiokhin and Hruschka, 2017). In these examples, the (often few) respondents interviewed each specific day

(or period) of the fieldwork are treated as a measurement of the same unit and, therefore, the data set is essentially a time-series. The reduced number of daily observations might create a problem of high variance and, depending on the organization of the fieldwork, also of bias in the measurements. In the examples we discuss in this paper the survey data is primarily analyzed at the individual level, so we do not deal with repeated measurements of the same unit, but with different observations in the pre and post event periods. This is the core difference between the ITS and the UESD designs, and it has implications for the type of concerns and threats to identification, as well as for the estimation strategies.

On similar grounds, one could also mention event studies used in finance, which analyze the behavior of firms' stock prices in response to corporate events (MacKinlay, 1997). But, again, the nature of the data used in the UESD and the comparison across different units interviewed before and after the event makes the UESD a distinct type of research design.

3 Assumptions and threats to causal identification

The identification of valid causal estimates based on the comparison of respondents interviewed before and after the day of the event t_e hinges on several identifying assumptions. Arguably, there are two key and potentially problematic assumptions in this design.

The first one is excludability: any difference between respondents interviewed before and after the event shall be the sole consequence of the event. That is, the timing of the interview t should affect the outcome variable Y only through the event T . In the case we are dealing with, together with the theoretically relevant event, other things could happen that make the pre- and post-event contexts different and, therefore, influence the respondents' outcome.

The second key assumption is temporal ignorability: for any individual, the potential outcomes must be independent from the moment of the interview. Therefore, for time t to be a valid instrument of the event T , assignment to different values of t should be independent from the potential outcomes of Y . This means that, in order to obtain an unbiased estimator, selection of the moment of the interview should be *as good as random*.³

³The Stable Unit Treatment Value Assumption (SUTVA) does also apply here. However, it is probably an unproblematic assumption. The only way one could think of interference across subjects is in the unlikely case of two individuals (i_1 and i_2) who know each other and who have discussed the content of the survey after i_1 was interviewed and before i_2 's interview (after the event).

3.1 Excludability

To estimate the causal effect of the event one must assume that the timing of the survey interview does not affect the outcome through any other channel except for the event of interest. Any factor, other than the event itself that correlates with time can potentially bias the causal estimates. We consider three potential threats to the exclusion restriction: collateral events, simultaneous events, and unrelated time trends. We also discuss another particular case in which the exclusion restriction might be violated: the endogenous timing of the event.

3.1.1 Collateral events

We define collateral events as the succession of reactions triggered by the unexpected event of interest. For example, in the case of a terrorist attack, generally the government will issue public statements and adopt specific policy responses. The opposition then will decide whether to back the government or criticize it for the security failure, and the news media will shape the width and tone of the coverage.

All these reactions and counter-reactions, rather than the unexpected event *per se*, might drive the public opinion's response. This can be seen as a problem of an imprecise treatment. Under these circumstances, it is impossible to narrowly interpret the effect as a consequence of the event itself, and it should rather be interpreted as the joint effect of the event *and* the subsequent reactions. In fact, it is often legitimate to question, counterfactually, if we would observe the same effect on Y had some of these collateral events not occurred.

In some applications the presence of collateral events might not be a problem because the event did not spur any relevant reactions, or because the reactions are a constitutive characteristic of the class of events being analyzed, or they are part of the researchers' focus of interest. However, they often pose a generalizability problem: the estimated causal effect of the event (e.g. a specific terrorist attack) may not be generalizable to the class of events of the same type (e.g. terrorism in general) had the collateral events been different. In such cases, it may be advisable to analyze additional events of the same class to average out the specificities of each individual event (see e.g. Balcells and Torrats-Espinosa, 2018, analysis of multiple terrorist attacks).

3.1.2 Simultaneous events

A potentially more troublesome situation can arise when other, unrelated events, take place at the same time. This might pose a problem of compound treatments, that violates the *Compound Treatment Irrelevance Assumption*.⁴ Applied to the UESD this assumption implies that out of all the treatments that covary with the timing of the interview, the event of interest is the only treatment that affects the potential outcomes. If, instead, there are reasons to believe that any of these other events might have had an effect on the outcome of interest, the exclusion restriction would be violated. For example, the 2016 Nice terrorist attack in France took place during Bastille Day celebrations, so any effect on a political outcome might be caused both by the attack and the national commemoration.

Unlike the case of collateral events, this problem cannot be reduced to an issue of precision, interpretation, and generalizability of the treatment effects. It is true, however, that in most existing applications of the UESD, the events analyzed were highly salient episodes. Therefore, it is often unlikely that the observed effect would have been driven by other, relatively minor events that were not nearly as salient.

3.1.3 Unrelated time trends

Another violation of the exclusion restriction can be generated by the presence of time-varying variables that, aside from the event itself, are systematically related to the outcome (temporal stability assumption). Let t be a continuous time variable, and T a dichotomous variable that takes value 1 if the respondent was interviewed after the event, and 0 otherwise. In the presence of a monotonic effect of t on Y , many arbitrary partitions of t will yield statistically significant effects in the same direction as the pre-existing trend. Time trends can exist for many different reasons, often unrelated to the unexpected event. For example, variables such as subjective well-being are related to calendar effects (Csikszentmihalyi and Hunter, 2003).

3.1.4 Endogenous timing of the event

A particular case in which some of the aforementioned threats to excludability can emerge is if the timing of the event is endogenous. Some unexpected events are human-crafted and, therefore, an actor decides when does the event take place. For example, the precise time

⁴See Keele and Titiunik (2016) and Eggers et al. (2018) for a discussion of compound treatments in the context of geographical borders and population thresholds.

at which a corruption scandal breaks in the news, or the moment at which a terrorist group perpetrates an attack are events that, although unexpected for almost all citizens, might be actually planned by politically motivated actors.

If the decision about the moment when the event occurs is endogenous to the outcome variable Y , and therefore T is correlated with the error term, the excludability assumption will be violated. Take for example the case of a partisan news media outlet. Once they receive information about a corrupt act committed by the incumbent, they might strategically manipulate the moment at which the information is disclosed in order to maximize (or minimize) damage to the incumbents' popularity. They might, for example, break the news at a time in which another, more salient event is taking place to play a diversionary strategy and reduce the effect. If this is the case, assignment to treatment and control cannot be treated as exogenous, and, most probably, either a pre-existing trend or some other events that correlate with t are having an effect on Y .

3.2 Ignorability

Formally, the ignorability assumption states that respondents' treatment status is independent of their potential outcomes $Y_i(0), Y_i(1) \perp\!\!\!\perp T_i$. However, in the UESD assignment to the *treatment* and *control* groups is not under the researchers' control and not at random. Treatment assignment is the result of a combination of the unexpected event and a set of decisions related to data collection taken by the fieldwork operative.

In simple random samples with full response all individuals would have an identical probability of being interviewed at any time during the fieldwork, before or after t_e . A suitable setting for this to hold would be a telephone survey with random digit dialing, a simple random sampling design and full response. In this case, any temporal partition of the sample would be as good as random. However, this is seldom the case in most surveys. Another favorable survey design for the ignorability assumption to hold is the rolling cross-section design, used in surveys such as the National Annenberg Election Survey (NAES) or the Canadian Election Study (CES), that explicitly guarantees that the date in which each respondent is interviewed is random (Johnston and Brady, 2002). Next we discuss how variation in survey sampling methods, fieldwork procedures, and the propensity and availability of sampled units to respond will, in most cases, lead to violations of the ignorability assumption.

3.2.1 Fieldwork organization: Imbalances on observables

Threats to the ignorability assumption often stem from complex sampling designs, such as cluster, stratified, or multi-stage samples. In these designs the primary sampling units are often contacted and interviewed sequentially. For example, in face-to-face surveys the fieldwork often follows a geographical pattern for efficiency reasons. Respondents in area A are interviewed first, then interviewers move to area B, and so on. Any correlation between subject location and time of the interview will lead to a violation of the ignorability assumption, since location can be correlated with many other respondents' characteristics that might bias the findings.⁵

Quota sampling is another strategy that can generate imbalances. Often the fieldwork goes on until a given quota is completed. Then, researchers seek to complete the remaining quotas and, therefore, any additional interview (or invitation to respond in the case of online surveys) will target the respondents with profiles that match the incomplete quotas. Often, there is a systematic pattern in the pace of quota completion, and the quotas corresponding to the types of respondents that are easier to reach and survey (such as inactive population that stays at home) will fill up earlier. If this is the case, timing of the interview will be correlated with respondents' characteristics and, therefore, splitting the sample at a specific point in time might not satisfy the ignorability assumption.

When information on how the fieldwork unfolded is available, researchers can use relevant observed characteristics of the respondents (e.g. place of residence or the variables used to set up the quotas) to relax the ignorability assumption and rely on a more plausible *Conditional Ignorability Assumption*. This assumption states that treatment status is independent of individuals' potential outcomes conditional on a set of covariates: $Y_i(0), Y_i(1) \perp\!\!\!\perp T_i \mid X_i$. This is similar to the common practice of observational studies in which statistical control of potential confounders is the only strategy to ensure *conditional ignorability*. However, in the case of the UESD the exogenous nature of the cutoff point increases the plausibility that, once we take into account the relevant observables related to how the fieldwork was organized, treatment status is orthogonal to the potential outcomes.

⁵Another potentially problematic situation emerges when the unexpected event of interest causes changes in the fieldwork, and, for example, leads to temporary suspension or delay of data collection in certain areas.

3.2.2 Unobservable confounders: *reachability*

Another threat to causal identification originates from potential imbalances between the treatment and control groups on unobservable characteristics. In the context of the UESD, this is likely to be related to the different levels of *reachability* of sampled units. Some respondents are more elusive to survey researchers and require a greater effort to be contacted. Differences in reachability are related to respondents' ease of contact and their inability or reluctance to participate. For example, those who are older and out of the labor market are likely to be interviewed early during the fieldwork period because they are more often at home and have fewer time constraints (Brehm, 1993; Stoop, 2004). Similarly those who are more interested in the topics covered in the survey should be more willing to respond when they are first contacted (Brehm, 1993). As a consequence, those who respond early to the survey might be different than those who respond later.

The extent to which these differences in respondents' reachability might pose a threat for the UESD will depend on the survey design. For example, in the case of rolling cross-section designs confounders related to reachability should be less of a concern, since this design guarantees that the day when respondents are interviewed is in principle random. In any survey in which the sampled units have any possibility to opt into the survey at different points in time though, reachability will lead to differences between those who are interviewed early and late during the fieldwork. Take as an example an online survey in which a link is distributed among all the sampled units and respondents can decide when they answer the survey (within a specified time frame). Biases related to reachability can also arise due to the common practice of making repeated attempts to recontact and interview sampled households/individuals that were not found or refused to cooperate when they were first contacted.⁶

We illustrate this point through the 2014 round of the ESS, a survey in which substitution of sampled units is not allowed and, therefore, interviewers make repeated attempts to contact non-responding units.⁷ Figure 1 summarizes the relationship between the number of attempted contacts with sampled units and the time during the fieldwork when these units were finally interviewed. In 18 out of the 21 countries included in the ESS this relationship is positive

⁶A similar situation will arise if non-responding units are not recontacted but simply substituted by units with similar characteristics (e.g. with the same gender or living in the same block).

⁷Replication materials for all the analyses conducted in this paper are available at the *Political Analysis Dataverse*: doi.org/10.7910/DVN/RDIIVL (Hernández et al., 2019).

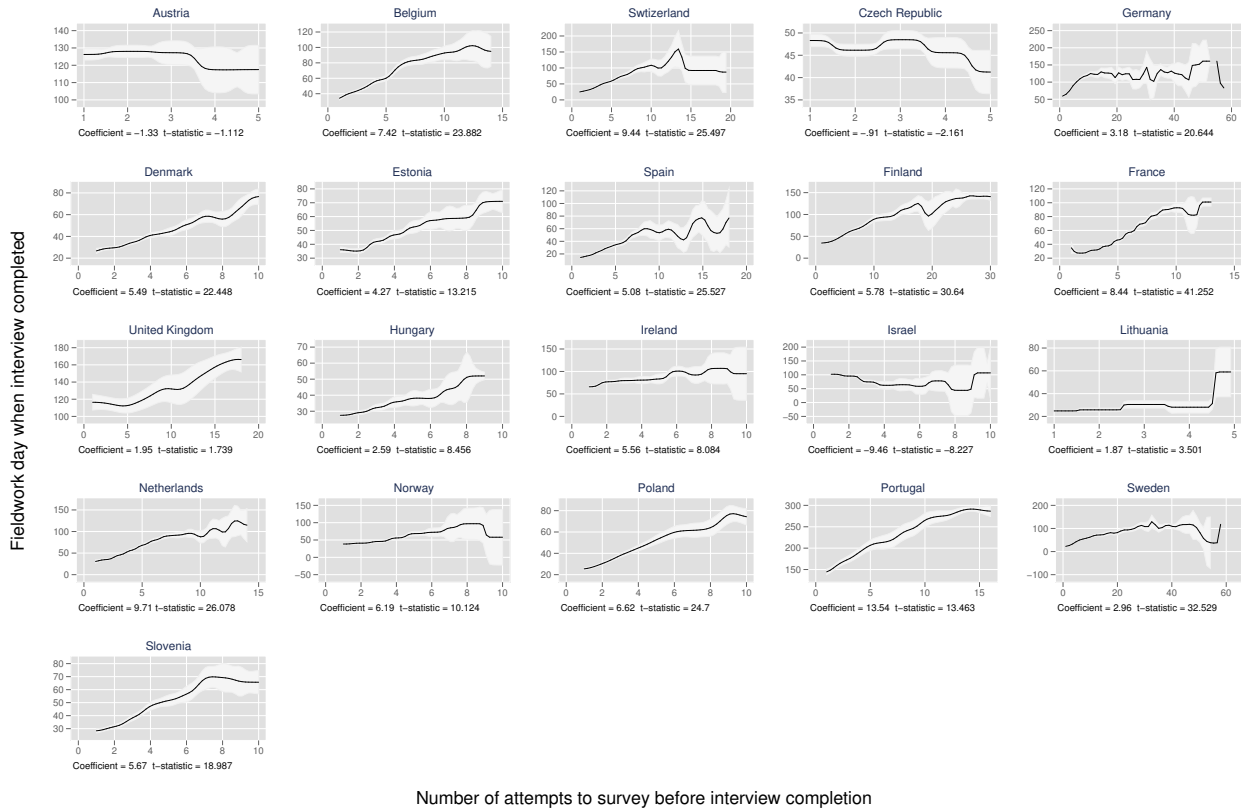
and statistically significant. In most cases, the higher the number of contact attempts needed to survey a subject, the higher the likelihood that she will be interviewed later during the fieldwork. Therefore, the subjects who are more elusive and harder to reach will have a higher probability of being interviewed *after* the event. If reachability is correlated with the potential outcomes, its correlation with treatment assignment will bias the estimates of the causal effect.

To exemplify the potential biases that the differential reachability of respondents might generate, in Table A1 in the Supporting Information we assess the relationship between the timing of interviews and respondents' observable characteristics in the ESS. As expected, those who were older, less educated, and out of the labor market (unemployed, retired, or doing housework) were interviewed earlier during the fieldwork of this survey. In contrast, when we conduct a similar analysis with the 2015 rolling cross-section of the CES and the 2008 NAES respondents' characteristics appear to be less related to the timing of their interview. However, even in these rolling cross-sectional surveys interview timing is related to some demographic characteristics of respondents (see Section A.2 in the Supporting Information for a discussion of these findings). Therefore, it is advisable to always account for potential biases related to reachability, even in the *a priori* most favorable sampling conditions.

3.2.3 Attrition

Survey research suffers from severe and increasing problems of non-response (Berinsky, 2017). This is particularly problematic when non-response is correlated with potential outcomes. In UESD studies, this might be even more troublesome if non-response is related to the treatment: *as a result of* being exposed to certain events some people might become more likely to refrain from responding to particular survey questions (item non-response), or even be less likely to participate in surveys altogether (unit non-response). The opposite can also happen when events unevenly increase the willingness of particular individuals to respond. In such cases, the general problem of non-response being correlated with the outcome is aggravated, as this correlation might be, in turn, different in the treatment and control groups. Take, as an illustration, the example of a coup that topples an autocratic government. There would be reasons to believe that individuals with certain political preferences –correlated with the outcome variable of interest– are now more (less) likely to respond to the survey than they were before. Differences in Y between both groups might reflect the substantively relevant effect of

Figure 1: Number of attempts to survey before interview completion and day of fieldwork when interview was completed



Note: Kernel-weighted local polynomial smoothing with 95% confidence intervals. Coefficient and t-statistic of OLS model regressing fieldwork day when interview was completed on the number of attempts to survey.

the event itself but, also, of the fact that different types of people are responding to the survey.

3.3 Noncompliance

So far we assumed that all respondents interviewed after the event were exposed to it (full compliance). Assignment to the treatment group $t_i > t_e$ was assumed to perfectly correspond to actually receiving the treatment $T = 1$. In most existing applications of the UESD this is a plausible assumption, since they refer to highly salient events.

