Protest policing and public perceptions of police. Evidence from a natural experiment in Germany.

Christof Nägel¹

Amy Nivette¹

¹Department of Sociology, Utrecht University, The Netherlands

Corresponding author:

Christof Nägel

Email: c.r.e.nagel@uu.nl

Address: Padualaan 14, 3584 CH Utrecht, The Netherlands

INTRODUCTION

Does excessively violent protest policing change public attitudes toward the police? Understanding when, to what extent, and in which contexts high-profile police actions influence attitudes towards the police is important because police perceptions are associated with public cooperation (Tyler & Huo, 2002), compliance with the law (Tyler & Jackson, 2014), criminal behaviour (Dawson, 2018; Eisner & Nivette, 2013) and, eventually, trust in the state itself (Jeong & Han, 2020). The current study assesses changes in public attitudes toward the police after the so-called ‘Black Thursday’ in Germany; the violent police dispersal of a protest movement against a new railway station project in Stuttgart (known as Stuttgart 21) on September 30, 2010. We take advantage of the overlap between the fieldwork period of the European Social Survey [ESS] 2010 (September 15, 2010-February 3, 2011) and the protests. This special type of natural experiment, known as an ‘unexpected event during survey design’ [UESD] (Muñoz et al., 2019), allows us to estimate causal effects of incidents of police violence on public opinion.

This study aims to contribute to knowledge on protest policing and public attitudes in two ways. First, we examine the impact of ‘vicarious experiences’ (Weitzer, 2017) on attitudes using a case study of police violence in the context of mass protests in Stuttgart, Germany. Given that the vast majority of research on perceptions of police, and high-profile police incidents in general, has been conducted in the United States (Jefferis et al., 1997; Kochel, 2019; Kochel & Skogan, 2021; Lasley, 1994; Reny & Newman, 2021; Tuch & Weitzer, 1997; Weitzer, 2002; White et al., 2018), but see (Curtice, 2021; Hohl et al., 2012; Kääriäinen et al., 2016; Nägel & Lutter, 2021; Thomassen et al., 2014), we lack a more general understanding of change in public attitudes towards the police in other national and institutional contexts (Roché & Oberwittler, 2017; Sahin et al., 2017; Sato, 2017). Compared to the US and

other European countries, trust in police in Germany is generally high (Guzy & Hirtenlehner, 2015). Previous research suggests that countries with a high average level of trust in the police are less likely to experience a severe legitimacy crisis even after a high-profile cases of police misconduct (Kääriäinen et al., 2016; Thomassen et al., 2014). The incident in Stuttgart therefore provides an important opportunity to evaluate changes in public attitudes towards the police in a ‘high-trust’ context.

Second, this study overcomes previous methodological limitations to estimate causal effects of incidents on public attitudes. Previous research on the effects of high-profile policing incidents on attitudes toward the police is either limited by (1) a lack of causal inference perspectives (Graziano, 2018), (2) a lack of nationally representative survey data (Hohl et al., 2012; Kaminski & Jefferis, 1998; Kochel, 2019; Kochel & Skogan, 2021; Lasley, 1994; White et al., 2018) and/or (3) a lack of appropriate measuring instruments (Nägel & Lutter, 2021). These issues call into question the internal, external and construct validity of previous findings. To our knowledge, the only papers to address all three issues are those by Curtice (2021) assessing violent police repression in Uganda and Reny and Newman (2021) assessing the impact of the killing of George Floyd and subsequent protests on public attitudes towards the police. The present study utilises a unique quasi-experimental design, high quality survey data, and extensive measurement instruments on perceptions of police in order to address these most common limitations. In addition, previous research suggests that opinion-mobilizing events tend to affect public opinion only briefly (Ares. & Hernández, 2017; Zaller, 1992), including attitudes towards the police (Nägel & Lutter, 2021; Reny & Newman, 2021). The fact that the focal event occurred early in the fieldwork period means that we are also able to investigate the evolution of potential effects over time.

HIGH-PROFILE INCIDENTS AND ATTITUDES TOWARD THE POLICE

Historically, event-based studies on perceptions of police first gained traction after the Rodney King beating in Los Angeles in March 1991. The incident not only sparked an extensive discussion about police violence against members of ethnic minorities, but also motivated researchers to assess the effect of police violence on public opinion (Graziano et al., 2009; Jefferis et al., 1997; Kaminski & Jefferis, 1998; Lasley, 1994; Weitzer, 1999, 2002; Weitzer & Tuch, 2004). Interest in these ‘indirect effects’ of police misconduct and violence, or so-called ‘vicarious experiences’ (Weitzer, 2017), was renewed following the death of Michael Brown in August 2014 (Kochel, 2019). Most recently, the killing of George Floyd in May 2020 further fuelled the debate both from a societal and academic point of view (Reny & Newman, 2021). The underlying question was whether and to what extent these incidents resulted in lasting effects on public attitudes towards the police. Prior research suggests that attitudes towards the police depend on the quality of treatment during interactions, whereby unfair, disrespectful, and unequal treatment can lead to more negative perceptions of police legitimacy (Mazerolle et al., 2013; Tyler et al., 2014; Walters & Bolger, 2019). Perceptions of police legitimacy concern ‘whether a power-holder is justified in claiming the right to hold power over other citizens’ (Bottoms & Tankebe, 2012). However, some events may result in little change in public opinion as they are filtered differently through (social) media outlets, resulting in varying degrees of public awareness and reactions (Cheng, 2021). In addition, incidents can be closely intertwined with protests, which can amplify public awareness, contestation, and ultimately polarization (Reny & Newman, 2021).

The results from these American-based studies told a largely similar story: in general, incidents of police violence led to significant but short-lived decreases in trust in police. However, the results were not unambiguous. For example, one study found little evidence that the death of

Freddie Gray was associated with more negative attitudes toward police in Baltimore (White et al., 2018). Likewise, studies on police misconduct and public opinion outside of the United States tend to show more heterogeneous effects. On the one hand, a study on the impact of the 2011 London disorder showed little to no effect on attitudes towards the police (Hohl et al., 2012), whereas analyses of police misconduct in Finland (Kääriäinen et al., 2016) and police failures during a terrorist attack in Norway (Thomassen et al., 2014) suggest that trust may even increase following certain controversial incidents. On the other hand, Nägel & Lutter (2021) found that trust in police significantly declined following a serious incident of police violence coupled with protests in France. Similar to studies in the US, the authors found that the negative effect was relatively short-lived for the general population, but more persistent for ethnic minorities.

PROTEST POLICING AND ATTITUDES TOWARD THE POLICE

Several incidents coincide with large-scale protests, where police are often under close public and media scrutiny (Greer & McLaughlin, 2010; Maguire & Oakley, 2020; Rosie & Gorringer, 2009). Protest policing refers to police treatment and control of protesters (Della Porta, 1999) which may involve a range of peaceful (e.g. communication, differentiation) or more aggressive (e.g. arrest, assault, tear gas, water cannons) tactics (Maguire & Oakley, 2020). While the public tends to be more supportive of peaceful policing tactics, recent research suggests that public perceptions of protest policing depends on situational characteristics and the goals of the protest (e.g. Black Lives Matter, pro-confederate monuments, see Metcalfe & Pickett, 2021). Specifically, in situations where protestors were armed, committed violence

and property damage, and interrupted traffic, respondents were more likely to fear protestors and support aggressive policing tactics.