In some cases, though, some subjects assigned to the treatment group might not be actually exposed to the event due to lack of direct experience or lack of access to information. This is potentially relevant in applications of the UESD in which the events are not extremely salient, for example Nobel Prize awards (De Vries, 2018). Under these circumstances, we would be facing a problem of one-sided noncompliance, in which some of the subjects are *never-takers*: despite having been assigned to the treatment group $t_i > t_e$, they did not receive it ($T = 0$). Since exposure to information or direct experience with the event might be correlated with

potential outcomes, in the presence of this type of non-compliance the difference between the subjects interviewed before and after the event ($Y_i(t_i > t_e) - Y_i(t_i < t_e)$) cannot be interpreted as the average treatment effect (ATE).

3.4 Heterogeneous effects and post-treatment bias

UESD studies commonly present heterogeneous effects as their central findings. This is often for good reasons, as these will be the theoretically relevant effects: the unconditional effect of an event might be of little theoretical interest. Moreover, the heterogeneous effects are often relevant to shed light on the causal mechanisms at work.

While general concerns about the causal interpretation of heterogeneous effects in experimental research might apply here (Stokes, 2014), in the case of the UESD special attention should be paid to potential post-treatment biases (Montgomery et al., 2018). In cross-sectional surveys the conditioning variables will always be measured after the event has occurred for the treatment group. While it is unlikely that some of the variables that could be used as moderators (e.g. gender or age) are affected by an event, many other theoretically relevant moderators could be shaped by it. If this is the case, the estimation of heterogeneous effects can be biased.

Think, as an illustration, of a study about the impact of a terrorist attack on trust in government, which tests whether the effect is different among those who identify with the incumbent party. In the unconditional comparison, the respondents interviewed before the attack might represent a valid counterfactual. However, once we condition for party identification we compare those that identified with the party in government before the attack and those that identify with it *in spite of* or *because of* the attack. If the attack had an effect on government support, the group of government identifiers (the moderator variable) would be different before and after the event, and therefore the estimate would not have a valid causal interpretation.

4 Estimation and assessment of assumptions

In the absence of threats to ignorability and excludability, a difference in means between the groups interviewed before and after the event would yield an unbiased estimation. However, this naive approach ignores the potential violations of these assumptions. Next, we propose a series of best practices in the form of various estimation strategies and robustness checks that can be used to lend credibility to the causal estimates.

4.1 Event description and qualitative assessment of assumptions

In-depth background knowledge is fundamental in order to validate the ignorability and excludability assumptions. A close and substantive examination of the event itself, of the circumstances surrounding it, and of the survey instrument should be the first step in any UESD study. Quantitative information will often prove useful for this, but researchers need to acquire in-depth case-based knowledge focused especially on the macro-level dynamics that generated the event and assigned units into treatment (see Kocher and Monteiro, 2016).

4.1.1 Event characterization

To understand the process of treatment assignment the event should be thoroughly characterized, paying special attention to its timing and salience. Assessing the salience of the event is of the utmost importance to assess the likelihood of noncompliance. The circumstances and causal chain leading to the event should also be discussed, paying special attention to the need of assessing and validating the definitional claim of the unexpected nature of the event. This examination should also address the possibility that, even if unexpected for most citizens, the event could have been strategically timed. Researchers should also describe the reactions that it generated in order to understand the role of collateral events.

Contextualizing the event and describing the circumstances surrounding its occurrence is also relevant in order to rule out the presence of other, simultaneous, events that might confound the effect of time. Sources such as Internet search trends and an analysis of print media can provide valuable evidence about the salience of the event and the potential presence of simultaneous events that could confound the estimation.

4.1.2 Background information on the survey

Detailed background knowledge about how the survey was conducted is also relevant to understand the process by which units are assigned to treatment. Special attention should be paid to the survey sampling procedure, the procedures to contact sampled units, and its strategy to deal with unit non-response. In the case of multi-stage sampling and/or quota sampling, researchers must collect information about how the fieldwork unfolded because that might create systematic differences in the type of respondents that are interviewed before and after t_e . In the case of disruptive events, one should also check if the event generated any changes in the fieldwork organization. Researchers must also assess if the survey procedures for

recontacting individuals or for refusal conversion might produce systematic biases, especially if refusal conversion strategies start at a specific point in time during the fieldwork. All these questions should be discussed in relation to the way in which the survey is administered –either in person, by phone, mail, or online–, as each of these methods will have particularities when it comes to the procedures to contact and re-contact sampled units.

4.2 Assessing and addressing violations of the ignorability assumption

4.2.1 Balance tests

To assess the plausibility of the ignorability assumption one can analyze balance on pre-treatment covariates between the groups. A difference of means between the two groups provides information on whether assignment to treatment and control is statistically independent of these covariates. If imbalances appear in characteristics related to the outcome of interest, this might generate a statistical dependence between assignment to treatment and the average outcome. Case-specific considerations should drive the decision of which covariates to include in these tests. Usually one would consider the sociodemographic characteristic of respondents, as well as their place of residence (e.g. region, municipality).

4.2.2 Bandwidths and statistical power

Selecting the appropriate bandwidth is of central concern in the literature on Regression Discontinuity Designs (RDD) (Imbens and Lemieux, 2008). Generally, in most RDD applications the bandwidth will be as narrow as possible, given that the forcing variable is substantively related to the potential outcomes. In the UESD, the equivalent to the RDD forcing variable is the timing of the interview, and in most applications it will not be related to the potential outcomes. Therefore, individuals interviewed around the day of the event will not necessarily be more similar to each other, and narrower bandwidths will increase variance (reduce N and statistical power) but not necessarily reduce bias. There are other downsides to the use of narrow bandwidths. First, they might also compromise the generalizability of the results, as the effects will tend to be very local. Second, the effects of certain types of events can take some time to unfold, and a narrow bandwidth might miss part of the effect, or even lead to a false negative. All these considerations should be taken into account when deciding on the appropriate bandwidth of days considered around the event date.

Under certain conditions, though, narrower bandwidths might reduce bias. In the presence

of time trends, collateral events, or confounders correlated with time of the interview, the UESD will present an analogous situation to the RDD. In these cases, and especially in surveys with prolonged fieldworks, researchers can narrow down the bandwidth h around the cutoff point and only use respondents interviewed during a certain period before and after the event to increase the plausibility of the ignorability assumption and the exclusion restriction. In these situations a narrower bandwidth could be an effective way of achieving conditional independence, since the potential outcomes of those interviewed around the day of the event should be more likely to be independent from treatment assignment: $Y_i(0), Y_i(1) \perp\!\!\!\perp T_i \mid h_i \rightarrow 0$, where h_i is unit's i absolute distance (in days) from the day when the event took place. Therefore, in some cases narrower bandwidths might provide an effective way to substantiate the *as-if random* treatment assignment assumption for that specific subset of the sample. Narrower bandwidths can also reduce the probability that other events or time trends drive the estimated effects. Narrower bandwidths should, therefore, be presented as a robustness check whenever possible. In surveys with prolonged fieldwork, where there is a greater potential for a correlation between the time of the interview and potential outcomes, this might even be the primary estimation strategy.

When conducting this robustness check (or when estimating the effects with a sub-sample around the event date) one should always consider the trade-off between narrower bandwidths and statistical power. In the UESD the sample is always split into pre- and post-event subgroups. As a consequence, the power of statistical tests is lower, and the study can be underpowered. This is especially the case if we are dealing with small samples, when the event occurs early/late during prolonged fieldworks, or when narrowing the bandwidth of days considered around the event date. Therefore, in all UESD studies, but especially in these cases, we recommend conducting power analyses to determine how many cases in the control and treatment groups would be adequate in order to detect a meaningful effect in each specific case. Hence, when discussing results based on multiple bandwidths we recommend a graphical representation that beyond the effect estimates, summarizes for each bandwidth the number of units included in the treatment and control groups and their corresponding statistical power (for meaningful effect sizes).

4.2.3 Covariate adjustment

The UESD might not always satisfy the ignorability assumption, even after narrowing the bandwidth around the date of the event. Therefore, to satisfy the *Conditional Ignorability Assumption* one can either pre-process the data through matching techniques to improve covariate balance between the treatment and control groups, and/or use controls in a regression framework. Matching techniques might be preferable in most situations, as they make less assumptions on the functional form of the relationships, and in some cases, such as the entropy balancing method (Hainmueller, 2012a) can balance the mean, variance and skewness of the covariates across groups.

This is a quite common practice in the literature. However, it is important to carefully select the covariates X_i that are included in the analyses. The covariates X_i used will generally be the same as those included in the balance tests. It is very important also to consider all those covariates that can take into account the survey design and the unfolding of the fieldwork, such as those related to the reachability and place of residence of respondents, as well as to their willingness to participate in the survey process. In the case of surveys using quota or stratified samples researchers must consider all the variables that are used to establish the quotas or stratum. However, it is also important to pay attention to the covariates that shall be avoided. In the UESD all covariates are measured post-treatment, so variables that can be potentially affected by the event should not be included to avoid post-treatment bias (Montgomery et al., 2018). For example, in studies analyzing the effects of a scandal on the attitudes toward the incumbent (e.g. Ares and Hernández, 2017), researchers might wish to include a covariate measuring if respondents identify with that party. However, precisely as a result of the scandal, respondents might be less likely to identify with the incumbent party, and introducing such a covariate could generate post-treatment bias.

4.2.4 Attrition

Events might alter respondents' likelihood of providing a valid answer to the questions measuring the outcome of interest (item non-response). Differences in the proportion of respondents who offer a *don't know* answer or refuse to answer the question measuring the outcome of interest across the control and treatment groups should be analyzed. If that is the case, researchers should carefully investigate whether there is any particular group of respondents that has mod-

ified their non-response pattern.

Events might also alter individuals' likelihood of responding to the survey altogether (unit non-response). Analyzing unit non-response patterns before and after the event poses greater challenges, as it requires access to survey paradata containing information about the moment when non-respondents were contacted. If this is available, one can first assess if the event affects the rates of failed interview attempts. However, a higher rate of unit non-response after the event might just reflect the fact that as the fieldwork progresses survey operatives are more likely to recontact those who are more difficult to reach or less willing to participate in the first place. Likewise, similar rates of non-response before and after the event might still pose a problem for causal identification if the profiles of non-respondents are different. A comparison of the characteristics of non-respondents before and after the event will provide valuable clues as to whether the event might have altered the likelihood of participating in the survey for a particular group.

4.3 Assessing and addressing violations of the exclusion restriction

The exclusion restriction implies that the timing of the survey only affects the outcome of interest through exposure to the event. The most relevant threats to this assumption are generated by simultaneous events and pre-existing time trends. One can assess the plausibility of this assumption through falsification tests, which test for the presence of an effect of the event and of the timing of the survey where it should not exist.

4.3.1 Inspection of pre-existing time trends

To address the possibility that pre-existing time trends, unrelated to the event of interest, could bias the findings researchers can test for the existence of such a trend before the event took place. Placebo treatments constructed at arbitrary points at the left of the cutoff point ($t_p < t_e$) should not affect the outcome. We follow the RDD literature and suggest to split the control group subsample at its empirical median and test for the absence of an effect at that point (Imbens and Lemieux, 2008). This strategy has the advantage of producing control ($t_i < t_p$) and treatment placebo ($t_p < t_i < t_e$) groups with a similar number of respondents. More generally, it is also advisable to examine the relationship between the timing of interviews and the outcome variable all throughout the pre-event period to check that these two variables are not systematically related during that period.

4.3.2 Falsification tests based on other surveys or units

Researchers can also conduct falsification tests that increase the plausibility of the excludability assumption by drawing on other surveys. Many surveys are repeated periodically. This provides the opportunity for conducting a falsification test of the effect of a placebo event that occurs on the same date as the event of interest but, for example, one year before/after it. These tests can help rule out that the effects of the event of interest are caused by aspects related to the timing of the survey fieldwork or to any other cyclical trends.

If the effects of the event of interest should be restricted to a specific region or country and contagion effects can be theoretically ruled out, one can also conduct placebo tests that exploit this circumstance. Surveys are generally conducted at the same time in different regions of a country or, also, in different countries. If this is the case, researchers can use these units as additional counterfactuals by testing for the effect of the event in a location where it should not exist (see Pollock et al., 2015). This test can be used to rule out that a global time-trend or any other simultaneous event that might be affecting the outcome of interest in different countries or regions bias the causal estimates. This of course does not apply to cases in which global or contagion effects are to be expected (see Legewie's (2013) study of the impact of a terror attack on citizens' attitudes toward immigrants in various distant countries).

4.3.3 Effects of the event on other outcome variables

One relevant threat for the excludability assumption is the occurrence of simultaneous events. While a thorough qualitative analysis will be useful to assess the plausibility of the assumption, researchers can also conduct falsification tests that involve testing for the effect of the timing of the event on variables that, theoretically, should not be affected by it.

The crucial question here is how to select these variables. Recall that t (timing of the interview) is an instrument for the event T , and therefore it should only affect Y through T . If one suspects that another event T' is driving the effect, a good falsification test would imply looking for an alternative outcome Y' that is theoretically unrelated to T but susceptible of being affected by other events T' that might also have an effect on Y . Therefore, ideally one would like to have variables that are as close as possible to Y except for the fact that we should expect them not to be affected by T . If we find an effect of t on Y' , this might be an indication of a violation of the exclusion restriction. While this test might lend additional plausibility to

assumption, it is not, by itself, either necessary nor sufficient to rule out the possibility that the estimated results for Y are being driven by another event.

4.4 Assessing compliance

The one-sided non-compliance problem discussed above refers to the situation in which the subjects in the treatment group have not been exposed to the treatment. In the UESD framework, all respondents interviewed after the event will be exposed, provided that they were aware of it. Unfortunately, this is not directly testable, as the unexpected nature of the event of interest means that survey questionnaires will not include specific manipulation checks.

However, sometimes one can rely on indirect pseudo-manipulation checks. Salient events often have an impact on the types of problems that respondents perceive to be important in their countries, and this might be reflected in their responses to the “most important problem(s)” questions, which are routinely asked in public opinion surveys (see e.g. Boomgaarden and de Vreese, 2007). If this is the case, a perceptible increase in the salience attributed to the problems to which the event is related should be observed for all types of respondents, regardless of their pre-treatment characteristics.

If non-compliance is a real possibility, the difference between the treatment and control groups should be interpreted as an *Intent-To-Treat* (ITT) effect, rather than as an ATE or an Average Treatment effect on the Treated (ATT). The subjects in the group interviewed after the event “have been given the opportunity” to receive the treatment, but often we cannot be entirely certain that they have been effectively exposed to it, and thus the interpretation as an ITT seems more appropriate. This is especially important in less salient events or in poor information environments, where the likelihood of noncompliance is higher.

4.5 Summary

In Table 1 we summarize the strategies that applied researchers can implement to assess the plausibility of these assumptions and, eventually, address some of their most common violations. Our analysis of these potential violations advises caution in the causal interpretation of the effects. In any UESD special attention should be devoted to the assessment of the observable implications of the ignorability and excludability assumptions and, in all instances, studies should conduct robustness checks that can help rule out some of the key violations of these assumptions. The list of strategies in Table 1 should be read as series of best practices to be

Table 1: Tests and robustness checks to assess and address violations of assumptions

Main Assumption	Test/Check	Aim	Expected result	Section
Ignorability and Excludability	In-depth description of the event and survey	Assess violation	Convincing case for unexpectedness, saliency, and uniqueness of the event, and randomness of timing of the interview after close examination	4.1
Ignorability	Balance tests	Assess violation	No significant treatment-control group differences in “pre-treatment” covariates	4.2.1
Ignorability	Multiple bandwidths	Address violation / Robustness	Significant treatment-control group differences in the outcome variable after narrowing the bandwidths	4.2.2
Ignorability	Covariate adjustment	Address violation	Significant treatment-control group differences in the outcome variable robust to accounting for relevant covariate imbalances	4.2.3
Ignorability	Analysis of non-response	Assess violation	No significant treatment-control group differences in: i) frequency of item non-response (outcome variable); ii) frequency of unit non-response; iii) characteristics of sampled units who do not participate in the survey	4.2.4
Excludability	Placebo treatments at different dates of the control group	Assess violation	No significant placebo effects in the control group	4.3.1
Excludability	Inspection of pre-existing time trends	Assess violation	Outcome variable temporally stable in the control group	4.3.1
Excludability	Falsification tests through other surveys/units	Assess violation	No significant placebo effects in other periods (same dates, other years) and in geographical areas unaffected by the event	4.3.2
Excludability	Falsification tests on other outcome variables	Assess violation	No significant treatment-control group differences in other outcome variables theoretically unrelated to the event	4.3.3
Compliance	Pseudo-manipulation checks	Assess violation	Significant treatment-control group differences in importance of problems related to the event, independently of pre-treatment characteristics	4.4
Compliance	Restrict interpretation of the causal effect	Address violation	Interpretation as ITT effect	4.4

Table 2: Empirical strategy and robustness checks of 44 published studies

Check	N	Proportion
Covariate adjustment	40	0.91
Balance tests	33	0.75
In-depth description of the event	25	0.57
Multiple bandwidths	17	0.39
Falsification tests on other outcome variables	13	0.30
In-depth description of the survey	12	0.27
Inspection of pre-existing time trends	11	0.25
Falsification tests through other surveys/units	9	0.20
Analysis of non-response	8	0.18
Placebo treatments at different dates of the control group	6	0.14
Pseudo-manipulation checks	6	0.14
Restricted interpretation of treatment effect	1	0.02

followed whenever possible rather than as a checklist for each and every UESD study. Often they will be not relevant or not possible to conduct. Take for instance Kim and Kim’s (2019) analysis of the impact of a public statement on the diffusion of a rumour, where a “pseudo-manipulation check” would overlap with the estimation of the main effect on the outcome variable.