Public perceptions of protest policing also depend on police conduct and treatment during interactions with protestors (Au and Hui 2014; Curtice, 2021; Newburn et al., 2018; Perry et al., 2017). Analysing the effects of protest policing on perceptions of police legitimacy in Hong Kong, Au and Hui proposed three mechanisms by which protest policing might influence attitudes: (1) degree of fairness and expediency in dealing with public processions; (2) the interaction between the police and protesters on site and the responsiveness in accommodating demands by protesters; and the (3) degree of neutrality displayed towards groups with different political views. (2014: 225). Generally, these mechanisms coincide with notions of fairness, voice, and neutrality or impartiality, which comprise key elements of procedural justice policing (Mazerolle et al., 2013; Trinkner & Tyler, 2016; Tyler & Blader, 2003). The use of excessive force or repressive tactics against protestors violates these principles and can damage the trustworthiness of police (Curtice & Behlendorf, 2021; Nägel & Lutter, 2021; Perry et al., 2017). Research that has used disaggregated measures of perceptions of police and legitimacy has found that the use of excessive force and violence during protests is associated with declines in perceptions of police fairness, feelings of obligation to obey, and normative alignment with police (Curtice, 2021; Metcalfe & Pickett, 2021).

In addition, the use of excessive force during protests can lead to political backlash among observers (Curtice & Behlendorf, 2021), as police are often considered representatives of the state there to uphold the status quo (Channing, 2021; Earl et al., 2003; Maguire & Oakley, 2020; Wang et al., 2020; Wood, 2020). In interviews conducted with participants of the 2012 'Occupy' protests in Israel, one protester stated that 'when the PPU's [paramilitary policing

units] are involved in policing the protest event, they give you this feeling – we do not work for you. You are the enemy [and] we have to keep you quiet’ (quoted in Perry et al., 2017: 620). Police presence at protests, and in particular the use of force against protesters, can therefore damage perceptions of impartiality alongside trustworthiness.

Taken together, previous research suggests that high-profile incidents of police violence during protests should result in more negative perceptions of police, and that these effects are typically short-lived. More specifically, we expect that the use of violence will lead to lower perceptions of police fairness, moral alignment, obligation to obey, and political impartiality (Research Question 1). In addition, we expect that any changes in attitudes towards the police will return to pre-event levels in the long-term (Research Question 2). This study examines these questions using a case study of excessive police violence during protests in Stuttgart, Germany.

Contextual setting

In order to understand the effects of the so called ‘Black Thursday’ it is necessary to explain the broader political context of the planned railway station in Stuttgart and the motivation for its opposition. The broader aim of the infrastructure project is to make changes to the Stuttgart railroad station. Among other things, Stuttgart's main station is to be converted from an overground terminus to an underground transit station. Since planning for Stuttgart 21 began, there have been protests against the project. Major criticisms concern for example the high costs, low economic viability, a lack of democratic legitimation and citizen participation, the difficulty in access for travellers, endangerment of the environment (through possible water contamination), the lack of protection of historical monuments, and general deficiencies in planning.

Although the planning of the project began in 1996, it was not until the police operation in the Stuttgart Schlossgarten on September 30, 2010, that the project came to the attention of the broader German public. On the morning of September 30, 2010, many protesters occupied the Schlossgarten in Stuttgart. Around 600 police officers acted with unusual force against allegedly over 1000 demonstrators who represented a broad cross-section of society (including for example pensioners and school children). According to information from citizens' initiatives and park protectors, more than 100 people were injured, including minors. Four demonstrators had serious eye injuries, and according to the protest organisers, more than 400 protesters suffered eye irritation. A 66-year-old man almost completely lost his eyesight as a consequence of being hit head-on in the eyes with a water cannon.¹ Following the events, the police justified their actions by claiming that the aggression had originated from the demonstrators. However, this account was refuted by police videos which included details on the timing of events. Later that day, work began on felling the first trees in the park to make way for the new station.

September 30, 2010 was hence remembered as 'Black Thursday.'² On November 20, 2010, about 10,000 people protested to demand clarification regarding the background and legal reappraisal of the police operation. Every year around September 30th, rallies, demonstrations, and similar events commemorate those affected by the police action. Thousands of people gather for these events.³

¹ See https://en.wikipedia.org/wiki/Dietrich_Wagner

² A major newspaper from Stuttgart provides a slide show of the events on the 'Black Thursday':
<https://www.stuttgarter-zeitung.de/gallery.chronologie-zu-stuttgart-21-schwarzer-donnerstag-im-schlossgarten-param~1~0~0~13~false.114e1e4e-5f61-4523-8f8c-9a387733c0ed.html>.

³ A more detailed description of the events - but only in German - can be found on the Wikipedia article

Figure 1. Evolution of weekly Google search trends before and after the event.

[https://de.wikipedia.org/wiki/Protest_gegen_Stuttgart_21#30. September 2010: R%C3%A4umung_des_Schlossgartens_%E2%80%93_%E2%80%93_Schwarzer_Donnerstag%E2%80%93](https://de.wikipedia.org/wiki/Protest_gegen_Stuttgart_21#30._September_2010:_R%C3%A4umung_des_Schlossgartens_%E2%80%93_%E2%80%93_Schwarzer_Donnerstag%E2%80%93)

A detailed description in English can also be found here:

<https://www.spiegel.de/international/germany/the-world-from-berlin-germany-shocked-by-disproportionate-police-action-in-stuttgart-a-720735.html>

Figure 1 complements the qualitative characterization of the political circumstances surrounding the event with a quantitative examination of related relative Google search trends for the year 2010 in Germany. Internet search data is considered a valuable resource to gain insights into public problem awareness (Muñoz et al., 2019). Google trends in particular are considered valid indicators of issue salience (Mellon, 2014) and have been analyzed alongside traditional survey data in many previous UESD studies (Ares & Hernández, 2017; Nägel & Lutter, 2021; Reny & Newman, 2021; Thompson, 2021). Google trends data can also serve as a proxy for political interest in the absence of survey data (Barrie, 2020).

The search trends demonstrate an interesting pattern: while the themes 'Polizei Stuttgart' ('police Stuttgart'), 'Stuttgart 21' and 'Wasserwerfer' (water cannon) peak in the week after the event, the more general search theme 'Protest' reaches the highest relative frequency in the week of the event. Except for 'water cannon', every search term shows some interest even in the weeks before the event. This provides an important clue for the present study, as it shows that the police were already associated with the infrastructure project before September 30, 2010, but the connection to excessive violence (water cannon) only appeared to be the focus of public interest after the event.

While interest in 'Protest' appears to be more persistent, interest in the other three search themes quickly decreases. This pattern suggests that even though police involvement was a very salient issue after September 30, 2010, the uptick of interest quickly faded. This trend is in line with the notion that the police are a 'Teflon Service' (Reiner, 2010) where even major scandals resolve quickly. This exploratory analysis furthermore highlights the importance of examining the existing survey data more closely with respect to interview timing.

Nevertheless, these Google trends demonstrate that ‘Black Thursday’ can be considered a significant event in the German public ‘psyche.’

METHODS⁴

Identification strategy

To answer the research questions, we make use of the UESD identification strategy, which is a quasi-experimental research design that can analyse shifts in public opinion after unexpected events that occur during the fieldwork of survey programs. Although this research design purports to permit causal inferences, these are subject to two key identification assumptions that need to be examined in detail: excludability and temporal ignorability (Muñoz et al., 2019). The first one assumes that the interview date, as an instrument Z , exclusively affects the outcome Y through the event X . The occurrence of other events influencing Y must therefore be excluded. The temporal ignorability assumption is concerned with the assignment to pre- or post-treatment groups, which should be as good as random. Muñoz et al. (2019) provide good practice advice for UESD that we adhere to as closely as possible in order to obtain unbiased estimates and efficient standard errors. Since these robustness checks are technical in nature, they are discussed in detail in the accompanying supplemental material. In general, the results are robust to these additional analyses.

⁴ All survey data can be downloaded from the ESS webpage. Google trends data can be assessed via the gtrendsR package in R or directly from the google trends webpage. All code to reproduce analysis and figures will be made publicly available in case of acceptance and linked in the paper.

Data, descriptive statistics and model specification

The European Social Survey [ESS] is a cross-national survey program that measures respondents' attitudes, beliefs and behaviour patterns in European countries on the basis of probability samples gathered through face-to-face interviews (European Social Survey, 2011). Detailed information concerning the ESS can be found in the supplementary materials (section A1.1).

The current study examines several operationalizations of attitudes towards the police that variously reflect public perceptions of fairness, moral alignment, obligation to obey, trust in police, and neutrality or impartiality. The variable *procedural fairness* was measured using three items that asked respondent how often police in their country 1) treat people with respect, 2) make fair, impartial decisions and 3) explain their decisions when asked ($\alpha = .60$). Responses were measured using a 4-point Likert-type scale ranging from 1 'not at all often' to 4 'very often'. *Moral alignment* was measured using three items that capture to what extent the police 1) 'have the same sense of right and wrong as me', 2) 'stand up for values that are important to people like me' and whether the respondent 3) 'generally support[s] how the police act' ($\alpha = .78$). Responses were measured on a 5-point Likert-type scale ranging from 1 'agree strongly' to 5 'disagree strongly'. The obligation to obey scale consists of 3 items measuring the respondent's felt duty to 1) 'back decisions made by the police, even if disagree', 2) 'do what police say, even when don't understand or agree' and 3) 'do what police say even if treated badly' ($\alpha = .86$). Responses are measured on a scale ranging from 0 'not at all my duty' to 10 'completely my duty'. Factor analyses confirm a one-factor solution for each construct (see Appendix A1.2, Table A1). The variable *trust in police* was measured using a single item that asks respondents to what extent they trust the following

institutions (the police) on a scale from 0 'no trust at all' to 10 'complete trust'. Finally, the measure of political impartiality was measured using a single item that asks to what extent the respondents agree or disagree with the statement that 'the decisions and actions of the police are unduly influenced by pressure from political parties and politicians'. Responses ranged from 1 'agree strongly' to 5 'disagree strongly'.

Five socio-demographic variables were included in order to assess and compare the composition of control and treatment groups: age (years), gender (1=female), education, income, and ethnic minority. Education was measured on a scale from 0-25, with 0 representing no education and 25 the completion of doctoral studies. Income was measured as household's total net income in deciles. We applied the natural logarithm to account for the skewed distribution. Ethnic minority captures whether the respondent is a member of an ethnic minority group (1=yes). Except for income, we only included observations with complete cases on all variables to keep the number of observations constant between different analyses. Results were virtually the same when we follow a listwise missing value strategy (see Tables A4 and A5 in the Supplementary Materials). Following previous research using the same design (Minkus et al., 2019), we avoided survey weights since they might further bias results in this particular design. Still, results were robust to applying the two possible ESS weights (design weights and post-stratification weights, see table A6 and A7 in the appendix).

Table 1. Descriptive Statistics

Control group						Treatment group						
Statistic	N	Mean	St. Dev.	Min	Max	Statistic	N	Mean	St. Dev.	Min	Max	Mean Δ
Political Impartiality	371	2.431	0.946	1	5	Political Impartiality	1,875	2.294	0.94	1	5	-0.137***
Trust in the Police	396	6.644	2.219	0	10	Trust in the Police	1,979	6.797	2.227	0	10	0.153
Procedural Fairness	350	2.938	0.432	1	4	Procedural Fairness	1,710	2.957	0.499	1	4.333	0.019
Moral Alignment	384	3.924	0.595	1.667	5	Moral Alignment	1,909	3.856	0.645	1	5	-0.068*
Obligation to Obey	386	6.631	2.293	0	10	Obligation to Obey	1,937	6.649	2.275	0	10	0.018
Age	396	52.571	18.127	15	90	Age	1,987	48.623	17.36	15	97	-3.948***
Female	396	0.48	0.5	0	1	Female	1,987	0.484	0.5	0	1	0.004
Education	396	12.571	3.494	2	23	Education	1,987	13.406	3.352	0	28	0.835***
Income	396	4.987	2.833	1	10	Income	1,987	5.582	2.711	1	10	0.595***
Ethnic Minority	396	0.975	0.157	0	1	Ethnic Minority	1,987	0.957	0.204	0	1	-0.018*

As can be seen in table 1, the data appear to be rather unbalanced between the treatment and control group: the treatment group was on average significantly younger, more educated, had a higher income, and was less likely to belong to an ethnic minority. As Muñoz et al. (2019) argue, controlling for these covariate imbalances can reduce bias. Furthermore, a narrower bandwidth is beneficial to achieve conditional independence since observations and their outcomes are more likely to be independent from treatment assignment. This, however, is a trade-off since a narrow bandwidth increases variance, allows exclusively for estimation of very local effects, and may even lead to type two errors since the treatment effect might need some time to evolve. The best strategy, therefore, was to choose a bandwidth with maximal balance between the groups. For this study, this meant a range of 15 days before and after the event. The only remaining imbalance was a younger ($M_c = 52.12$ years, $M_t = 48.54$; $p = .008$) post-treatment sample. To further reduce model dependence, we present findings from a ‘full bandwidth sample’ in the appendix (Supplementary Materials A1.3, Table A3). This specification again did not change our main results.

The +/- 15-day sample period can be used to answer RQ 1, namely that people interviewed after the event generally reported less positive attitudes towards the police. To address the second research question, we ran 5 x 18 regressions for each outcome by subsequently adding one week of observations to the post-treatment time-window (starting with a time window of one week). The control group stayed the same across all models (15 days before the event). This way, we could track the trajectories of potential treatment effects over time (see Ares & Hernández, 2017 for a similar approach).