Table 2 presents an overview of the empirical strategies and robustness checks of published UESD studies.⁸ The table reveals a high degree of heterogeneity in the empirical strategies implemented in existing UESD studies. While most of them conduct balance tests and some form of covariate adjustment, the use of other robustness checks and strategies to assess and substantiate the observable implications of the assumptions of this research design are far less common.

5 Empirical application: The Charlie Hebdo terrorist attacks and satisfaction with government

We illustrate the use of the UESD through an analysis of the effects of the Charlie Hebdo terrorist attacks on individuals’ satisfaction with government in France. While terrorism can affect political attitudes in countries other than those directly targeted (Legewie, 2013), we restrict the illustration to France, where according to the well-established “rally round the flag” hypothesis (see Hetherington and Nelson, 2003) we can expect a positive impact of these attacks on government approval.

⁸See Table B1 in the Supporting Information for an assessment of each study.

The Charlie Hebdo terrorist attacks occurred in Paris at approximately 11:30AM of January 7th, 2015. As a result of the attacks twelve people related to the Charlie Hebdo satirical magazine were killed. On January 8th, four hostages were also killed in a related incident involving an acquittance of the Charlie Hebdo attackers. The two Al-Qaeda terrorists that perpetrated the Charlie Hebdo attack were shot and killed by the police on January 9th. The attacks were followed by large-scale demonstrations, which gathered citizens and politicians from all parties as well as international leaders. These attacks coincided with the fieldwork of the 7th round of the ESS in France, which was in the field between October 31st, 2014 and March 3rd, 2015.⁹

We exemplify the use of the UESD through this event for three reasons. First, terrorist attacks are the type of events most often analyzed through the UESD. Second, the “rally round the flag” theory lead us to expect that satisfaction with government should increase in the aftermath of the attacks. Third, the ESS provides extensive documentation and complementary data that allow us to illustrate a series of good UESD practices.

5.1 Naive baseline estimation

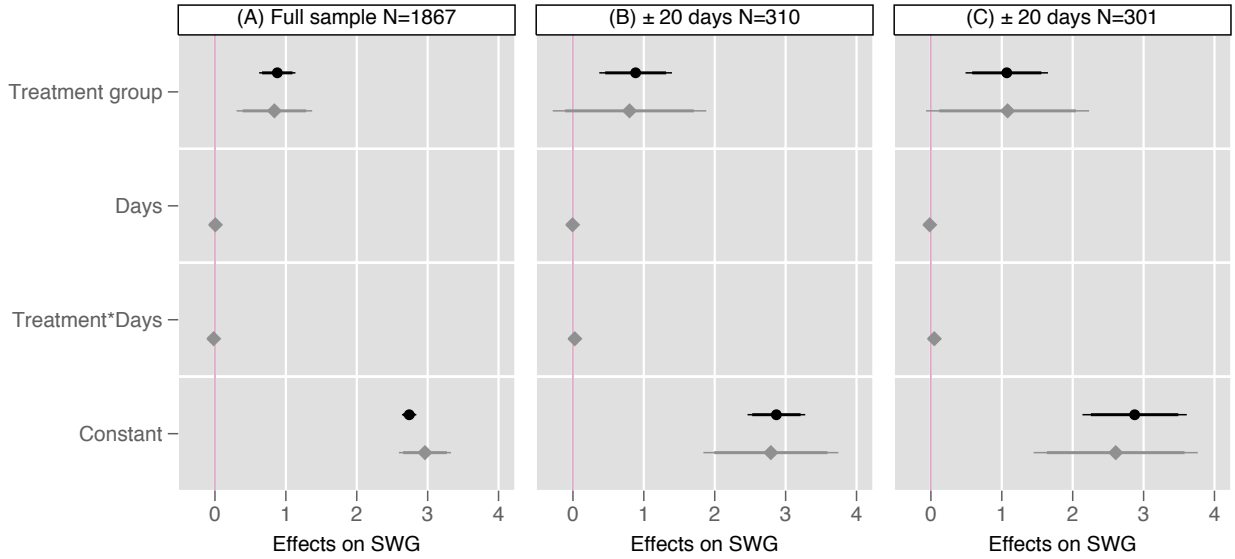
Panel A in Figure 2 summarizes the naive baseline estimation of the effects of the Charlie Hebdo attacks on individuals’ satisfaction with government (SWG), which is measured on a 0-10 scale. Since the event took place on January 7th at 11:30, we exclude all those interviewed during that day from the analyses. Therefore, the treatment group variable takes the value 0 for those interviewed before January 7th and the value 1 for those interviewed after that day.

The results of an OLS model that only includes the treatment indicator (black spikes and circle markers in Panel A) indicate that those interviewed after the attacks were more satisfied with the national government. The effect of the attacks is substantial, since a 0.88 average increase in satisfaction with government corresponds to a change of almost half a standard deviation on the pre-treatment distribution of the satisfaction with government variable.

Next we estimate if this change in satisfaction with government occurred immediately after the Charlie Hebdo attacks. For this purpose we specify a model that includes a running variable ranging from -68 to 46 (the fieldwork days before and after the attacks), with 0 corresponding to January 8th. This variable is interacted with the treatment indicator. In this interactive model

⁹An in-depth description of the event and detailed background information on the ESS can be found in Supporting Information C.

Figure 2: Effects of Charlie Hebdo attacks on satisfaction with government (SWG).



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. The models summarized in Panel C are estimated with entropy balancing weights and region fixed-effects.

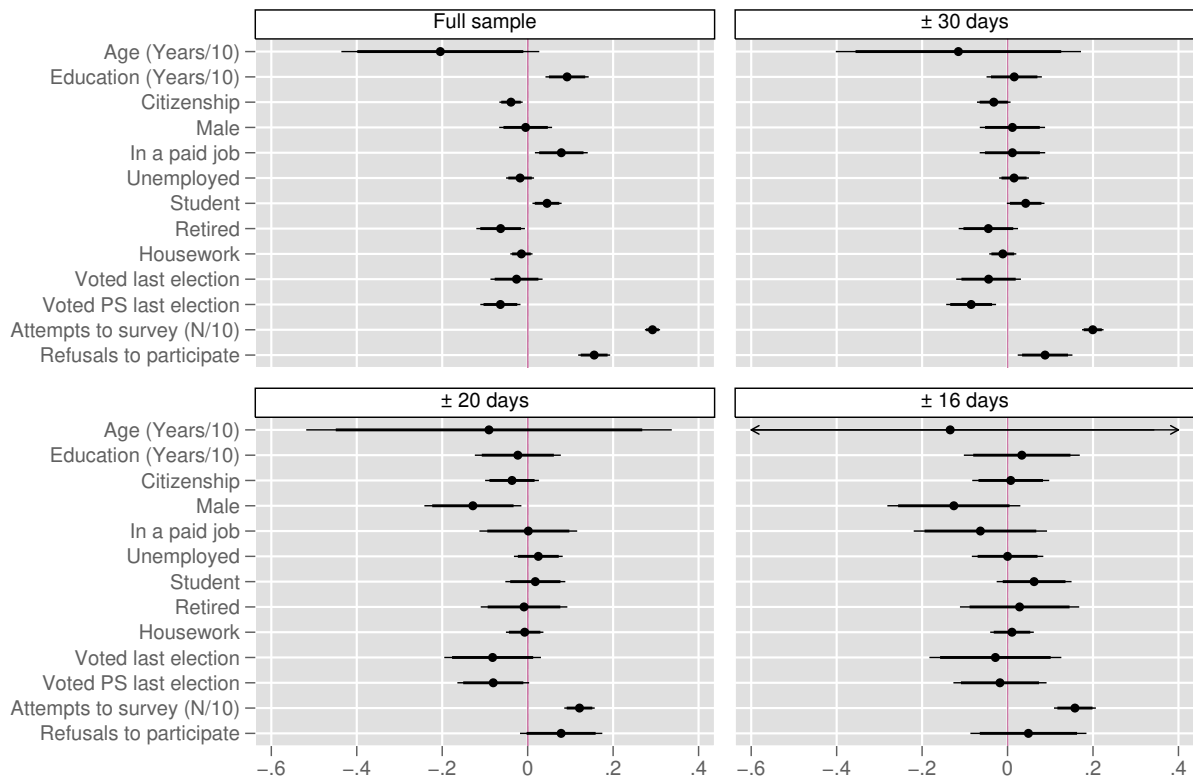
the constitutive term for the “treatment group” variable corresponds to the effect of the terrorist attacks when the running variable “days” equals 0 (i.e. on January 8th, the first day immediately after the attacks). The interactive term, in turn, indicates whether or not the effects of the attacks changed (weakened or strengthened) as time went by after the attacks. The “treatment group” coefficient in the interactive model (grey spikes and diamond markers in Panel A of Figure 2), reveals that the levels of satisfaction with government increased immediately after the attacks. Moreover, the interactive term of this model indicates that the effects of the attacks did not significantly change thereafter (see Finseraas and Listhaug, 2013, for a similar finding about the effects of terrorist attacks on fear of terror).

While all these results are in line with the rally “round the flag” hypothesis, we cannot interpret them causally unless we assess whether the key identifying assumptions of the UESD are met.

5.2 Balance tests, bandwidths, and covariate adjustment

In Figure 3 we summarize the covariate imbalances between the treatment and control groups using multiple bandwidths. In the balance tests we include the sociodemographic characteristics of respondents and two retrospective variables measuring if the respondent voted in

Figure 3: Balance tests: Covariate differences between the treatment and control groups.



Note: Entries report the difference in the mean of the covariates between the treatment and control groups. Thick and thin lines are 95% and 90% confidence intervals, respectively

the last election and if she voted for the party in government (PS). In addition, we also include two variables from ESS paradata related to the reachability of respondents: the number of times that an individual was contacted before the completion of the survey interview and the number of times she refused to participate.

In the full sample, the treatment and control groups are unbalanced in several pre-treatment covariates. Those in the treatment group were harder to reach, younger, more educated, more likely to be either employed or studying, and less likely to support the party in government. By narrowing the bandwidth these imbalances are substantially reduced.¹⁰

To conduct our analyses of the Charlie Hebdo attacks we rely on the ± 20 days bandwidth for several reasons. First, the balance tests indicate that relying on a wider bandwidth leads to larger imbalances in pre-treatment covariates. Second, our assessment of the aftermath of the attacks indicates that until January 25th the French media coverage was clearly dominated

¹⁰For example, for the ± 20 bandwidth we only find statistically significant differences for gender, vote for the incumbent party (at $p < 0.1$), and the number of attempts to survey. However, even when relying on narrower bandwidths there are still some regional imbalances between the two groups. Balance tests including respondents' region of residence can be found in Figure D1 of the Supporting Information.

by news related to the attacks. Third, the ± 20 bandwidth provides sufficient statistical power to detect a meaningful increase in satisfaction with government (equivalent to one third of the standard deviation of this variable in the control group).^{11,12}

The effects of the terrorist attacks drawing on the ± 20 bandwidth are summarized in Panel B of Figure 2. Results indicate that the average level of satisfaction with government was significantly higher after the attack. Moreover, the interactive model (grey spikes) suggests that the boost in satisfaction occurred immediately after the attack, although, in this case, the “treatment group” coefficient does not reach conventional levels of statistical significance.

Regarding covariate adjustment, we rely on entropy balancing to pre-process the data and produce covariate balance between the treatment and control groups by re-weighting units appropriately (Hainmueller, 2012*b*).

To re-weight the observations we consider all the covariates X_i included in the balance tests, with the exception of the variable measuring the number of attempts to survey and the respondent’s region of residence.¹³ When re-weighting the sample through entropy balancing and including region fixed-effects the average effect of the Charlie Hebdo attacks is of 1.07 (Panel C in Figure 2). Moreover, the “treatment group” coefficient of the interactive model indicates that the average increase in satisfaction with government was almost of the same magnitude immediately after the attacks.

5.3 Additional robustness checks

In the Supporting Information we conduct a set of analyses to assess the plausibility of the temporal stability assumption through a placebo treatment (Table D3), analyze the relationship between the timing of interviews and the outcome variable during the whole survey fieldwork period (Figure D9), rule out that the attacks caused a problem of attrition (Table D4

¹¹A narrower bandwidth might not provide enough power to detect an effect of this size. See the power analysis in Figure D2 of the Supporting Information

¹²In Figure D3 of the Supporting Information we replicate the estimation of the effects using multiple bandwidths.

¹³We exclude the former variable from the entropy balancing because the distribution of attempts to survey is radically different between the control and treatment groups. Including this variable imposes an unrealistic constraint that produces re-balancing weights in the control group that lead to results that end up being driven by a few respondents from the control group (see section D.3 in the Supporting Information for further details). Concerning the respondents’ region of residence, the entropy balancing algorithm fails to converge when this variable is included (tolerance level of 0.015). To prevent biases related to regional imbalances, the models summarized in Panel C of Figure 2 estimate the effect of the attacks through specifications that besides the entropy balancing weights also include region fixed-effects. Finally, one could argue that vote recall questions might be affected by the terrorist attacks and could, therefore, induce post-treatment bias. In Figure D8 of the Supporting Information we replicate the analyses without balancing the two groups with regard to vote recall.

and Table D5), perform falsification tests based on other rounds of the ESS to rule out that the identified effects are driven by seasonal trends (Figure D10), assess the threat posed by simultaneous events through falsification tests using alternative outcomes (Figure D11).¹⁴

6 Concluding Remarks

In this paper we have described and systematized the UESD, focusing on the assumptions under which unexpected events that occur during the fieldwork of surveys can yield valid causal estimates. The impact of political events on public opinion can be analyzed through different designs. For example, one can correlate the recall of an event (or the importance attributed to it) and the outcome of interest through a cross-sectional survey. However, this empirical strategy is likely to be affected by endogeneity biases. Longitudinal studies with repeated observations of the same individuals can overcome some of these biases. Yet in these studies the outcomes of interest are often measured a long time after the event occurred. This might lead to an underestimation of the effects and a greater potential for bias due to the occurrence of other unrelated events. Priming survey experiments are an alternative in which researchers have full control over treatment assignment and administration. Therefore, they are more immune to the type of threats discussed here. However, they are generally based on artificially designed treatments, and subject to potentially short-lived effects (Gaines et al., 2006), which may raise concerns about their external validity (Barabas and Jerit, 2010). The UESD, this paper has argued, can be a compelling alternative, provided its key assumptions are shown to hold.

One of the drawbacks of most UESD studies is that they are event specific, and this can limit the generalizability of the findings. Each event is in some regards unique. With a single-event study we cannot rule out the possibility that the estimated effect is driven by one of its specific characteristics, or by collateral events. This can be addressed if the timing of collateral events allows for a separate estimation based on the change in the outcome of interest at the exact time in which the event took place. However, this is not always possible or advisable. Therefore, it is always necessary to discuss how the event analyzed might deviate from a typical

¹⁴There are further tests and robustness checks that we recommended but we cannot conduct in our application. First, salient terrorist attacks, especially those perpetrated by international terror groups, produce attitudinal effects beyond the borders of the country directly affected by them (Legewie, 2013). Hence, using other surveys administered simultaneously in other countries to conduct placebo tests does not seem advisable. Moreover, we cannot assess non-compliance through the ‘most important problem(s)’ questions as no such questions is included in the ESS.

case of that class of events.

The ideal way to increase the generalizability of UESD studies is to analyze more than one event of the same class in order to establish some regularities (e.g. Legewie, 2013). However, this alternative will often not be feasible, simply because no other event of the same class coincided with the fieldwork of a survey. A good alternative in these cases is to complement UESD studies with lab or survey experiments, which can be designed to study the effect of a similar event in isolation (e.g. Flores, 2018).

One question that remains is whether surveys with a short or prolonged fieldwork are more adequate to conduct UESD studies. The main benefit of a prolonged fieldwork is that one can assess in detail the existence of pre-existing time trends. However, when compared to shorter fieldworks, they have clear drawbacks as the differential reachability of respondents is likely to generate greater differences between the control and treatment groups. Like in our Charlie Hebdo example, this often imposes the need to focus on a subset of the sample interviewed around the day of the event, which might limit the generalizability of the findings. Moreover, in surveys with prolonged fieldworks we will generally find a smaller number of observations around the event date, and the likelihood of the occurrence of simultaneous events that might confound the effects is higher. On the contrary, in short fieldworks, the *as-if random* assignment to treatment is more likely to be plausible without the need for further adjustments, and the threat of simultaneous events is also likely to be lower. Therefore, the benefits of surveys with short fieldwork periods clearly outweigh those of surveys with prolonged fieldworks.

Awareness of the potential threats to causal identification affecting the UESD should not lead researchers to dismiss it as a potentially fruitful identification strategy. While it certainly falls short of a randomized controlled trial, the fact that it is based on naturally occurring treatments and often relies on representative population samples makes UESD studies potentially strong in terms of their external validity. The tests and robustness checks we propose should lead to a more careful implementation of this design, which shall, in turn, increase the credibility of causal claims based on UESD studies.