RESULTS

Figure 2 provides a visual representation of this quasi-natural experiment, resembling a classic regression discontinuity (rd) plot.⁵ The dashed line represents the 15-day bandwidth of the ‘maximum balanced local sample’. The histogram at the bottom of Figure 2 represents the average number of interviews per week. Fluctuations in interview frequency could influence outcomes, but the results suggest that the number of interviews around the ‘cut-off’ were relatively stable and should provide enough power to run the analyses. The control group (before the event) consists of 424 (17.8%) respondents, whereas there are 1959 (82.2%) respondents in the treatment group. The outcome evolution is represented by two simple linear regression lines, one before and one after the event for each of the five outcomes, the predictor being the date of the interview. While there appears to be little level or trend change for the outcomes procedural fairness, moral alignment or obligation to obey, there are visible discontinuities after September 30 in the outcomes political impartiality and trust in the police. While the former drops right after the ‘cut-off’, the latter increases. It is important to note that there exist considerable differences between the pre- and post-event groups which might influence results. The naïve before-after-comparisons in Figure 2 may therefore give biased estimates of a causal effect.

⁵ We refrain from using a straightforward rd design because there are very few observations in the direct vicinity of the cut-point. Recent research suggests that rd may be an inappropriate method for inference when the post-intervention sample size is small (Stommes et al., 2021).

Figure 2. Outcome evolution after ‘Black Thursday’. Histogram represents average number of interviews per week for complete weeks. Black dotted line is October 1, 2010 (One day after the event). Grey dotted line is October 15, 2010 and represents the shortened time window.

To achieve a more reliable assessment of effects of the event on attitudinal outcomes, we first regressed each outcome on the treatment variable in a baseline model and then included the covariates presented in Table 1 as well as federal state fixed effects. Although this is not a panel data set, the fixed effects take into account that the same regions were surveyed before and after the event. This allows us to control for the effect of omitted variables that are constant over time but vary between states.⁶

⁶ This also controls for a possible east/west effect which should always be considered when using nationally representative data for Germany.

Figure 3. Effect of 'Black Thursday' on outcomes. Note: Grey models summarise results from regressions that only include the treatment indicator. Black models summarise the results of models adjusting for all covariates listed in Table 1 as well as federal state fixed effects. Thick and thin lines are 90% and 95% confidence intervals, respectively.

Figure 3 summarises our findings for RQ 1. Both the baseline model and the model with controls and fixed effects show a significant decrease in perceived political impartiality of the police after the event with $B = -.24$ and $p = .001$ [95 % CI = $-.390, -.097$] in the long regression. This effect is substantial as it suggests a change of more than $\frac{1}{4}$ of a standard deviation of the variable's pre-treatment distribution. The results thus suggest that respondents interviewed after the event were significantly more likely to perceive the police as being under unjust political pressure. Regressing the second outcome, trust in the police, on the treatment, the controls and state dummies revealed a treatment effect of $B = 0.299$ ($p = .084$), however the 95% confidence interval does include zero [95 % CI = $-.041, .641$]. This result is noteworthy since it opposes the relationship we hypothesised. Concerning the outcomes procedural fairness, moral alignment and obligation to obey, none of the long regressions showed any effect that is significantly different from zero. Moral alignment appears to be negatively affected in the baseline model ($B = -0.089$) but the effect was small, statistically significant only at the 10% level and mediated by inclusion of covariates and fixed effects. Hence, apart from the event's effect on political impartiality, none of our findings support the hypothesised relationships.

To address RQ2 we focus on the two outcomes that show some over-time variation which might be attributable to the event (political impartiality and trust in the police). For both outcomes, we did so by re-running the regressions on the treatment effect and subsequently adding one week of observations at a time starting with a model of 7 days post-treatment and ending with a model 18 weeks after the event. Diminishing effect sizes could be an indication that any changes in public perceptions of police will return to pre-event levels in the long-term, especially when standard errors do not inflate at the same time. This

specification also provides a convenient way of analysing whether findings could merely be driven by sample bandwidths. Results are summarised in Figure 4.

Figure 4. Regression models run on each weekly independent cross section for the outcomes ‘Political Impartiality’ and ‘Trust in the Police’.

While the treatment effect on political impartiality is not significant in week one after the event (which might be a result of very limited power), it is greatest two weeks after the event and sharply drops in the next two weeks while remaining significant at $p < .05$. This effect appears to stay relatively stable but significant for another seven weeks until it again decreases in the last weeks of the fieldwork period. As such, it appears as though the effect on political impartiality indeed diminishes over time, and that this decrease is non-linear.

The results on trust in the police are somewhat more difficult to interpret. While the figure clearly shows a decrease in the positive treatment effect, the treatment dummy falls in and out of significance even for the most liberal threshold of 10% in a seemingly random

pattern. This suggests that any changes in trust over time are apparently little more than noise.

Summarizing our results, it seems that the events of ‘Black Thursday’ only had a substantial and consistent (in terms of statistical significance) effect on perceptions of political impartiality. It is also the only treatment effect that shows a clear decreasing pattern over time, although it does not reach pre-event levels during the period we can observe in the survey. Trust in the police is, on average, somewhat elevated after the event and this effect also diminishes over time. However, this result is not robust to all model specifications. The analyses on the other outcomes clearly point in one direction: the event did not alter perceptions of procedural fairness, moral alignment or obligation to obey.

To increase the reliability of our estimates and our overall research design, we ran virtually all sensitivity analyses recommended in the literature (Muñoz et al., 2019), the results of which can be found in the Supplemental Material. In addition to general analytical choices like handling of missing values or survey choices, we addressed the assumptions related to UESD design. Specifically, we assessed placebo outcomes and pre-existing time trends, evaluated the UK as a placebo comparison, examined placebo-treatments at the empirical median of the control group, and analysed non-response patterns. Overall, these analyses indicate that our results are generally robust.

DISCUSSION

This study aimed to assess potential change in public attitudes toward the police after a major policing controversy during protests against a planned railway station in Stuttgart, Germany. We found little to no change in several dimensions of perceptions of police and legitimacy, specifically measures of moral alignment, procedural fairness, and obligation to obey the

police. However, respondents interviewed after the event saw the police as more unduly influenced by political pressure. This effect was robust to various placebo specifications, to the temporal bandwidth of the analysis, and sizable across all modelling strategies. Surprisingly, people in the post-intervention group reported more trust in the police than respondents in the control group, but this result was not robust to all alternative time windows and model specifications. In the following discussion, we discuss two main implications for research on high-profile events and public attitudes towards the police.

First, the results suggest that, by and large, the events of ‘Black Thursday’ had little impact on public perceptions of police in Germany. On the one hand, this contradicts previous research that has found significant changes in attitudes towards the police following high-profile incidents of police violence (Kochel, 2019; Reny & Newman, 2021; Weitzer, 2002). On the other hand, the findings are consistent with research on high-profile events in ‘high trust’ societies (e.g. Finland, Norway, and to a lesser extent UK; Hohl et al., 2012; Kääriäinen et al., 2016; Thomassen et al., 2014), which found that attitudes showed little or even positive change following controversial incidents. This heterogeneity in effects may be due to differences in institutional contexts or event characteristics. For example, in ‘high trust’ societies where the trust in police is also high, public reactions to police misconduct may be more tolerant or only short-term. In these contexts, incidents of police violence also tend to be less common compared to countries like the United States or Brazil (Ceccato et al., 2018; Lawson, 2019). By contrast, where incidents of police misconduct are situated in polarised contexts where the legitimacy of police is highly contested, high-profile incidents can have a lasting impact particularly among members of affected minority groups (Nägel & Lutter, 2021; Reny & Newman, 2021). However, in contexts where police violence is more frequent and

support for the police is already low, these incidents may have relatively smaller or null effects (see e.g. White et al., 2019).

Second, we found that the event changed public opinion about police impartiality, whereby respondents reported that the police were more likely to be unduly influenced by politics. This is in line with previous research that shows that police are often seen as tools of the state when present at protests (Curtice & Behlendorf, 2021; Newburn et al., 2018). In particular, more aggressive actions taken during protests, such as the use of force and paramilitary-style policing, can lead to feelings of alienation among protesters (Perry et al., 2017). From a methodological standpoint, the significant effect on political impartiality could also be considered a ‘manipulation check’ (see Muñoz et al., 2019), to test whether respondents were exposed to the event in the sense of a natural experiment. Such an interpretation would underline reliability of null- or even contrary effects on the other outcomes.

Interestingly, while some argue that perceived political impartiality is a key element of police legitimacy (Tankebe, 2013), the current findings show that changes in neutrality did not ‘spill over’ to influence changes in correlated constructs, such as procedural fairness. This suggests that the incident was perceived as a political rather than police crisis, and that the public likely makes some distinction between police actions in different contexts (i.e. protests vs on the street) and how these actions reflect on the police more generally. These effects may also depend on the extent and content of media coverage, which can play a role in shaping narratives about police misconduct and violence (Cheng, 2021). More research is needed to understand how police actions during protests are differentially filtered through (social) media outlets, and to what extent this influences broader political and legal attitudes.

Taken together, the impact of incidents of police violence on public attitudes may be moderated variously by media attention, the type of event, and/or the societal and institutional context in which police are embedded. For example, high-profile incidents of police violence may have an inverse-U relationship with trust in police, whereby effects are smaller in societies with very low or high trust in police. These processes are important to dissect in order to understand how and under what circumstances the public responds to such incidents, and what strategies police must take to effectively repair public opinion and mitigate losses of legitimacy over time (e.g. reconciliation, apologies, structural reform see McLean, 2021; O'Brien et al., 2019). Future research should move beyond single-incident case studies to comparatively assess how and when high-profile incidents lead to changes in attitudes towards the police, accounting for prior societal characteristics as well as post-incident organizational and media responses.

Limitations and conclusions

While there are several strengths to this study, notably the utilization of a quasi-experimental design and multiple dimensions of perceptions of police, there are also a number of limitations to address. First, natural experiments are often praised for 'maximizing' external validity, but the current study is nevertheless a single case study concerning a particular event in time on attitudes in Germany. More comparative research utilizing natural experiments is needed in order to assess and establish general mechanisms that contribute to changes in attitudes. In addition, controlled experiments and qualitative research can complement existing research to describe the evolution of media coverage and police reactions and reconciliation following different events (see Cheng, 2021).

Other limitations include those inherent to natural experiments in general and UESD in particular. Although we discuss and test the excludability and temporal stability assumption in close detail in the supplementary materials, it is still possible that our results could be driven or levelled by unobserved heterogeneity. Finally, as has been discussed in the UESD literature (Muñoz et al., 2019), there is an obvious one-sided compliance problem in UESD studies, since it is more difficult in this particular setting to be certain that respondents have effectively been exposed to the treatment, thus the effect should be treated as an Intent-to-treat, rather than an Average-treatment-effect. The interpretation of the political impartiality item as a pseudo-manipulation check can lend more confidence that respondents were aware of the events of the 'Black Thursday'.

Overall, the current paper demonstrates that public perceptions of police can be responsive to police actions, however the results likely depend on societal and situational contexts. It is important to note that, although opinion-mobilizing effects might be short-lived (Nägel & Lutter, 2021; Reny & Newman, 2021), high-profile policing incidents can threaten the validity of measuring instruments. Researchers should pay close attention to fieldwork dates and overlap with potentially highly salient events that can influence attitudinal measures. In order to avoid bias stemming from these sources, it may therefore be advisable to choose short fieldwork periods or to examine the survey periods in close detail for possible confounding events.

REFERENCES

- Ares, M., & Hernández, E. (2017). The corrosive effect of corruption on trust in politicians: Evidence from a natural experiment. *Research & Politics*, 4(2), 205316801771418. <https://doi.org/10.1177/2053168017714185>
- Barrie, C. (2020). Searching Racism after George Floyd. *Socius: Sociological Research for a Dynamic World*, 6, 237802312097150. <https://doi.org/10.1177/2378023120971507>
- Bottoms, A., & Tankebe, J. (2012). BEYOND PROCEDURAL JUSTICE: A DIALOGIC APPROACH TO LEGITIMACY IN CRIMINAL JUSTICE. *The Journal of Criminal Law and Criminology*, 102(1), 119–170.
- Ceccato, V., Melo, S. N., & Kahn, T. (2018). Trends and Patterns of Police-Related Deaths in Brazil. In K. Carrington, R. Hogg, J. Scott, & M. Sozzo (Eds.), *The Palgrave Handbook of Criminology and the Global South* (pp. 521–550). Springer International Publishing. https://doi.org/10.1007/978-3-319-65021-0_26
- Channing, I. (2021). Policing Extreme Political Protest: A Historical Evaluation of Police Prejudice. *Policing: A Journal of Policy and Practice*, 14(4), 900–915. <https://doi.org/10.1093/police/pay010>
- Cheng, T. (2021). Social media, socialization, and pursuing legitimation of police violence*. *Criminology*, 1745-9125.12277. <https://doi.org/10.1111/1745-9125.12277>
- Curtice, T. B. (2021). How Repression Affects Public Perceptions of Police: Evidence from a Natural Experiment in Uganda. *Journal of Conflict Resolution*.
- Curtice, T. B., & Behlendorf, B. (2021). Street-level Repression: Protest, Policing, and Dissent in Uganda. *Journal of Conflict Resolution*, 65(1), 166–194. <https://doi.org/10.1177/0022002720939304>