References

- Ares, Macarena and Enrique Hernández. 2017. “The corrosive effect of corruption on trust in politicians: Evidence from a natural experiment.” *Research & Politics* 4(2):1–8.
- Balcells, Laia and Gerard Torrats-Espinosa. 2018. “Using a natural experiment to estimate the electoral consequences of terrorist attacks.” *Proceedings of the National Academy of Sciences* 115(42):10624–10629.
- Barabas, Jason and Jennifer Jerit. 2010. “Are survey experiments externally valid?” *American Political Science Review* 104(2):226–242.
- Berinsky, Adam J. 2017. “Measuring Public Opinion with Surveys.” *Annual Review of Political Science* 20:309–329.
- Bischof, Daniel. 2017. “New graphic schemes for Stata: plotplain and plottig.” *The Stata Journal* 17(3):748–759.
- Boomgaarden, Hajo G. and Claes H. de Vreese. 2007. “Dramatic Real-world Events and Public Opinion Dynamics: Media Coverage and its Impact on Public Reactions to an Assassination.” *International Journal of Public Opinion Research* 19(3):354–366.
- Boydston, Amber E., Jessica T. Fezell and Rebecca A. Glazier. 2018. “In the wake of a terrorist attack, do Americans’ attitudes toward Muslims decline?” *Research & Politics* 5(4):1–7.
- Branton, Regina, Valerie Martinez-Ebers, Tony E. Carey and Tetsuya Matsubayashi. 2015. “Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies.” *American Journal of Political Science* 59(2):390–402.
- Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. University of Michigan Press.
- Coupe, Tom. 2017. “The impact of terrorism on expectations, trust and happiness – the case of the November 13 attacks in Paris, France.” *Applied Economics Letters* 24(15):1084–1087.
- Csikszentmihalyi, Mihaly and Jeremy Hunter. 2003. “Happiness in everyday life: The uses of experience sampling.” *Journal of happiness studies* 4(2):185–199.
- De Vries, Catherine E. 2018. *Euroscepticism and the Future of European Integration*. Oxford University Press.
- Dinesen, Peter Thisted and Mads Meier Jæger. 2013. “The effect of terror on institutional trust: New evidence from the 3/11 Madrid terrorist attack.” *Political Psychology* 34(6):917–926.
- Eggers, Andrew C, Ronny Freier, Veronica Grembi and Tommaso Nannicini. 2018. “Regression discontinuity designs based on population thresholds: Pitfalls and solutions.” *American Journal of Political Science* 62(1):210–229.
- ESS Round 7: European Social Survey Round 7 Data. Data file edition 2.1. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.* N.d.
- Finseraas, Henning and Ola Listhaug. 2013. “It can happen here: the impact of the Mumbai terror attacks on public opinion in Western Europe.” *Public Choice* 156(1-2):213–228.
- Flores, René D. 2018. “Can Elites Shape Public Attitudes Toward Immigrants?: Evidence from the 2016 US Presidential Election.” *Social Forces* 96(4):1649–1690.

- Gaines, Brian J, James H Kuklinski and Paul J Quirk. 2006. "The logic of the survey experiment reexamined." *Political Analysis* 15(1):1–20.
- Geys, Benny and Salmai Qari. 2017. "Will you still trust me tomorrow? The causal effect of terrorism on social trust." *Public Choice* 96(3-4):289–305.
- Hainmueller, Jens. 2012a. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis* 20(1):25–46.
- Hainmueller, Jens. 2012b. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1):25–46.
- Hernández, Enrique, Albert Falcó-Gimeno and Jordi Muñoz. 2019. "Replication Data for: Unexpected Event during Surveys Design: Promise and Pitfalls for Causal Inference." doi.org/10.7910/DVN/RDIIVL.
- Hetherington, Marc J and Michael Nelson. 2003. "Anatomy of a rally effect: George W. Bush and the war on terrorism." *PS: Political Science & Politics* 36(1):37–42.
- Hofstetter, C. Richard. 1969. "Political Disengagement and the Death of Martin Luther King." *The Public Opinion Quarterly* 33(2):174–179.
- Imbens, Guido W. and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics* 142(2):615–635.
- Jakobsson, Niklas and Svein Blom. 2014. "Did the 2011 Terror Attacks in Norway Change Citizens' Attitudes Toward Immigrants?" *International Journal of Public Opinion Research* 26(4):475–486.
- Jensen, Carsten and Elias Naumann. 2016. "Increasing pressures and support for public health-care in Europe." *Health Policy* 120(6):698–705.
- Johnston, Richard and Henry E Brady. 2002. "The rolling cross-section design." *Electoral Studies* 21(2):283–295.
- Keele, Luke and Rocío Titiunik. 2016. "Natural Experiments Based on Geography." *Political Science Research and Methods* 4(1):65–95.
- Kim, Jin Woo and Eunji Kim. 2019. "Identifying the effect of political rumor diffusion using variations in survey timing." *Quarterly Journal of Political Science* forthcoming.
- Kocher, Matthew A. and Nuno P. Monteiro. 2016. "Lines of Demarcation: Causation, Design-Based Inference, and Historical Research." *Perspectives on Politics* 14(4):952–975.
- Larsen, Erik Gahner. 2018. "Welfare Retrenchments and Government Support: Evidence from a Natural Experiment." *European Sociological Review* 34(1):40–51.
- Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118(5):1199–1245.
- Lever, H. 1969. "The Johannesburg Station Explosion and Ethnic Attitudes." *The Public Opinion Quarterly* 33(2):180–189.

- MacKinlay, A Craig. 1997. "Event studies in economics and finance." *Journal of economic literature* 35(1):13–39.
- Metcalf, Robert, Nattavudh Powdthavee and Paul Dolan. 2011. "Destruction and distress: Using a quasi-experiment to show the effects of the September 11 attacks on mental well-being in the United Kingdom." *The Economic Journal* 121(550):F81–F103.
- Minkus, Lara, Emanuel Deutschmann and Jan Delhey. 2018. "A Trump effect on the EU's popularity? The US presidential election as a natural experiment." *Perspectives on Politics* forthcoming:1–18.
- Montgomery, Jacob M., Brendan Nyhan and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3):760–775.
- Motta, Matthew. 2018. "The Polarizing Effect of the March for Science on Attitudes toward Scientists." *PS: Political Science and Politics* 51(4):782–788.
- Perrin, Andrew J. and Sondra J. Smolek. 2009. "Who trusts? Race, gender, and the September 11 rally effect among young adults." *Social Science Research* 38(1):134–145.
- Pierce, Lamar, Todd Rogers and Jason A. Snyder. 2016. "Losing Hurts: The Happiness Impact of Partisan Electoral Loss." *Journal of Experimental Political Science* 3(1):44–59.
- Pollock, William, Jason Barabas, Jennifer Jerit, Martijn Schoonvelde, Susan Banducci and Daniel Stevens. 2015. "Studying media events in the European social surveys across research designs, countries, time, issues, and outcomes." *European Political Science* 14(4):394–421.
- Silber Mohamed, Heather. 2013. "Can protests make Latinos "American"? Identity, immigration politics, and the 2006 marches." *American Politics Research* 41(2):298–327.
- Silva, Bruno Castanho. 2018. "The Non-Impact of the 2015 Paris Terrorist Attacks on Political Attitudes." *Personality and Social Psychology Bulletin* 44(6):838–850.
- Solaz, Hector, Catherine E. De Vries and Roosmarijn A. de Geus. 2018. "In-Group Loyalty and the Punishment of Corruption." *Comparative Political Studies* OnlineFirst:1–31.
- St. Clair, Travis, Thomas D Cook and Kelly Hallberg. 2014. "Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment." *American Journal of Evaluation* 35(3):311–327.
- Stokes, Susan C. 2014. A defense of observational research. In *Field experiments and their critics: Essays on the uses and abuses of experimentation in the social sciences*, ed. Dawn Teele. New Haven: Yale University Press pp. 33–57.
- Stoop, Ineke A. L. 2004. "Surveying Nonrespondents." *Field Methods* 16(1):23–54.
- Tiokhin, Leonid and Daniel Hruschka. 2017. "No evidence that an Ebola outbreak influenced voting preferences in the 2014 elections after controlling for time-series autocorrelation: A commentary on Beall, Hofer, and Schaller (2016)." *Psychological science* 28(9):1358–1360.
- Zepeda-Millán, Chris and Sophia J Wallace. 2013. "Racialization in times of contention: How social movements influence Latino racial identity." *Politics, Groups, and Identities* 1(4):510–527.

Supporting Information for:
Unexpected Event during Surveys Design:
Promise and Pitfalls for Causal Inference

Jordi Muñoz

Albert Falcó-Gimeno

Enrique Hernández

University of Barcelona

University of Barcelona

Universitat Autònoma de Barcelona

Accepted for publication at *Political Analysis*

*This research has received financial support from the Spanish Ministry of Science, Innovation, and Universities through the research grant CSO2017-89085-P. We are grateful to Lucas Leeman, Silja Häusermann, Macarena Ares, Guillem Rico, Salomo Hirvonen, the editor and anonymous reviewers of *Political Analysis*, and participants at the University of Zurich IPZ publication seminar, the Autonomous University of Barcelona DEC Seminar, and the 2018 EPSA conference for helpful comments and suggestions. We are especially grateful to Erik Gahner Larsen for his detailed and insightful comments, and for generously sharing with us his collection of UESD references. Enrique Hernández also thanks the University of Zurich Political Science Department for hospitality during the spring of 2018. Replication materials are available at the *Political Analysis Dataverse*: doi.org/10.7910/DVN/RDIIVL (Hernández et al., 2019).

Contents

A	Reachability	A3
A.1	Analysis of reachability in 21 European countries: ESS	A3
A.2	Reachability in the rolling cross-section design	A4
B	Assessment of the empirical strategies of UESD studies	A6
C	Charlie Hebdo attacks: Event and survey description	A9
C.1	Characterization of the Charlie Hebdo attacks	A9
C.2	Background information on ESS-7 fieldwork in France	A11
D	Additional analyses of Charlie Hebdo attacks	A12
D.1	Regional imbalances	A12
D.2	Bandwidths and statistical power	A13
D.3	Entropy balancing details	A17
D.4	Entropy balancing excluding vote recall	A20
D.5	Pre-existing time trends	A21
D.6	Attrition and non-response	A23
D.7	Falsification tests based on other surveys	A25
D.8	Falsification tests based on other outcome variables	A26

A Reachability

A.1 Analysis of reachability in 21 European countries: ESS

To assess the potential problems related to the differential reachability of respondents that might affect studies exploiting the UESD we first assess how the observable characteristics of respondents are related to the time during the fieldwork when they were interviewed in the 21 European countries included in ESS-7.

Table A1: Correlates of fieldwork day when survey interview was completed: 2016 ESS

	(1)	(2)
Age (years)	-0.10 (0.02)	-0.07 (0.02)
Education (years)	0.31 (0.07)	0.23 (0.07)
Citizenship	-0.20 (1.13)	-0.19 (1.10)
Male	0.67 (0.50)	0.42 (0.49)
In a paid job	0.62 (0.82)	-0.61 (0.80)
Unemployed	-4.30 (1.22)	-3.70 (1.19)
Student	-0.73 (1.08)	-1.31 (1.05)
Retired	-2.86 (1.04)	-2.20 (1.01)
Housework	-1.94 (0.75)	-1.97 (0.73)
Attempts to survey		3.81 (0.10)
Refusals to participate		16.99 (0.92)
Constant	128.43 (2.12)	121.70 (2.07)
Country fixed-effects	Yes	Yes
Observations	39697	39697
Countries	21	21
R^2	0.347	0.383

Standard errors in parentheses

Model 1 in Table A1 summarizes the results of an OLS regression model fitted to analyze

the relationship between the sociodemographic characteristics of respondents and the timing of their survey interview. The results indicate that those who are older, less educated, and out of the labor market (unemployed, retired or doing housework) are, on average, more likely to be interviewed earlier during the fieldwork period. These sociodemographic groups should be easier to reach for an interview because presumably they should spend longer periods of time at home and might have more time available to answer surveys. Therefore, they are more likely to be interviewed earlier during the fieldwork.

The second model incorporates variables that are directly related to the differential reachability of respondents: the number of times respondents refused to participate in the survey, and the number of attempts to survey before an interview could be completed. The results indicate that refusing to participate in the survey and a higher number of attempts to survey increase the likelihood of being interviewed later during the fieldwork. Moreover, even when accounting for these variables related to the reachability of respondents the sociodemographic characteristics of those interviewed later during the fieldwork are still significantly different.

A.2 Reachability in the rolling cross-section design

In this section we replicate the analysis of the covariates of interview day using two well-known surveys that explicitly take on the rolling cross-section design: the 2015 Canadian Election Study (CES) and the National Annenberg Election (NAES) Survey 2008 Phone Edition. The rolling cross-section design is explicitly aimed at making the interview date of each respondent random, and it is typically used in the context of election campaigns with the intention of studying campaign effects. The sample is randomly split into groups that are released to the fieldwork operators on a daily basis.

This strategy should prevent imbalances in both observables and unobservables from occurring. However, in practice we might still observe some imbalances: within each block of respondents, those that are more likely to refuse and harder to reach will be concentrated towards the end of the block since they have a few days to be contacted. This means that for any arbitrary partition of the sample one might still find differences. In table A2 we summarize the results of the same analysis conducted with the ESS for the two rolling cross sections. The CES offers information on attempts to survey and refusals, while the NAES does not. In both cases we see some minor imbalances in observable characteristics of the respondents, but it

is especially relevant to notice that in the CES there is a strong correlation between date of interview and refusals and attempts to survey. Therefore, we conclude that in any case it is advisable to check and, eventually, address potential problems of ignorability of the treatment assignment in the UESD context, even in the a priori most favorable sampling conditions.

Table A2: Correlates of interview day. 2015 CES & 2008 NAES

	CES		NAES
	(1)	(2)	(3)
Male	0.13 (0.37)	0.07 (0.36)	-2.39 (0.89)
Age	0.00 (0.02)	0.00 (0.01)	-0.01 (0.00)
Education	0.03 (0.09)	0.07 (0.09)	1.07 (0.19)
Born in country	0.40 (0.53)	0.65 (0.52)	1.48 (1.61)
Paid work	0.95 (0.58)	0.89 (0.57)	0.44 (1.41)
Retired	0.42 (0.72)	0.30 (0.70)	3.30 (1.54)
Unemployed	-1.04 (1.10)	-1.31 (1.09)	11.91 (3.03)
Student	4.59 (1.08)	4.59 (1.06)	-3.97 (4.20)
Housework	0.58 (1.27)	0.51 (1.25)	-1.30 (2.41)
Disabled/Ill	-1.24 (1.49)	-1.28 (1.47)	-3.80 (2.63)
# Refusals to participate		6.81 (0.60)	
# Contacts to participate		1.47 (0.47)	
Constant	20.52 (1.17)	17.98 (1.25)	152.27 (2.32)
R-squared	0.007	0.041	0.001
N	4011	4011	56181

Standard errors in parentheses

B Assessment of the empirical strategies of UESD studies

Table B1 summarizes a detailed assessment of the empirical strategies and robustness tests performed in each of the articles published using the UESD (see Table 2 for the aggregated figures). The criteria used to classify each study for each of the tests/checks are outlined in Table B2. To classify each of the studies we take into consideration both the main text and the appendixes or supplemental information of each study (when available).

Table B1 catalogs the published social science studies that rely on the UESD of which we are aware. These studies analyze the impact of different type of events on multiple outcomes, and they all do so applying the UESD as defined in Section 2. This classification is based on a systematic search of bibliographic databases (Web of Science and Google Scholar), as well as consultation with researchers who have used the UESD.

Table B1, therefore, provides a comprehensive overview of the empirical strategy and robustness checks conducted in UESD studies. The UESD definition provided in Section 2 excludes some published studies that examine the impact of events through surveys, though. This is the case, for example, of studies that rely on a survey fielded some time before the event and another survey fielded some time after the event (or multiple surveys stacked around the event), but without an overlap between the fieldwork and the event (see e.g. Bisgaard and Slothuus, 2018; Brouard et al., 2018). Similarly, we exclude studies that track the association between media coverage of multiple –and sometimes unrelated– events during a prolonged period of time, and time-trends in public opinion (see e.g. Zaller and Hunt, 1995).¹ Studies in which a survey is purposefully fielded to capture the effects of an expected event by including questions directly related to the event are also excluded (see e.g. Motta, 2018). While these studies analyze the impact of events on public opinion through surveys, their design has certain particularities and methodological challenges (e.g. time-series dynamics) not directly addressed in our paper, which focuses on the increasingly common practice of estimating the individual-level effects of unexpected events that occur during the fieldwork of a single survey.