- Dawson, A. (2018). Police Legitimacy and Homicide: A Macro-Comparative Analysis. *Social Forces. Advance Online Publication*. <https://doi.org/10.1093/sf/soy043>
- Earl, J., Soule, S. A., & McCarthy, J. D. (2003). Protest under Fire? Explaining the Policing of Protest. *American Sociological Review*, 68(4), 581. <https://doi.org/10.2307/1519740>
- Eisner, M., & Nivette, A. (2013). Does Low Legitimacy Cause Crime? A Review Of The Evidence. In J. Tankebe & A. Liebling (Eds.), *Legitimacy and Criminal Justice: An International Exploration* (pp. 308–325). OUP Oxford.
<https://doi.org/10.1093/acprof:oso/9780198701996.003.0016>
- European Social Survey. (2011). *European Social Survey (ESS), Round 5—2010*.
<https://doi.org/10.21338/NSD-ESS5-2010>
- Graziano, L. M. (2018). News media and perceptions of police: A state-of-the-art-review. *Policing: An International Journal*, 10(2), 109. <https://doi.org/10.1108/PIJPSM-11-2017-0134>
- Greer, C., & McLaughlin, E. (2010). We Predict a Riot?: Public Order Policing, New Media Environments and the Rise of the Citizen Journalist. *British Journal of Criminology*, 50(6), 1041–1059. <https://doi.org/10.1093/bjc/azq039>
- Guzy, N., & Hirtenlehner, H. (2015). Trust in the German Police: Determinants and Consequences for Reporting Behavior. In G. Meško & J. Tankebe (Eds.), *Trust and Legitimacy in Criminal Justice* (pp. 203–229). Springer International Publishing.
https://doi.org/10.1007/978-3-319-09813-5_10
- Hohl, K., Stanko, B., & Newburn, T. (2012). The Effect of the 2011 London Disorder on Public Opinion of Police and Attitudes Towards Crime, Disorder, and Sentencing. *Policing: A Journal of Policy and Practice*, 7(1), 12–20. <https://doi.org/10.1093/police/pas055>

- Jefferis, E. S., Kaminski, R. J., Holmes, S., & Hanley, D. E. (1997). The effect of a videotaped arrest on public perceptions of police use of force". *Journal of Criminal Justice*, 25, 381–389.
- Jeong, J., & Han, S. (2020). Trust in police as an influencing factor on trust in government: 2SLS analysis using perception of safety. *Policing: An International Journal of Police Strategies & Management*, 43(4), 541–558. <https://doi.org/10.1108/PIJPSM-11-2019-0172>
- Kääriäinen, J., Isotalus, P., & Thomassen, G. (2016). Does public criticism Erode trust in the police? The case of Jari Aarnio in the Finnish news media and its effects on the public's attitudes towards the police. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 17(1), 70–85.
<https://doi.org/10.1080/14043858.2016.1144315>
- Kaminski, R. J., & Jefferis, E. S. (1998). The effect of a violent televised arrest on public perceptions of the police. *Policing: An International Journal of Police Strategies & Management*, 21(4), 683–706. <https://doi.org/10.1108/13639519810241692>
- Kochel, T. R. (2019). Explaining Racial Differences in Ferguson's Impact on Local Residents' Trust and Perceived Legitimacy: Policy Implications for Police. *Criminal Justice Policy Review*, 30(3), 374–405. <https://doi.org/10.1177/0887403416684923>
- Kochel, T. R., & Skogan, W. G. (2021). Accountability and transparency as levers to promote public trust and police legitimacy: Findings from a natural experiment. *Policing: An International Journal*, ahead-of-print(ahead-of-print).
<https://doi.org/10.1108/PIJPSM-04-2021-0062>

- Lasley, J. R. (1994). The impact of the Rodney king incident on citizen attitudes toward police. *Policing and Society*, 3(4), 245–255.
<https://doi.org/10.1080/10439463.1994.9964673>
- Lawson, E. (2019). *TRENDS: Police Militarization and the Use of Lethal Force. Political Research Quarterly*, 72(1), 177–189. <https://doi.org/10.1177/1065912918784209>
- Maguire, E. R., & Oakley, M. (2020). Policing Protests. Lessons from the Occupy Movement, Ferguson & Beyond: A Guide for Police. *Harry Frank Guggenheim Foundation*, 84.
- Mazerolle, L., Bennett, S., Davis, J., Sargeant, E., & Manning, M. (2013). Procedural justice and police legitimacy: A systematic review of the research evidence. *Journal of Experimental Criminology*, 9(3), 245–274. <https://doi.org/10.1007/s11292-013-9175-2>
- McLean, K. (2021). Justice-restoring responses: A theoretical framework for understanding citizen complaints against the police. *Policing and Society*, 31(2), 209–228.
<https://doi.org/10.1080/10439463.2019.1704755>
- Mellon, J. (2014). Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience. *Journal of Elections, Public Opinion and Parties*, 24(1), 45–72. <https://doi.org/10.1080/17457289.2013.846346>
- Metcalfe, C., & Pickett, J. T. (2021). Public fear of protesters and support for protest policing: An experimental test of two theoretical models. *Criminology*, 30.
<https://doi.org/10.1111/1745-9125.12291>
- Minkus, L., Deutschmann, E., & Delhey, J. (2019). A Trump Effect on the EU's Popularity? The U.S. Presidential Election as a Natural Experiment. *Perspectives on Politics*, 17(02), 399–416. <https://doi.org/10.1017/S1537592718003262>

- Muñoz, J., Falcó-Gimeno, A., & Hernández, E. (2019). Unexpected Event during Survey Design: Promise and Pitfalls for Causal Inference. *Political Analysis, OnlineFirst*, 1–21.
<https://doi.org/10.1017/pan.2019.27>
- Nägel, C., & Lutter, M. (2021). The 2017 French riots and trust in the police: A quasi-experimental approach. *European Journal of Criminology*, 147737082199897.
<https://doi.org/10.1177/1477370821998974>
- Newburn, T., Diski, R., Cooper, K., Deacon, R., Burch, A., & Grant, M. (2018). ‘The biggest gang’? Police and people in the 2011 England riots. *Policing and Society*, 28(2), 205–222. <https://doi.org/10.1080/10439463.2016.1165220>
- O’Brien, T. C., Meares, T. L., & Tyler, T. R. (2019). Reconciling Police and Communities with Apologies, Acknowledgements, or Both: A Controlled Experiment. *Annals of the American Academy of Political and Social Science*, 687, 202–215.
- Perry, G., Jonathan-Zamir, T., & Weisburd, D. (2017). The Effect of Paramilitary Protest Policing on Protestors’ Trust in the Police: The Case of the “Occupy Israel” Movement: Effects of Paramilitary Policing. *Law & Society Review*, 51(3), 602–634.
<https://doi.org/10.1111/lasr.12279>
- Porta, D. della. (1999). Protest, Protesters, and Protest Policing: In S. Tarrow, M. Giugni, D. McAdam, & C. Tilly (Eds.), *How Social Movements Matter* (NED-New edition, Vol. 10, pp. 66–96). University of Minnesota Press; JSTOR.
<http://www.jstor.org/stable/10.5749/j.cttt706.9>
- Reiner, R. (2010). *The politics of the police* (4.). Oxford Univ. Press.
- Reny, T. T., & Newman, B. J. (2021). The Opinion Mobilizing Effect of Social Protest Against Police Violence: Evidence from the 2020 George Floyd Protests. *American Political*

Science Review.

https://www.dropbox.com/s/189v1jq3iltqlqe/reny_newman_apsr_sharing.pdf?dl=0

Roché, S., & Oberwittler, D. (2017). Towards a broader view of police-citizen relations: How societal cleavages and political contexts shape trust and distrust, legitimacy and illegitimacy. In D. Oberwittler & S. Roché (Eds.), *Routledge Frontiers of Criminal Justice. Police-Citizen Relations Across the World: Comparing sources and contexts of trust and legitimacy* (1st ed., pp. 3–26). Taylor and Francis.

Rosie, M., & Gorringer, H. (2009). What a Difference a Death Makes: Protest, Policing and the Press at the G20. *Sociological Research Online*, 14(5), 68–76.