¹While these studies analyze the impact of campaign events, they do not usually focus on isolating the effect of a single event. Instead they include multiple (expected and unexpected) events and track public opinion reactions to those. Therefore, many of the issues discussed in our paper do not apply to these studies

Table B1: Empirical strategy and robustness checks of published studies drawing on the UESD

	In-depth description of the event	In-depth description of the survey	Balance tests	Multiple bandwidths	Covariate adjustment	Analysis of non-response	Placebos in the control group	Inspection of pre-trends	Falsification through other surveys/units	Falsification through other outcomes	Pseudo-manipulation checks	Restricted interpretation of treatment effect
Ares and Hernández (2017)	✓		✓	✓	✓				✓	✓		
Balcells and Torrats-Espinosa (2018)	✓	✓	✓	✓	✓	✓						
Berger (2010)				✓	✓			✓			✓	
Bonilla and Grimmer (2013)	✓		✓		✓						✓	
Boomgaarden and de Vreese (2007)	✓		✓								✓	
Boydston et al. (2018)			✓		✓							
Bozzoli and Müller (2011)	✓				✓							
Branton et al. (2015)			✓		✓							
Burlacu et al. (2018)	✓				✓							
Carey Jr et al. (2014)					✓							
Coupe (2017)					✓						✓	
Das et al. (2009)						✓						
De Vries (2018)			✓		✓							
Dinas et al. (2016)		✓	✓	✓	✓	✓		✓	✓	✓		
Dinesen and Jæger (2013)			✓	✓	✓							
Finseraas and Listhaug (2013)		✓	✓	✓	✓		✓		✓	✓		
Finseraas et al. (2011)	✓	✓	✓	✓	✓			✓		✓		
Flores (2018)	✓	✓	✓	✓	✓					✓		
Geys and Qari (2017)			✓	✓	✓		✓			✓	✓	
Hariri et al. (2015)	✓	✓	✓	✓	✓				✓			
Hofstetter (1969)	✓		✓									
Holbrook et al. (2012)	✓				✓							
Jakobsson and Blom (2014)	✓		✓	✓	✓				✓			
Jensen and Naumann (2016)			✓	✓	✓		✓	✓		✓		
Kim and Kim (2019)	✓	✓	✓		✓	✓		✓	✓	✓		✓
Krosnick and Kinder (1990)	✓		✓		✓							
Larsen (2018)	✓		✓	✓	✓		✓	✓		✓		
Legewie (2013)	✓		✓		✓	✓						
Lever (1969)	✓				✓			✓				
Metcalfe et al. (2011)		✓	✓		✓	✓		✓	✓			
Minkus et al. (2018)	✓		✓		✓		✓	✓	✓			
Muñoz and Anduiza (2019)	✓	✓	✓	✓	✓	✓	✓	✓		✓		
Nagoshi et al. (2007)			✓									
Ojeda (2016)					✓							
Perrin and Smolek (2009)			✓		✓							
Pierce et al. (2016)			✓	✓	✓							
Reeves and de Vries (2016)	✓	✓	✓	✓	✓	✓			✓	✓		
Schaffner and Roche (2017)	✓	✓	✓		✓					✓		
Schulz et al. (2008)	✓		✓		✓							
Silber Mohamed (2013)	✓	✓	✓		✓							
Silva (2018)			✓	✓	✓						✓	
Slothuus (2016)	✓				✓			✓				
Solaz et al. (2018)	✓		✓	✓	✓					✓		
Zepeda-Millán and Wallace (2013)					✓							

Table B2: Criteria for the classification of UESD studies

Test/Check	Criteria
In-depth description of the event	Discusses when did the event happen, the circumstances leading to the event, the circumstances surrounding it, the post-event context, and provides some direct or indirect evidence of the salience or news coverage of the event.
In-depth description of the survey	Provides details about: the survey sampling procedure, the duration and development of the fieldwork, the fieldwork protocols, and the strategies to deal with unit non-response (e.g. timing of refusal conversion strategies).
Balance tests	Analyzes balance on pre-treatment covariates between the control and treatment group(s).
Multiple bandwidths	Discusses results based on multiple bandwidths around the time of the event.
Covariate adjustment	In order to estimate the effect of the event uses some form of covariate adjustment such as controlling through regression or any other method to improve covariate balance between the treatment and control groups.
Analysis of non-response	Analyzes item and/or unit non-response before and after the event.
Placebo treatments at different dates of the control group	Conducts placebo tests at arbitrary points at the left of the cutoff point ($t_p < t_e$)
Inspection of pre-existing time trends	Examines the relationship between the timing of interviews and the outcome of interest during the pre-event period in order to rule out the presence of pre-existing time trends.
Falsification tests through other surveys/units	Conducts a falsification test of a simulated event that occurs on the same date of the event (or at a similar time point during the fieldwork) using another edition/round of the survey used to analyze the event of interest.
Falsification tests through other outcome variables	Conducts falsification test by testing for the effects of t on an alternative outcome Y' that is theoretically unrelated to T
Pseudo-manipulation checks	Analyzes if there is a perceptible increase in the salience respondents attribute to problems related to the event of interest (through a “most important problem” type of question). We consider that a study performs this pseudo-manipulation check even if the analysis of the “most important problem question” (or a similar question) is not explicitly conducted as a robustness check to assess non-compliance.
Restricted interpretation of the treatment effect	Interprets effect of the event as an <i>Intent-To-Treat</i> (ITT)

C Charlie Hebdo attacks: Event and survey description

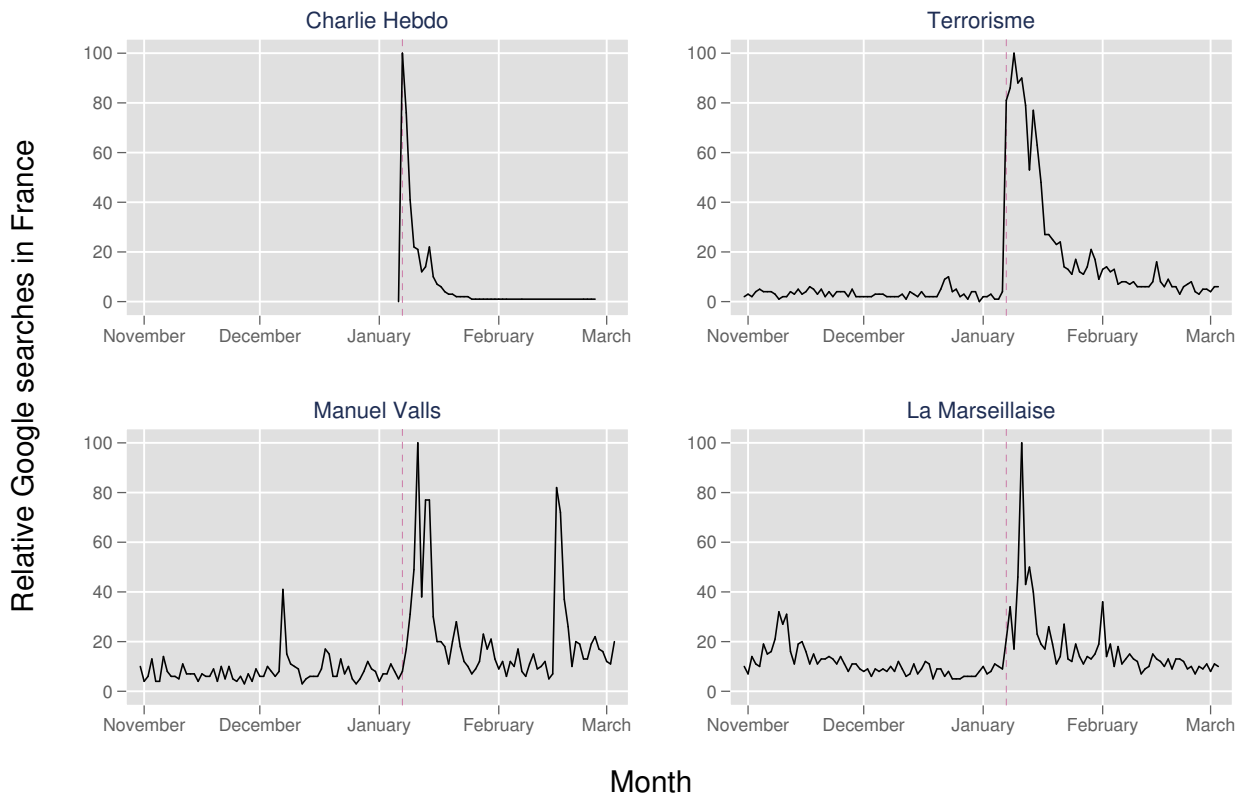
C.1 Characterization of the Charlie Hebdo attacks

The Charlie Hebdo terrorist attacks occurred in Paris on January 7th, 2015. At approximately 11:30 of January 7th, 2015, the brothers Chérif and Saïd Kouachi, identifying as members of Al-Qaeda, burst into the offices of the satirical magazine Charlie Hebdo and killed 12 people and injured 11 others. Immediately after the attack, a manhunt began. The manhunt lasted until January 9th when the two terrorists were shot and killed by the police. In parallel to the manhunt, an acquaintance of the Kouachi brothers (Amedi Coulibaly) shot a policeman on January 8th, and the day after he killed 4 hostages he had taken at a Kosher supermarket in Paris while he was in contact with the Kouachi brothers. Coulibaly was also shot and killed by the police on January 9th. At the time these were the most serious terrorist attacks in France since 1995, which also had an important symbolic impact since, until then, French citizens did not think of themselves as a target of lethal terrorism (Brouard et al., 2018; Moran, 2017).

The attacks sparked multiple reactions. On January 11th a demonstration against the attacks gathered more than 2 million people in Paris, including many presidents and prime ministers of allied countries who stood next to French leaders such as President Hollande, Prime Minister Valls, or former President Nicolas Sarkozy. In this context of intense display of national unity, political leaders such as President Hollande or Prime Minister Valls framed the shootings as an attack on core French values (Moran, 2017). In fact, the national government was “keen to rally public opinion under the ‘Je suis Charlie’ banner” (Moran, 2017, p. 318). These displays of national unity were accompanied by the deployment of military troops in public spaces and the announcement of numerous anti-terror measures by Prime Minister Valls (Willsher, 2015). This climate of national unity was exemplified on January 13th when, after singing the national anthem, all members of parliament voted in favor of continuing air strikes against Daesh in Iraq.

The Charlie Hebdo attacks were a highly salient political event. In the weeks that followed, the attacks and the reactions they generated were front page news in most French newspapers. In fact, in the ESS media claims dataset, which provides information about the news that are most salient during the fieldwork period as published in the newspapers *Le Figaro* and

Figure C1: Google trends for key search terms during the survey fieldwork



Note: The dashed red line denotes January 7th

Libération, almost all entries (claims) are related to the Charlie Hebdo attacks until January 20th (Fernee et al., 2014). During those days, the actor with the highest number of entries in the claims dataset is the head of the government, PM Manuel Valls. The Internet search trends summarized in Figure C1 also reflect that the terms ‘Charlie Hebdo’ and ‘Terrorisme’ rapidly rose in salience in the aftermath of the attacks. Google trends also reveal a clear spike in the number of searches about ‘Manuel Valls’ and the French national anthem ‘La Marseillaise’ immediately after the shootings.

The Charlie Hebdo attacks were preceded by a context of unpopularity of the French government, which some attributed to the implementation of austerity policies (see Moran, 2017). Moreover, before the January 7th shootings, three other serious terrorist incidents occurred in December. While these events led to heightened security and the Prime Minister claimed that ‘we never faced a greater risk in matter of terrorism’ already in December 22nd, we would argue that the Charlie Hebdo attack was still an unexpected event, the specific timing of which was nowhere near foreseeable. The ESS media claims dataset reveals, that the terrorist attacks and the reactions it generated dominated the news in the weeks following January 7th. It was not

until around January 25th that other major political topics, like the ‘Law for growth, activity and equal economic opportunity (the ‘Macron Law’)’ gained prominence again.

C.2 Background information on ESS-7 fieldwork in France

The fieldwork of 7th round of the ESS was conducted between October 31st, 2014 and March 3rd, 2015 through face-to-face computer-assisted personal interviews (CAPI). The ESS sampling guidelines establish that samples must be representative of all persons aged 15 or older, that sampling frames of individuals, households and addresses can be used, that individuals are selected by random probability methods at each stage, and that quota sampling or substitution of non-responding sampled units is not permitted at any point.²

In the case of the 7th round in France, the ESS was not granted access to the census sampling frame and the sample was selected by drawing of addresses within each Primary Sampling Unit (municipality or group of municipalities) and then randomly selecting dwellings and individuals within dwellings. Sampled units were contacted in person by interviewers, and non-contact and refusal conversion strategies began immediately after the first failed visit. The absence of sampling quotas, the use of random probability methods at each stage, and the immediate start of refusal conversion strategies are likely to minimize some of the potential biases related to the fieldwork of the survey. However, the long duration of the fieldwork, the use of a 4-stage sampling design, and the late occurrence of the event generate some challenges that must be taken into account when estimating and interpreting the effects of the Charlie Hebdo attacks.

With regard to the development of the fieldwork, it is relevant to note that the fieldwork was not interrupted by the Charlie Hebdo terrorist attacks. Interviews were conducted on January 7th after 11:30, as well as in January 8th and the following days. Even in the region most directly affected by the attacks (Ile de France) interviews were conducted during all those days.

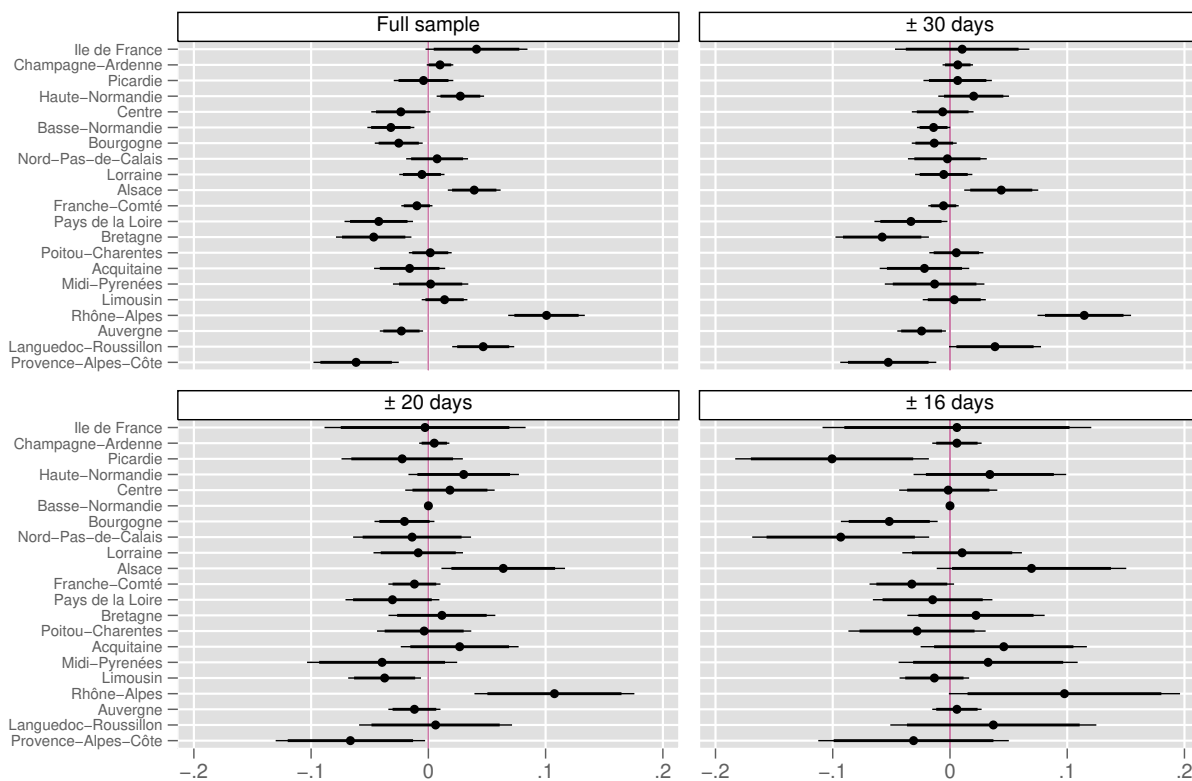
²See ESS sampling guidelines.

D Additional analyses of Charlie Hebdo attacks

D.1 Regional imbalances

Figure D1 summarizes the regional imbalances between the treatment and the control groups. The balance tests indicate that even when relying on a ± 20 days bandwidth there are still some statistically significant imbalances in the regions of Lorraine, Limousin, Rhne-Alpes and Provece-Alpes-Cte. Relying on an even narrower bandwidth of ± 16 days would not eliminate these imbalances. In fact, for the ± 16 days bandwidth some of these imbalances are even larger than for the ± 20 bandwidth.

Figure D1: Balance tests: Covariate differences between the treatment and control groups.