<https://doi.org/10.5153/sro.2047>

Sahin, N., Braga, A. A., Apel, R., & Brunson, R. K. (2017). The Impact of Procedurally-Just Policing on Citizen Perceptions of Police During Traffic Stops: The Adana Randomized Controlled Trial. *Journal of Quantitative Criminology*, 33(4), 701–726.

<https://doi.org/10.1007/s10940-016-9308-7>

Sato, M. (2017). Police legitimacy and public cooperation: Is Japan an outlier in the procedural justice model. In D. Oberwittler & S. Roché (Eds.), *Routledge Frontiers of Criminal Justice. Police-Citizen Relations Across the World: Comparing sources and contexts of trust and legitimacy* (1st ed., pp. 108–126). Taylor and Francis.

Stommes, D., Aronow, P. M., & Sävje, F. (2021). On the reliability of published findings using the regression discontinuity design in political science. *ArXiv:2109.14526 [Stat]*.

<http://arxiv.org/abs/2109.14526>

Tankebe, J. (2013). Viewing Things Differently: The Dimensions of Public Perceptions of Police Legitimacy: Public Perceptions of Police Legitimacy. *Criminology*, 51(1), 103–135. <https://doi.org/10.1111/j.1745-9125.2012.00291.x>

- Thomassen, G., Strype, J., & Egge, M. (2014). Trust no Matter What? Citizens' Perception of the Police 1 Year after the Terror Attacks in Norway. *Policing*, 8(1), 79–87.
<https://doi.org/10.1093/police/pat032>
- Thompson, J. (2021). *Support for same-sex marriage after the Pulse nightclub shooting: Evidence from a quasi-experiment* [Preprint]. Open Science Framework.
<https://doi.org/10.31219/osf.io/64qm9>
- Trinkner, R., & Tyler, T. R. (2016). Legal Socialization: Coercion versus Consent in an Era of Mistrust. *Annual Review of Law and Social Science*, 12(1), 417–439.
<https://doi.org/10.1146/annurev-lawsocsci-110615-085141>
- Tuch, S. A., & Weitzer, R. (1997). The polls – trends: Racial differences in attitudes toward the police". *Public Opinion Quarterly*, 61, 642–63.
- Tyler, T. R., & Blader, S. L. (2003). The Group Engagement Model: Procedural Justice, Social Identity, and Cooperative Behavior. *Personality and Social Psychology Review*, 7(4), 349–361. https://doi.org/10.1207/S15327957PSPR0704_07
- Tyler, T. R., Fagan, J., & Geller, A. (2014). Street Stops and Police Legitimacy: Teachable Moments in Young Urban Men's Legal Socialization: Street Stops and Police Legitimacy. *Journal of Empirical Legal Studies*, 11(4), 751–785.
<https://doi.org/10.1111/jels.12055>
- Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*.
<http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1069888>

- Tyler, T. R., & Jackson, J. (2014). Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law*, 20(1), 78–95.
- Walters, G. D., & Bolger, P. C. (2019). Procedural justice perceptions, legitimacy beliefs, and compliance with the law: A meta-analysis. *Journal of Experimental Criminology*, 15(3), 341–372. <https://doi.org/10.1007/s11292-018-9338-2>
- Wang, P., Joosse, P., & Cho, L. L. (2020). The Evolution of Protest Policing in a Hybrid Regime. *The British Journal of Criminology*, azaa040. <https://doi.org/10.1093/bjc/azaa040>
- Weitzer, R. (1999). Citizen's perceptions of police misconduct: Race and neighborhood context". *Justice Quarterly*, 16, 819–846.
- Weitzer, R. (2002). Incidents of police misconduct and public opinion. *Journal of Criminal Justice*, 30(5), 397–408. [https://doi.org/10.1016/S0047-2352\(02\)00150-2](https://doi.org/10.1016/S0047-2352(02)00150-2)
- Weitzer, R. (2017). *Recent trends in police-citizen relations* (D. Oberwittler & S. Roché, Eds.; 1st ed.). Taylor and Francis.
- White, C., Weisburd, D., & Wire, S. (2018). Examining the Impact of the Freddie Gray Unrest on Perceptions of the Police. *Criminology & Public Policy*, 17(4), 829–858. <https://doi.org/10.1111/1745-9133.12404>
- Wood, L. (2020). Policing counter-protest. *Sociology Compass*, 14(11), 1–10. <https://doi.org/10.1111/soc4.12833>
- Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.

SUPPLEMENTAL ONLINE MATERIALS

A1. GENERAL APPENDIX

This appendix which accompanies the paper “Protest Policing and public perceptions of police: Evidence from a natural experiment in Germany” is structured as follows: First, Chapter A1 presents general information (such as a general description of the ESS5, results from factor analyses, tabulations of coefficient plots and the alternative model specifications mentioned in the paper), before Chapters A2 and A3 go on to examine the specific UESD assumptions.

A.1.1. DESCRIPTION OF THE DATA SOURCE

As stated in the paper, the ESS a cross-national survey that reviews attitudes, beliefs and behavior patterns in European countries every two years on the basis of probability samples gathered through face-to-face interviews. We use the ESS5 which has been conducted throughout the year 2010 in 28 European countries. The specific file we use is the ESS5-2010 Edition 3.4 which was released on 1 December 2018. Apart from the other rounds of the ESS, the ESS5 includes 45 questions on Trust in Justice which as of 2021 have not been repeated in other rounds. Further information on this specific survey round can be found via the ESS website: <https://www.europeansocialsurvey.org/data/download.html?r=5>.

A.1.2. FACTOR ANALYSES

Table A1. Standardized loadings (pattern matrix) based upon correlation matrix (method is minimal residual).

Items	MR1	MR2	MR3
'How often do police treat people in country with respect'	-0.03	-0.01	0.71
'How often do police make fair, impartial decisions'	0.02	-0.01	0.68
'How often do the police explain their decisions and actions when asked'	0.04	0.02	0.48
'Police have the same sense of right and wrong as me'	-0.02	0.67	0.00
'Police stand up for values that are important to people like me'	0.02	0.85	0.06
'I generally support how the police act'	-0.04	0.67	-0.13
'Duty to: back decisions made by police, even if disagree'	0.71	-0.04	0.02
'Duty to: do what police say, even when don't understand or agree'	0.98	0.03	-0.03
'Duty to: do what police say even if treated badly'	0.79	-0.04	0.04

A.1.3. FULL REGRESSION TABLES

Table A2 shows the full regression results from the models that are summarized in Figure 3 in the paper. Table A3 gives results from the same models but using the whole fieldwork period of the ESS as the sample. Table A4 and A5 use the shortened and prolonged fieldwork period but without fixing the number of missing values (we use listwise deletion). Table A6 and A7, again, only use the +/- 15 days bandwidth but apply design and post-stratification weights respectively. Finally, Table A8 and A9 do the same but for the whole fieldwork period. Across all different model specifications, the effect on political impartiality is sizeable and statistically significant for the most conservative significance level. The effect on trust in the police, however, is not. Depending on the modelling strategy, the effect is either only significant for $p < .1$ or non-significant. Only Table A4 shows a significant effect at the 5% level in the baseline model. Adding covariates and fixed effects again results in an effect only significant at the 10% level, adding further doubts about the reliability of this finding.