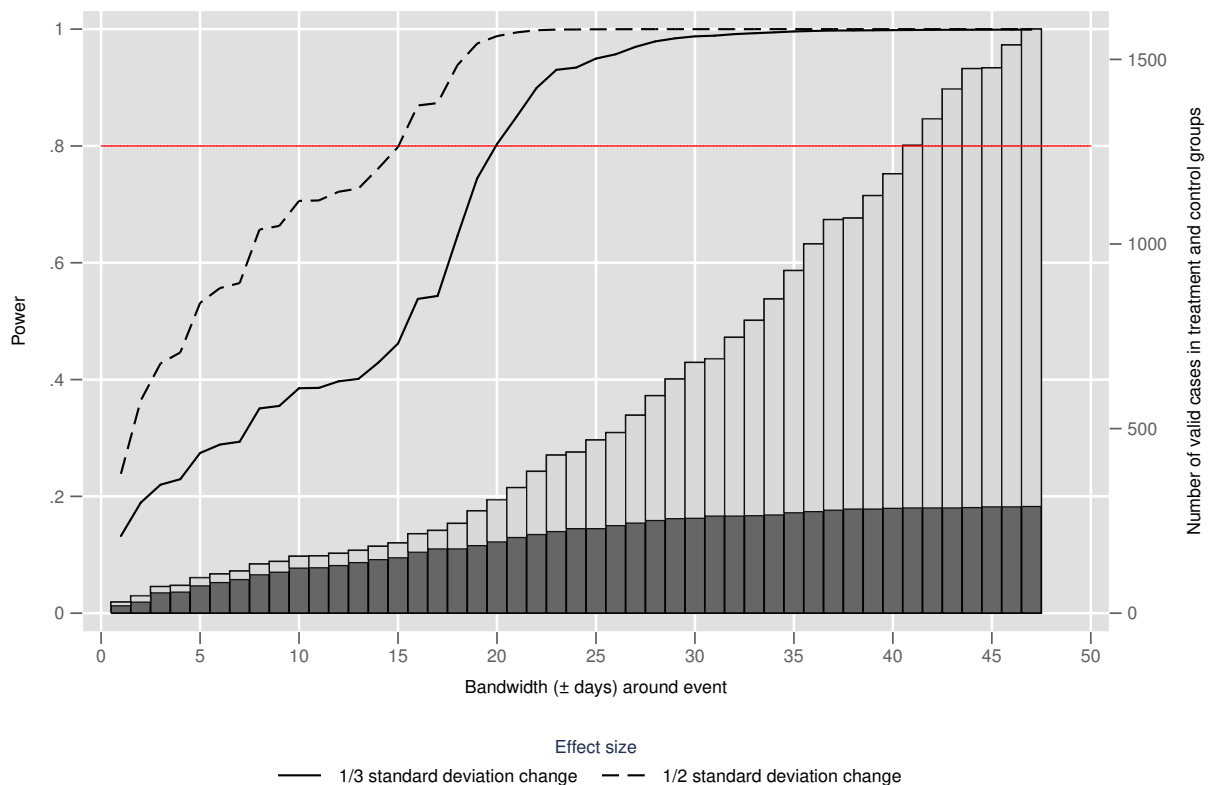


Note: Entries report the difference in the mean of the covariates between the treatment and control groups. Thick and thin lines are 95% and 90% confidence intervals, respectively

D.2 Bandwidths and statistical power

Figure D2 summarizes the power calculations for two different effect sizes. To determine the effect sizes for the power calculation we take as a reference the distribution and the standard deviation of the “Satisfaction with government” variable among those interviewed before the event (which equals 2.0 in this case). As a lower bound for a meaningful effect we consider an increase in the mean level of satisfaction with government equivalent to one third of a standard deviation (+0.67). The ± 20 bandwidth is the first bandwidth with a power higher than 0.8 for this effect size. We also estimate the power for a larger effect size: an increase in the mean level of satisfaction with government equivalent to one half of a standard deviation (+1.0). The ± 16 bandwidth is the first bandwidth with a power higher than 0.8 for this effect size. In addition to the power calculations, the stacked histogram of Figure D2 summarizes the number of cases included in the treatment and control groups for each bandwidth.

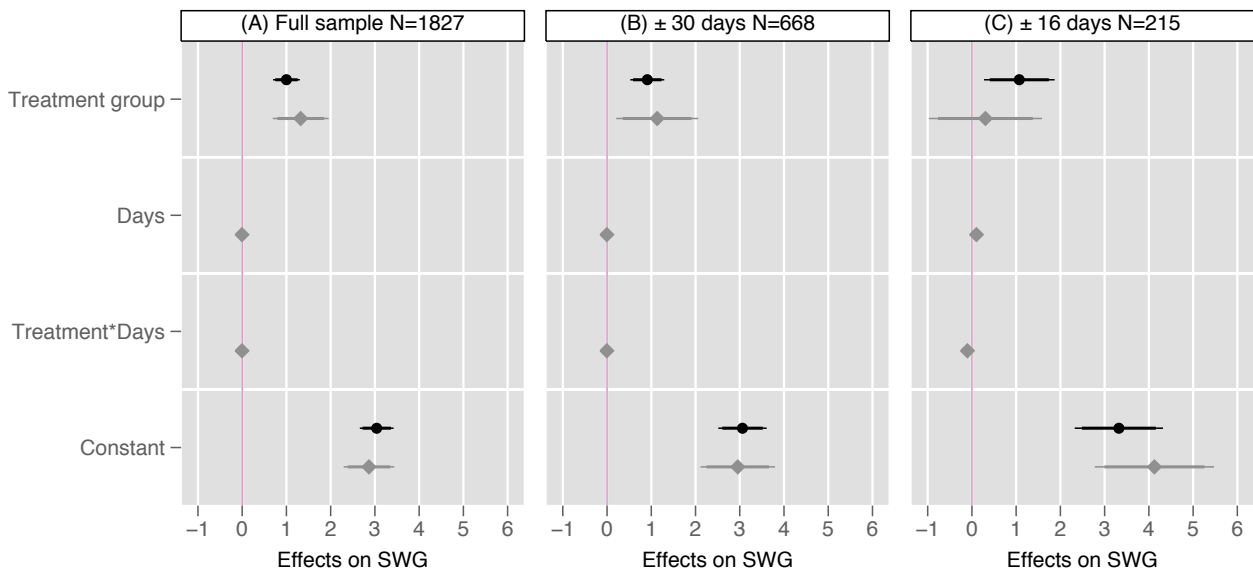
Figure D2: Statistical power for various effect sizes and bandwidths



Note: Power calculations for a 0.05 significance level and based on the standard deviation of the control group. In the histogram the total height of the stacked bars refers to the total number of valid cases included with each bandwidth, and the black and grey bars refer to the number of cases in the treatment and control groups, respectively. Valid cases are those respondents who provide a valid answer for all the covariates used in the estimation of the effects of the Charlie Hebdo attacks

Figure D3 summarizes the effect of the Charlie Hebdo attacks through the bandwidths considered in the balance tests summarized in Figure 3. The table replicates the main models used to estimate the effects of the terrorist attacks summarized in Panel C of Figure 2 (in the main text), which include region-fixed effects and entropy balancing weights. Before estimating these models we have re-estimated the entropy balancing weights for each bandwidth. To achieve convergence with all bandwidths reported in this appendix, the entropy balancing weights used in Figure D3 and Figure D5 only adjust for the first and second moment conditions of the covariates (mean and variance).

Figure D3: Effects of Charlie Hebdo attacks on satisfaction with government (SWG).

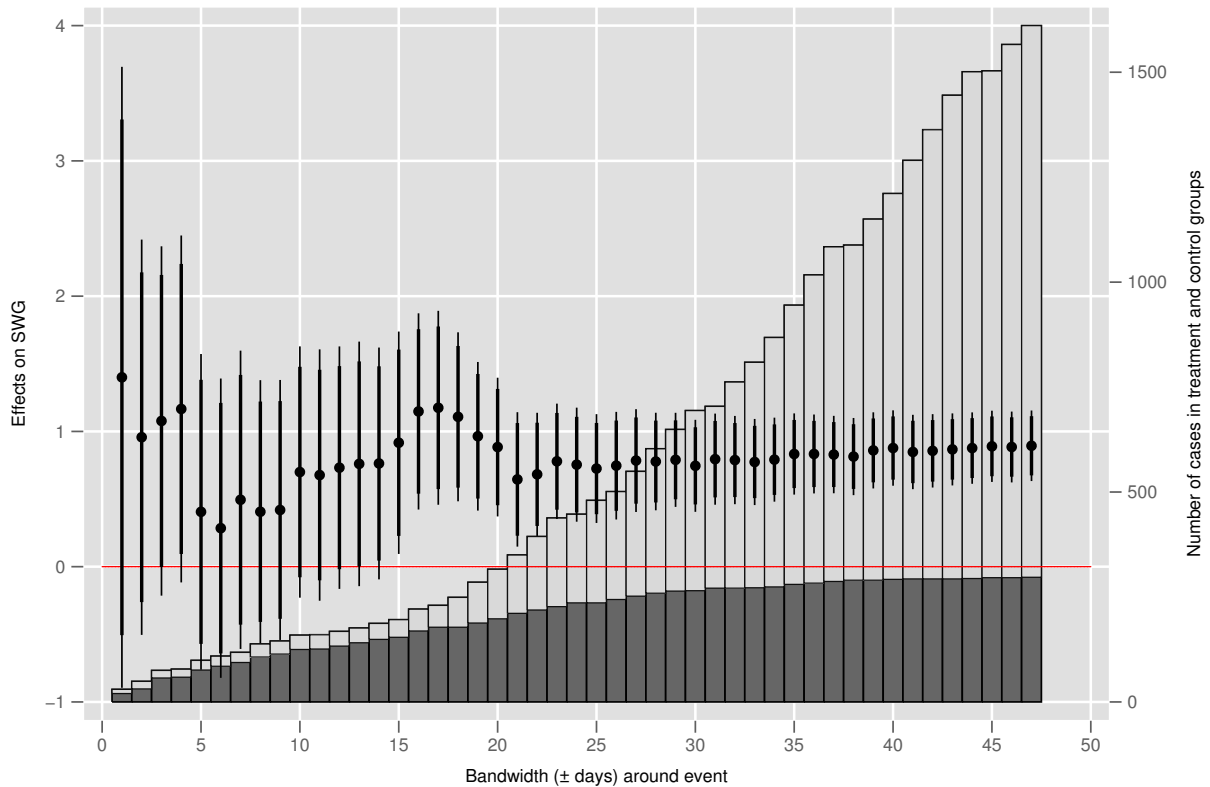


Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. All models are estimated with entropy balancing weights and region fixed-effects.

Next, we summarize the results of the Charlie Hebdo attacks increasing the bandwidth around the event day by ± 1 days until reaching the last day of the fieldwork. However, as the histograms in the figures show, in the case of narrow bandwidths there is a very limited number of cases, especially in the control group. For this reason, Figure D4 summarizes these results without any type of covariate adjustment. In this case, we observe that the estimated effect of the attacks is consistently positive, although for bandwidths smaller than ± 15 the results are not statistically significant. Figure D5 summarizes the results for those bandwidths for which we can balance the control and treatment groups through entropy balancing (with a

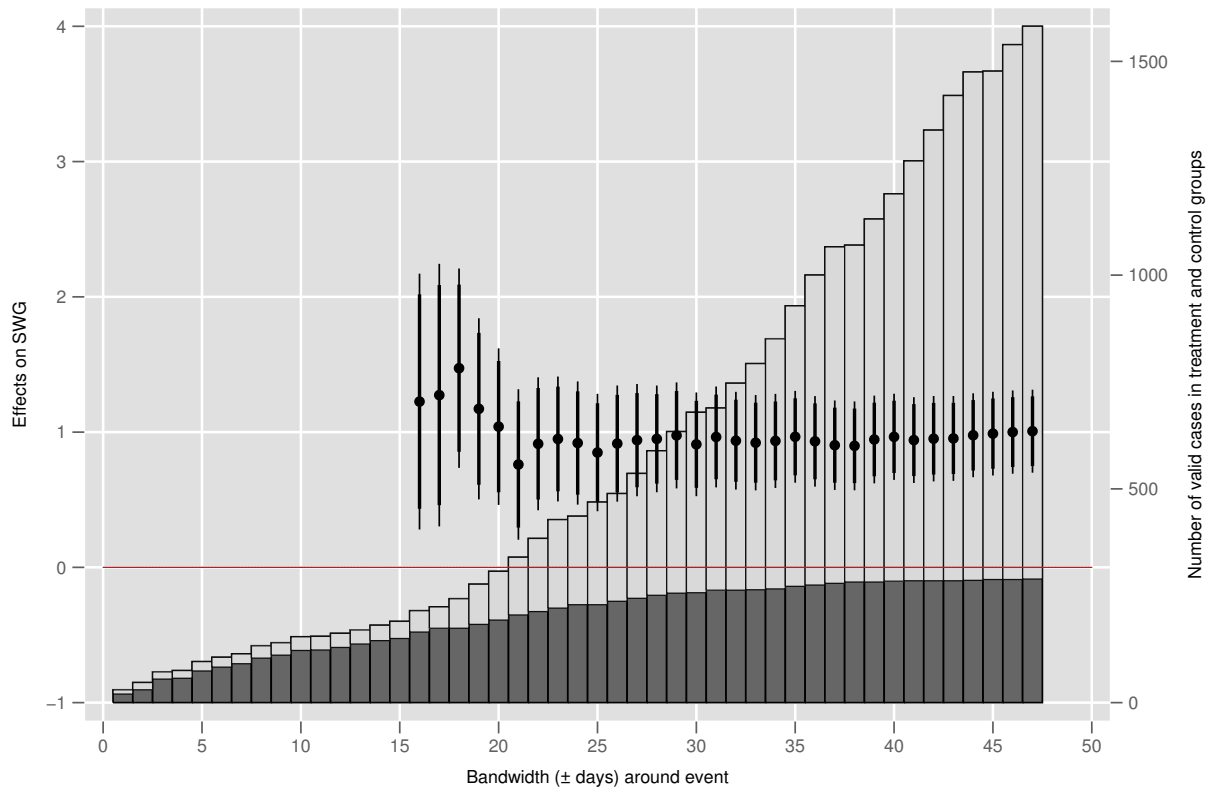
0.015 tolerance) at least when it comes to the mean and variance of the covariates discussed throughout the paper. This limits the analyses to the bandwidths equal or larger than ± 16 . In this case we observe that the estimated effects of the Charlie Hebdo attacks are consistently positive and statistically significant.

Figure D4: Effects of the Charlie Hebdo attacks for multiple bandwidths: no covariate adjustment



Note: Results from OLS models that only include the treatment indicator and without any type of covariate adjustment. Thick and thin lines are 90% and 95% confidence intervals, respectively. In the histogram the total height of the stacked bars refers to the total number of cases included with each bandwidth, and the black and grey bars refer to the number of cases in the treatment and control groups, respectively.

Figure D5: Effects of the Charlie Hebdo attacks for multiple bandwidths: entropy balancing and region fixed effects



Note: Results from OLS models estimated with entropy balancing weights and region fixed-effects. Entropy balancing weights adjust for two moment conditions (mean and variance) but not for skewness. Thick and thin lines are 90% and 95% confidence intervals, respectively. In the histogram the total height of the stacked bars refers to the total number of valid cases included with each bandwidth, and the black and grey bars refer to the number of cases in the treatment and control groups, respectively. Valid cases are those respondents who provide a valid answer for all the covariates used in the estimation of the effects of the Charlie Hebdo attacks

D.3 Entropy balancing details

Table D1 summarizes the distribution of the observable covariates without the entropy balancing weighting, and Table D2 summarizes this distribution in the reweighted control group (i.e. after adjusting for the first, second and third moment of all the covariates with a tolerance level of 0.015).

Table D1: Covariate distribution (no weighting)

	Treatment group			Control group		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Age (years)	48.052	367.133	0.270	48.921	315.418	0.089
Education (years)	13.332	18.983	0.124	13.570	18.814	-0.067
Citizenship	0.902	0.089	-2.696	0.956	0.042	-4.455
Male	0.420	0.245	0.325	0.544	0.250	-0.176
In a paid job	0.539	0.250	-0.156	0.553	0.249	-0.212
Unemployed	0.078	0.072	3.155	0.053	0.050	4.007
Student	0.114	0.102	2.429	0.096	0.088	2.733
Retired	0.259	0.193	1.100	0.246	0.187	1.182
Housework	0.036	0.035	4.961	0.044	0.042	4.455
Voted last election	0.554	0.248	-0.219	0.649	0.230	-0.625
Voted PS last election	0.135	0.117	2.140	0.219	0.173	1.357
Refusals to participate	0.192	0.208	2.352	0.114	0.120	3.052

Initially, one would also want to balance the treatment and control groups with regards to the number of attempts to survey before the interview was completed. This would allow for a better adjustment of the potential differences in reachability between the two groups. However, like any other method of covariate adjustment, entropy balancing is not an infallible method for achieving covariate balance (Hainmueller and Xu, 2013). If the two groups are too different with regard to a certain characteristic (e.g. the number of attempts to survey) there will not be enough information to identify the counterfactual, and the control group weights generated thorough the entropy balancing might be unrealistic (Hainmueller and Xu, 2013). This seems to be the case for the variable identifying the numbers of attempts to survey each individual. Figure D6 reveals that in the treatment group we can find a substantial number of

respondents who participated in the survey after 8 or more attempts to survey them. However, in the control group only two respondents were surveyed after 8 attempts, and no respondent was surveyed after more than 8 attempts. As a consequence, if one reweights the control group with entropy balancing weights that include the variables measuring the number of attempts to survey the weights of some of the subjects in the control group are unrealistic and are likely to distort the results. Figure D7 summarizes the number of respondents for each value of the ‘attempts to survey’ variable and the average balance weights for each of these values when considering this variable for the entropy balancing. For example, while the average weight of the 18 respondents who were surveyed after 1, 2 or 3 attempts equals 0.07, the average weight for the two respondents that were surveyed after 8 times equals 32.42. Therefore, when using these weights, the results of the control group could be driven by the two observations that were surveyed after 8 attempts. Since this seems to be a case of what Hainmueller and Xu (2013) characterize as unrealistic weights we exclude the variable measuring the number of attempts to survey from the entropy balancing. Researchers relying on the UESD and adjusting for covariate imbalances should always assess these potential threats and decide which covariates can be adjusted.

Table D2: Covariate distribution (reweighted control group)

	Treatment group			Control group		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Age (years)	48.052	367.133	0.270	48.074	367.291	0.267
Education (years)	13.332	18.983	0.124	13.338	18.993	0.119
Citizenship	0.902	0.089	-2.696	0.902	0.089	-2.704
Male	0.420	0.245	0.325	0.419	0.246	0.328
In a paid job	0.539	0.250	-0.156	0.539	0.251	-0.157
Unemployed	0.078	0.072	3.155	0.078	0.072	3.159
Student	0.114	0.102	2.429	0.113	0.101	2.438
Retired	0.259	0.193	1.100	0.258	0.193	1.106
Housework	0.036	0.035	4.961	0.036	0.035	4.967
Voted last election	0.554	0.248	-0.219	0.555	0.249	-0.221
Voted PS last election	0.135	0.117	2.140	0.134	0.117	2.147
Refusals to participate	0.192	0.208	2.352	0.192	0.208	2.348

Figure D6: Distribution of attempts to survey in the treatment and control groups

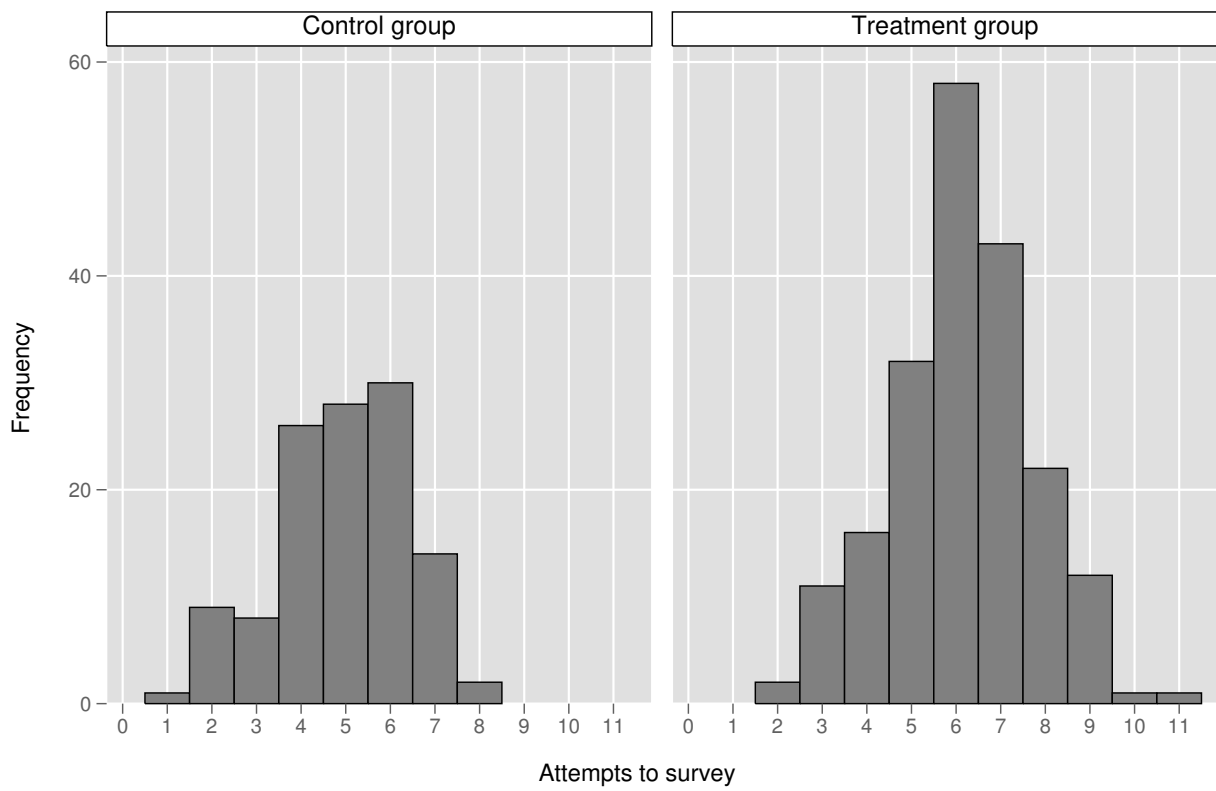
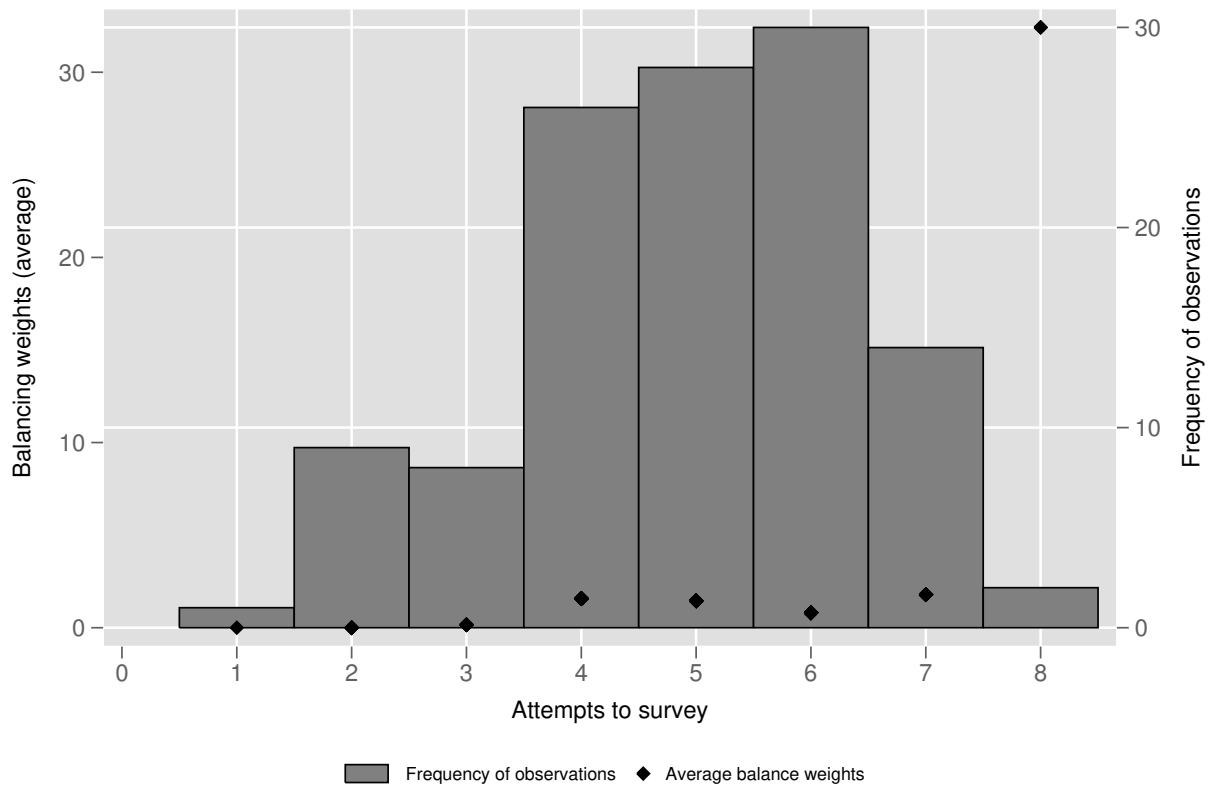


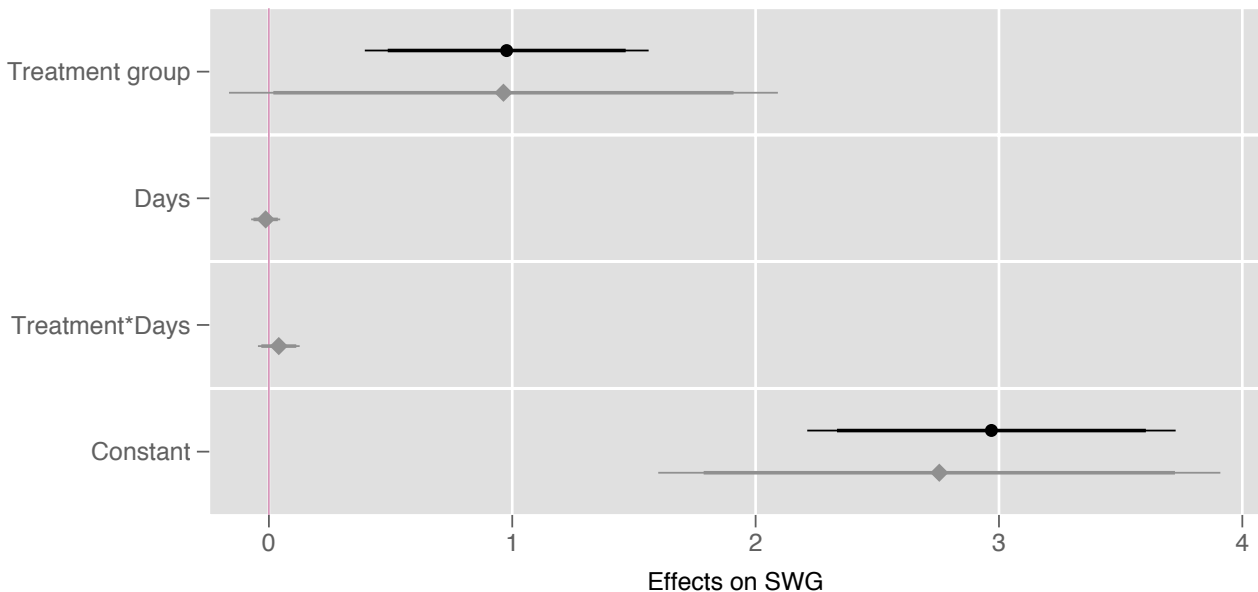
Figure D7: Average entropy balancing weights for each level of attempts to survey in the control group



D.4 Entropy balancing excluding vote recall

Whether respondents voted or not in the previous election, and the party they voted for is reported by the respondents themselves. Despite the fact that vote recall questions refer to an election held prior to the event, vote memories are often known to be biased by current preferences and social desirability (Himmelweit et al., 1978; Van Elsas et al., 2013). Since we expect the Charlie Hebdo attacks to affect evaluations of government, respondents might, therefore, modify their answers to these questions as a consequence of the terrorist attack. Respondents might be less willing to report that they voted for the party in government (or that they turned out to vote) after the attacks. We believe that this is a remote possibility since the survey questions ask about past behaviors and not about a future behavior (e.g. intention to turn out to vote) or an attitudinal orientation towards the party in government (e.g. party identification or closeness to the PS). In any case, to assess the potential post-treatment bias that this variable could induce we re-estimate the main models used to assess the impact of the terrorist attacks without balancing the control and treatment groups with respect to the vote recall variables. The results of Figure D8 are very similar to those of the models summarized in Panel C of Figure 2.

Figure D8: Effects of Charlie Hebdo attacks on satisfaction with government (SWG).



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. The models are estimated with entropy balancing weights and region fixed-effects.

D.5 Pre-existing time trends

To assess the plausibility of the temporal stability assumption we estimate the effect of a placebo treatment at the left of the cutoff point ($t_p < t_e$) using the empirical median of the control group as the placebo treatment point, which corresponds to December 20. Therefore, the “placebo treatment” variable takes the value 0 for those interviewed between December 17-20 and the value 1 for those interviewed between December 22 and January 6.³ The results summarized in Table D3 reveal that the change in satisfaction with government is small and not statistically significant, both focusing on the average levels at the two sides of the placebo cutoff (Model 1) and on the jump at the cutoff (Model 2).

Table D3: Placebo treatments. Dependent variable: satisfaction with government. OLS models

	Control group		Pre-event period	
	(1)	(2)	(3)	(4)
Placebo treatment	-0.21 (0.50)	-0.15 (1.31)		
Days		-0.28 (0.37)	0.00 (0.00)	0.03 (0.02)
Placebo treatment * Days		0.37 (0.38)		
Days ²				0.00 (0.00)
Constant	3.38 (0.52)	2.61 (1.21)	3.14 (0.22)	3.58 (0.38)
Region fixed-effects	Yes	Yes	Yes	Yes
Observations	115	115	1574	1574
R^2	0.172	0.197	0.018	0.019

Standard errors in parentheses.

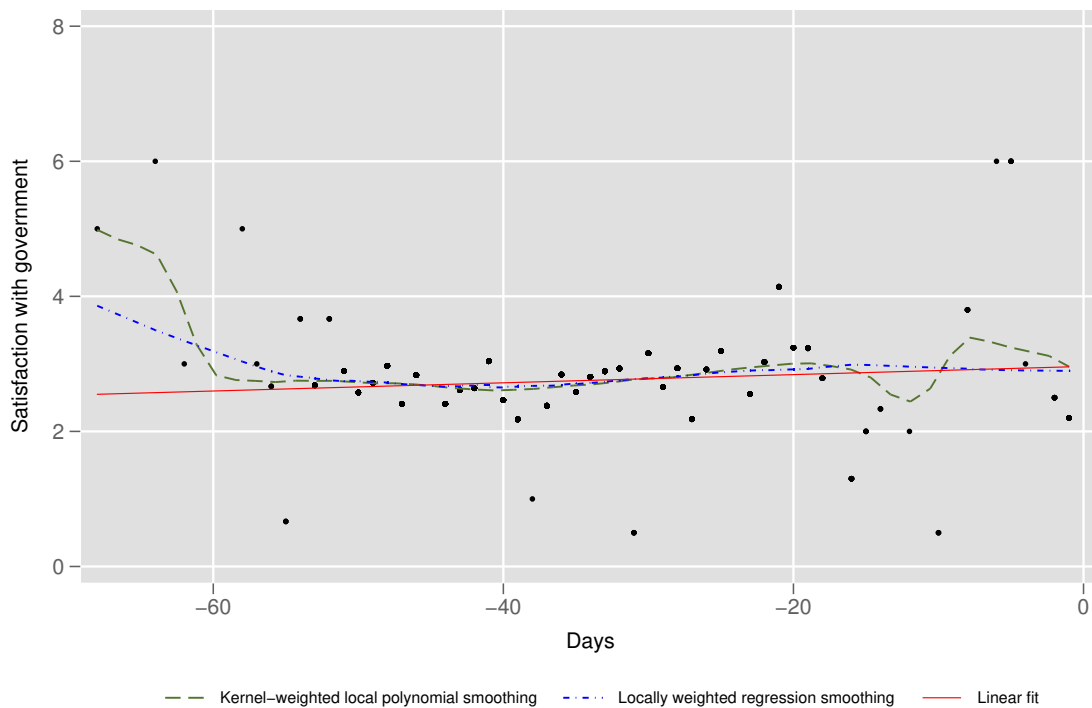
Note: In model 2 the variable days takes the value 0 on December 22.

We also analyze the relationship between the timing of interviews and the outcome variable during the whole survey fieldwork period before the event. Figure D9 illustrates the relationship between satisfaction with government and the timing of interviews in the pre-event period, and

³There were no interviews on December 21.

Models 3 and 4 of Table D3 test for the existence of a linear and quadratic relationship between these two variables. While these variables are not systematically related, the daily average of satisfaction with government reveals some substantial jumps at points other than the event date. Therefore, in this case, it would not be advisable to draw conclusions about the effects of the event exclusively on the basis of analyses conducted at a specific point in time (e.g. the day immediately after the event).

Figure D9: Average satisfaction with government by day in pre-event period (0 = January 8th).



D.6 Attrition and non-response

The rate of non-response for the “satisfaction with government” question is not affected by the Charlie Hebdo attacks. The proportion of item non-response (including refusals and *don't know* answers) in the control period is of 0.025 and of 0.015 in the treatment period. A two-sample t-test (not shown) indicates that this difference is not statistically significant.

Table D4 summarizes the number of sampled units who refused to answer the survey or were not surveyed for any other reason during the control and treatment periods. The proportion of refusals over the total number of attempted interviews is very similar in both periods. When analyzing all forms of unit non-response the results also indicate that the proportion of sampled units that do not respond is lower after the event.

Table D4: Survey non-response during the control and treatment periods

	Refusals	Unit non-response	Total attempted interviews
Control period	159 (11%)	930 (67%)	1396
Treatment period	223 (11%)	1,156 (55%)	2090

Note: The category “Unit non-response” also includes those who refused to participate in the survey. Percent over total attempted interviews in each period in parentheses

In order to judge if attrition might bias the findings, though, it is more appropriate to analyze the characteristics of non-respondents before and after the event. To analyze if attrition and non-response might be affected by the event one would ideally rely on information originating from the sample frame. The sample frame generally provides information such as the age and gender of the sampled units, even of those that did not finally participate in the survey. However, in the case of the 7th round of the ESS in France, the survey organization was not granted access to the sampling frame. In the absence of this information we rely on the interviewers’ judgment of the age, gender and contextual conditions of the house of the sampled units that were not interviewed during each period (we use the latter as a proxy for the socioeconomic status of non-respondents). Relying on the interviewers’ judgment of the characteristics of respondents imposes an additional limitation on our analysis of non-response, since interviewers can only judge the age and gender of those sampled units who refuse to participate in the survey. It is

not possible for interviewers to judge the age and gender of the sampled units that can not be contacted (unit non-response).

These analyses, summarized in Table D5, reveal no major differences in the characteristics of those who refused or could not be interviewed during the control and treatment periods. This suggests that our estimations are unlikely to be biased by attrition related to the attacks.⁴

Table D5: Characteristics of sampled units who refused or did not respond during the control and treatment periods

	Refusals			Unit non-response		
	Control	Treatment	t-stat.	Control	Treatment	t-stat.
Age (4 categories)	3.09	3.07	0.197			
Gender	0.41	0.45	-0.680			
House physical condition	3.08	3.03	0.513	2.99	2.93	1.156
Amount litter and rubbish	3.80	3.80	0.135	3.78	3.77	0.420
Amount vandalism and graffiti	3.83	3.79	0.673	3.85	3.82	0.991
Observations	127	191		467	693	

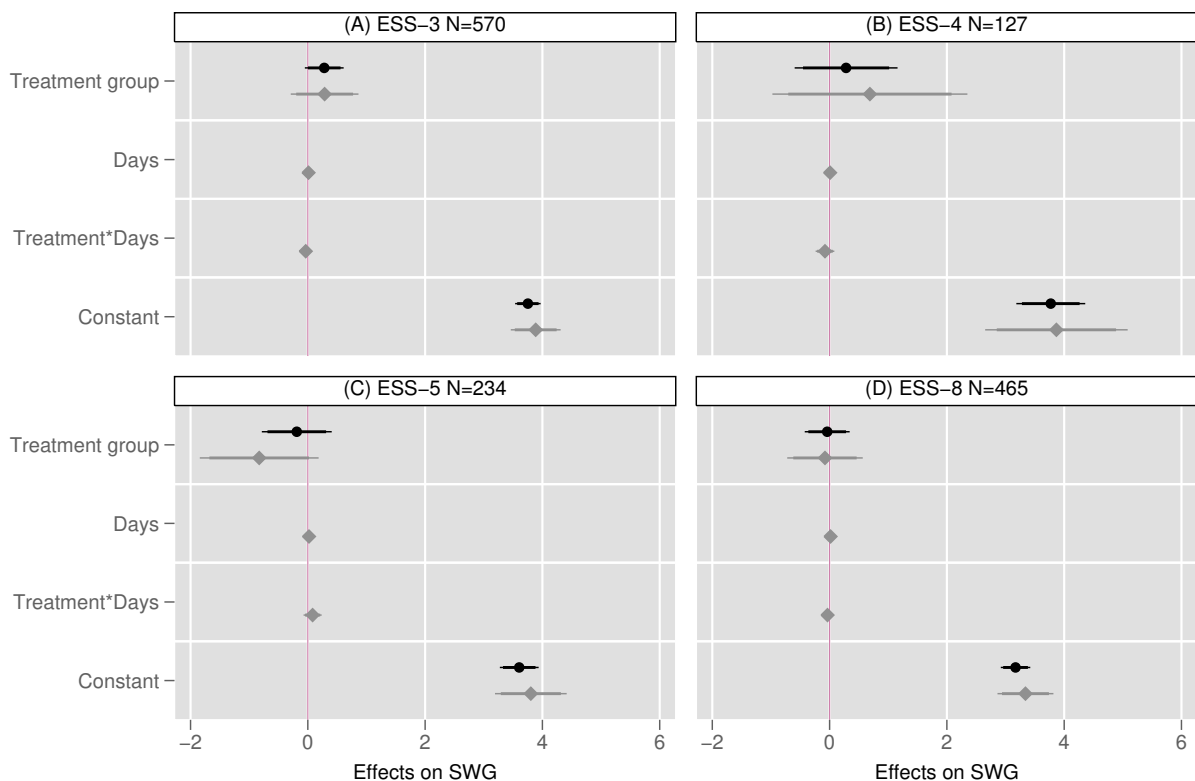
Note: Sampled units that refused or did not respond both during the treatment and control periods are excluded from this table. One observation per sampled unit. All variables are estimated by the interviewer.

⁴The difference in the number of observations between Table D5 and Table D4 is explained by the fact that in Table D5 each sampled unit is included just once, while in Table D4 each sampled unit might be included more than once.

D.7 Falsification tests based on other surveys

To test for the possibility that the estimated effects of the Charlie Hebdo attacks could be driven by aspects related to the timing of the survey administration or by any other cyclical trends related to people’s satisfaction with government, we use data from four rounds of the French ESS in which the fieldwork period included January 7th.⁵ For each round, we generate a placebo independent variable that mirrors our original treatment indicator, which compares those interviewed before and after January 7th with a bandwidth of ± 20 days. The results of these tests summarized in Figure D10 reveal that in none of these rounds the placebo treatment has a statistically significant effect on satisfaction with government that is comparable in size to the effect that we identify for the Charlie Hebdo attacks.

Figure D10: Placebo tests with rounds 3, 4, 5, and 8 of the European Social Survey.



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively.

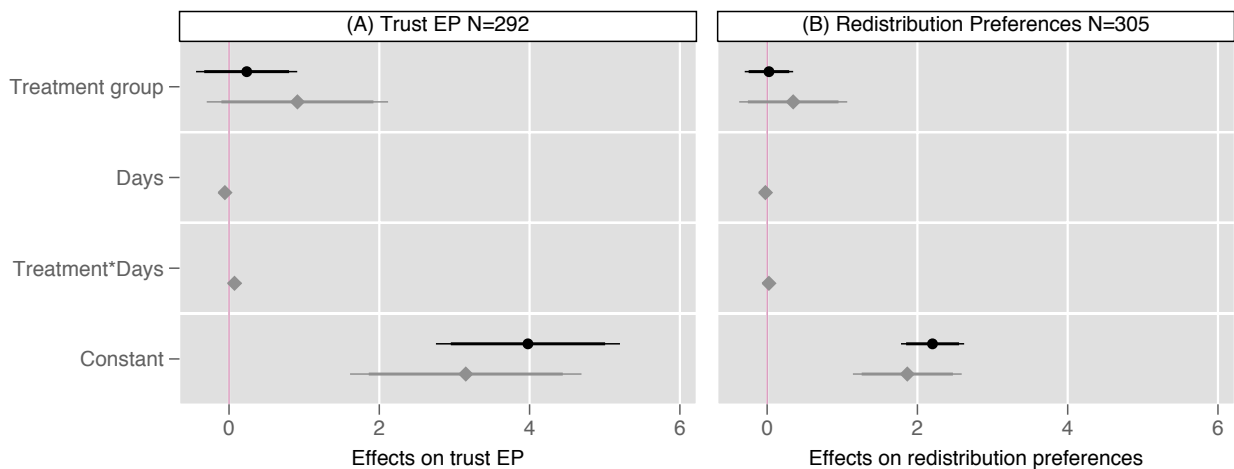
⁵The second round of the ESS in France also includes January 7th. However, we do not consider this round because the date of the interview was recorded differently than in the rounds that followed.

D.8 Falsification tests based on other outcome variables

We also test if the attacks had any impact on a series of variables that are close to our outcome of interest but should not be directly affected by the attacks. These variables measure respondents' trust in the European Parliament (EP) and their redistribution preferences. The first of these variables is selected because like our key variable of interest (satisfaction with government) it measures political support, but for an international institution that should not be directly affected by a "rally round the flag" effect. The second variable is selected in order to capture potential changes in preferences related to the support of a left-wing government.

The results of these tests are summarized in Figure D11. The effects of the treatment indicator are of reduced magnitude and not statistically significant when using any of these alternative outcomes.

Figure D11: Falsification tests: Effect of the Charlie Hebdo attacks on theoretically unrelated dependent variables.



Note: Black spikes (circle markers) summarize results from OLS models that only include the treatment indicator. Grey spikes (diamond markers) summarize the results of models that include an interaction between the treatment indicator and the running variable measuring fieldwork days. Thick and thin lines are 90% and 95% confidence intervals, respectively. The models are estimated with entropy balancing weights and region fixed-effects.

References in Supporting Information

- Ares, Macarena and Enrique Hernández. 2017. "The corrosive effect of corruption on trust in politicians: Evidence from a natural experiment." *Research & Politics* 4(2):1–8.
- Balcells, Laia and Gerard Torrats-Espinosa. 2018. "Using a natural experiment to estimate the electoral consequences of terrorist attacks." *Proceedings of the National Academy of Sciences* 115(42):10624–10629.
- Berger, Eva M. 2010. "The Chernobyl disaster, concern about the environment, and life satisfaction." *Kyklos* 63(1):1–8.
- Bisgaard, Martin and Rune Slothuus. 2018. "Partisan Elites as Culprits? How Party Cues Shape Partisan Perceptual Gaps." *American Journal of Political Science* 62(2):456–469.
- Bonilla, Tabitha and Justin Grimmer. 2013. "Elevated threat levels and decreased expectations: How democracy handles terrorist threats." *Poetics* 41(6):650–669.
- Boomgaarden, Hajo G. and Claes H. de Vreese. 2007. "Dramatic Real-world Events and Public Opinion Dynamics: Media Coverage and its Impact on Public Reactions to an Assassination." *International Journal of Public Opinion Research* 19(3):354–366.
- Boydston, Amber E., Jessica T. Feezell and Rebecca A. Glazier. 2018. "In the wake of a terrorist attack, do Americans' attitudes toward Muslims decline?" *Research & Politics* 5(4):1–7.
- Bozzoli, Carlos and Cathérine Müller. 2011. "Perceptions and attitudes following a terrorist shock: Evidence from the UK." *European Journal of Political Economy* 27(S1):S89–S106.
- Branton, Regina, Valerie Martinez-Ebers, Tony E. Carey and Tetsuya Matsubayashi. 2015. "Social Protest and Policy Attitudes: The Case of the 2006 Immigrant Rallies." *American Journal of Political Science* 59(2):390–402.
- Brouard, Sylvain, Pavlos Vasilopoulos and Martial Foucault. 2018. "How terrorism affects political attitudes: France in the aftermath of the 2015–2016 attacks." *West European Politics* 41(5):1073–1099.
- Burlacu, Diana, Ellen M. Immergut, Maria Oskarson and Björn Rönnerstrand. 2018. "The politics of credit claiming: Rights and recognition in health policy feedback." *Social Policy & Administration* 52(4):880–894.
- Carey Jr, Tony E., Regina P. Branton and Valerie Martinez-Ebers. 2014. "The influence of social protests on issue salience among latinos." *Political Research Quarterly* 67(3):615–627.
- Coupe, Tom. 2017. "The impact of terrorism on expectations, trust and happiness – the case of the November 13 attacks in Paris, France." *Applied Economics Letters* 24(15):1084–1087.
- Das, Enny, Brad J. Bushman, Marieke D. Bezemer, Peter Kerkhof and Ivar E. Vermeulen. 2009. "How terrorism news reports increase prejudice against outgroups: A terror management account." *Journal of Experimental Social Psychology* 45(3):453–459.
- De Vries, Catherine E. 2018. *Euro-scepticism and the future of European integration*. New York: Oxford University Press.
- Dinas, Elias, Erin Hartman and Joost van Spanje. 2016. "Dead man walking: The affective roots of issue proximity between voters and parties." *Political Behavior* 38(3):659–687.

- Dinesen, Peter Thisted and Mads Meier Jæger. 2013. "The effect of terror on institutional trust: New evidence from the 3/11 Madrid terrorist attack." *Political Psychology* 34(6):917–926.
- Ferneer, Henk, Ineke Stoop and Eric Harrison. 2014. "Coding media claims in the European Social Survey, Round 7." Manuscript, https://www.europeansocialsurvey.org/docs/round7/methods/ESS7_media_claims_guidelines_ed_01_0.pdf.
- Finseraas, Henning, Niklas Jakobsson and Andreas Kotsadam. 2011. "Did the murder of Theo van Gogh change Europeans' immigration policy preferences?" *Kyklos* 64(3):396–409.
- Finseraas, Henning and Ola Listhaug. 2013. "It can happen here: the impact of the Mumbai terror attacks on public opinion in Western Europe." *Public Choice* 156(1-2):213–228.
- Flores, René D. 2018. "Can Elites Shape Public Attitudes Toward Immigrants?: Evidence from the 2016 US Presidential Election." *Social Forces* 96(4):1649–1690.
- Geys, Benny and Salmai Qari. 2017. "Will you still trust me tomorrow? The causal effect of terrorism on social trust." *Public Choice* 96(3-4):289–305.
- Hainmueller, Jens and Yiqing Xu. 2013. "Ebalance: A Stata package for entropy balancing." *Journal of Statistical Software* 54(7):1–18.
- Hariri, Jacob Gerner, Christian Bjørnskov and Mogens K. Justesen. 2015. "Economic shocks and subjective well-being: Evidence from a quasi-experiment." *The World Bank Economic Review* 30(1):55–77.
- Himmelweit, Hilde T, Marianne Jaeger Biberian and Janet Stockdale. 1978. "Memory for past vote: implications of a study of bias in recall." *British Journal of Political Science* 8(3):365–375.
- Hofstetter, C. Richard. 1969. "Political Disengagement and the Death of Martin Luther King." *The Public Opinion Quarterly* 33(2):174–179.
- Holbrook, Thomas M., Clayton Clouse and Aaron C. Weinschenk. 2012. "Bringing the president back in: The collapse of Lehman Brothers and the evolution of retrospective voting in the 2008 presidential election." *Political Research Quarterly* 65(2):263–274.
- Jakobsson, Niklas and Svein Blom. 2014. "Did the 2011 Terror Attacks in Norway Change Citizens' Attitudes Toward Immigrants?" *International Journal of Public Opinion Research* 26(4):475–486.
- Jensen, Carsten and Elias Naumann. 2016. "Increasing pressures and support for public health-care in Europe." *Health Policy* 120(6):698–705.
- Kim, Jin Woo and Eunji Kim. 2019. "Identifying the effect of political rumor diffusion using variations in survey timing." *Quarterly Journal of Political Science* forthcoming.
- Krosnick, Jon A and Donald R Kinder. 1990. "Altering the foundations of support for the president through priming." *American Political Science Review* 84(2):497–512.
- Larsen, Erik Gahner. 2018. "Welfare Retrenchments and Government Support: Evidence from a Natural Experiment." *European Sociological Review* 34(1):40–51.
- Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118(5):1199–1245.

- Lever, H. 1969. "The Johannesburg Station Explosion and Ethnic Attitudes." *The Public Opinion Quarterly* 33(2):180–189.
- Metcalf, Robert, Nattavudh Powdthavee and Paul Dolan. 2011. "Destruction and distress: Using a quasi-experiment to show the effects of the September 11 attacks on mental well-being in the United Kingdom." *The Economic Journal* 121(550):F81–F103.
- Minkus, Lara, Emanuel Deutschmann and Jan Delhey. 2018. "A Trump effect on the EU's popularity? The US presidential election as a natural experiment." *Perspectives on Politics* forthcoming:1–18.
- Moran, Matthew. 2017. "Terrorism and the *banlieues* : the *Charlie Hebdo* attacks in context." *Modern & Contemporary France* 25(3):315–332.
- Motta, Matthew. 2018. "The Polarizing Effect of the March for Science on Attitudes toward Scientists." *PS: Political Science and Politics* 51(4):782–788.
- Muñoz, Jordi and Eva Anduiza. 2019. "'If a fight starts, watch the crowd' The effect of violence on popular support for social movements." *Journal of Peace Research* forthcoming.
- Nagoshi, Julie L., Heather K. Terrell and Craig T. Nagoshi. 2007. "Changes in authoritarianism and coping in college students immediately after the terrorist attacks of September 11, 2001." *Personality and Individual Differences* 43(7):1722–1732.
- Ojeda, Christopher. 2016. "The effect of 9/11 on the heritability of political trust." *Political psychology* 37(1):73–88.
- Perrin, Andrew J. and Sondra J. Smolek. 2009. "Who trusts? Race, gender, and the September 11 rally effect among young adults." *Social Science Research* 38(1):134–145.
- Pierce, Lamar, Todd Rogers and Jason A. Snyder. 2016. "Losing Hurts: The Happiness Impact of Partisan Electoral Loss." *Journal of Experimental Political Science* 3(1):44–59.
- Reeves, Aaron and Robert de Vries. 2016. "Does media coverage influence public attitudes towards welfare recipients? The impact of the 2011 English riots." *The British Journal of Sociology* 67(2):281–306.
- Schaffner, Brian F. and Cameron Roche. 2017. "Misinformation and motivated reasoning: Responses to economic news in a politicized environment." *Public Opinion Quarterly* 81(1):86–110.
- Schulz, Peter J., Kent Nakamoto, Uwe Hartung and Carmen Faustinelli. 2008. "The death of Rosmarie Voser: The not-so-harmful consequences of a fatal medical error." *International Journal of Public Opinion Research* 20(3):386–397.
- Silber Mohamed, Heather. 2013. "Can protests make Latinos 'American'? Identity, immigration politics, and the 2006 marches." *American Politics Research* 41(2):298–327.
- Silva, Bruno Castanho. 2018. "The Non-Impact of the 2015 Paris Terrorist Attacks on Political Attitudes." *Personality and Social Psychology Bulletin* 44(6):838–850.
- Slothuus, Rune. 2016. "Assessing the Influence of Political Parties on Public Opinion: The Challenge from Pretreatment Effects." *Political Communication* 33(2):302–327.
- Solaz, Hector, Catherine E. De Vries and Roosmarijn A. de Geus. 2018. "In-Group Loyalty and the Punishment of Corruption." *Comparative Political Studies* OnlineFirst:1–31.

- Van Elsas, Erika J, Rozemarijn Lubbe, Tom WG van der Meer and Wouter van der Brug. 2013. "Vote recall: A panel study on the mechanisms that explain vote recall inconsistency." *International Journal of Public Opinion Research* 26(1):18–40.
- Willsher, Kim. 2015. "France boosts anti-terror measures in wake of Paris attacks." The Guardian <http://www.theguardian.com/world/2015/jan/21/france-anti-terror-measures-paris-attacks-manuel-valls>.
- Zaller, John and Mark Hunt. 1995. "The rise and fall of candidate Perot: The outsider versus the political system—part ii." *Political Communication* 12(1):97–123.
- Zepeda-Millán, Chris and Sophia J Wallace. 2013. "Racialization in times of contention: How social movements influence Latino racial identity." *Politics, Groups, and Identities* 1(4):510–527.