Table A2. Full Regression Models for the +/-15 days bandwidth

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.218*** (0.072)	-0.244*** (0.075)	0.307* (0.165)	0.300* (0.174)	-0.034 (0.038)	-0.026 (0.039)	-0.089* (0.048)	-0.066 (0.051)	0.092 (0.172)	0.112 (0.181)
Age		-0.002 (0.002)		0.005 (0.005)		0.004*** (0.001)		0.006*** (0.001)		0.007 (0.005)
Female		0.081 (0.071)		0.194 (0.164)		-0.055 (0.037)		-0.009 (0.048)		-0.292* (0.171)
Education Years		-0.002 (0.011)		-0.001 (0.025)		0.011* (0.006)		0.011 (0.007)		-0.014 (0.026)
log(Income)		0.135** (0.058)		0.254* (0.132)		-0.049 (0.030)		0.015 (0.039)		-0.094 (0.137)
Ethnic Minority		-0.100 (0.228)		0.189 (0.461)		-0.034 (0.104)		0.112 (0.142)		-0.177 (0.487)
Constant	2.430*** (0.047)		6.627*** (0.107)		2.933*** (0.024)		3.914*** (0.031)		6.621*** (0.111)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	687	687	728	728	638	638	709	709	712	712
R ²	0.013	0.875	0.005	0.910	0.001	0.977	0.005	0.975	0.0004	0.902
Adjusted R ²	0.012	0.871	0.003	0.907	-0.0003	0.976	0.003	0.975	-0.001	0.899

Note: Two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used restrict observations to respondents with complete information available for all variables included in each pair of models for a given dependent variable (e.g. Models 1 and 2 for Political Impartiality).

Table A3. Full Regression Models for the complete fieldwork bandwidth

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.138*** (0.052)	-0.186*** (0.053)	0.176 (0.119)	0.166 (0.121)	0.025 (0.028)	0.036 (0.028)	-0.057* (0.035)	-0.033 (0.035)	0.031 (0.123)	0.077 (0.126)
Age		-0.004*** (0.001)		0.007*** (0.003)		0.005*** (0.001)		0.008*** (0.001)		0.008*** (0.003)
Female		0.040 (0.040)		0.033 (0.091)		-0.019 (0.021)		0.017 (0.026)		-0.003 (0.095)
Education Years		0.013** (0.006)		0.001 (0.014)		0.002 (0.003)		0.002 (0.004)		-0.025* (0.015)
log(Income)		0.045 (0.033)		0.308*** (0.076)		0.016 (0.018)		0.068*** (0.022)		-0.020 (0.079)
Ethnic Minority		-0.137 (0.109)		-0.186 (0.235)		0.013 (0.058)		-0.001 (0.070)		-0.709*** (0.252)
Constant	2.430*** (0.047)		6.627*** (0.108)		2.933*** (0.025)		3.914*** (0.031)		6.621*** (0.112)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,246	2,246	2,375	2,375	2,060	2,060	2,293	2,293	2,323	2,323
R ²	0.003	0.863	0.001	0.907	0.0004	0.975	0.001	0.975	0.00003	0.897
Adjusted R ²	0.003	0.862	0.0005	0.906	-0.0001	0.974	0.001	0.975	-0.0004	0.897

Note: Two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used restrict observations to respondents with complete information available for all variables included in each pair of models for a given dependent variable (e.g. Models 1 and 2 for Political Impartiality).

A.1.3. ALTERNATIVE MODEL SPECIFICATIONS

Table A4. Full Regression Models for the complete fieldwork bandwidth with listwise missing values

	<i>Dependent variable:</i>									
	Political Impartiality	Trust in the Police	Procedural Fairness	Moral Alignment	Obligation to Obey					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.187*** (0.065)	-0.244*** (0.075)	0.376** (0.147)	0.300* (0.174)	-0.030 (0.034)	-0.026 (0.039)	-0.044 (0.045)	-0.066 (0.051)	0.156 (0.155)	0.112 (0.181)
Age		-0.002 (0.002)		0.005 (0.005)		0.004*** (0.001)		0.006*** (0.001)		0.007 (0.005)
Female		0.081 (0.071)		0.194 (0.164)		-0.055 (0.037)		-0.009 (0.048)		-0.292* (0.171)
Education Years		-0.002 (0.011)		-0.001 (0.025)		0.011* (0.006)		0.011 (0.007)		-0.014 (0.026)
log(Income)		0.135** (0.058)		0.254* (0.132)		-0.049 (0.030)		0.015 (0.039)		-0.094 (0.137)
Ethnic Minority		-0.100 (0.228)		0.189 (0.461)		-0.034 (0.104)		0.112 (0.142)		-0.177 (0.487)
Constant	2.470*** (0.042)		6.595*** (0.096)		2.923*** (0.022)		3.874*** (0.029)		6.590*** (0.101)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	854	687	909	728	785	638	881	709	889	712
R ²	0.010	0.875	0.007	0.910	0.001	0.977	0.001	0.975	0.001	0.902
Adjusted R ²	0.009	0.871	0.006	0.907	-0.0003	0.976	-0.00005	0.975	0.00002	0.899

Note: Two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used to restrict observations to respondents with complete information for all variables included in a specific model.

Table A5. Full Regression Models for the complete fieldwork bandwidth with listwise missing values

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.150*** (0.047)	-0.186*** (0.053)	0.195* (0.107)	0.166 (0.121)	0.032 (0.025)	0.036 (0.028)	-0.020 (0.032)	-0.033 (0.035)	0.014 (0.111)	0.077 (0.126)
Age		-0.004*** (0.001)		0.007*** (0.003)		0.005*** (0.001)		0.008*** (0.001)		0.008*** (0.003)
Female		0.040 (0.040)		0.033 (0.091)		-0.019 (0.021)		0.017 (0.026)		-0.003 (0.095)
Education Years		0.013** (0.006)		0.001 (0.014)		0.002 (0.003)		0.002 (0.004)		-0.025* (0.015)
log(Income)		0.045 (0.033)		0.308*** (0.076)		0.016 (0.018)		0.068*** (0.022)		-0.020 (0.079)
Ethnic Minority		-0.137 (0.109)		-0.186 (0.235)		0.013 (0.058)		-0.001 (0.070)		-0.709*** (0.252)
Constant	2.470*** (0.043)		6.595*** (0.097)		2.923*** (0.023)		3.874*** (0.029)		6.590*** (0.101)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,813	2,246	3,019	2,375	2,534	2,060	2,878	2,293	2,941	2,323
R ²	0.004	0.863	0.001	0.907	0.001	0.975	0.0001	0.975	0.00001	0.897
Adjusted R ²	0.003	0.862	0.001	0.906	0.0002	0.974	-0.0002	0.975	-0.0003	0.897

Note: Two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used to restrict observations to respondents with complete information for all variables included in a specific model.

Table A6. Full Regression Models for the +/-15 days bandwidth with design weights applied.

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.258*** (0.072)	-0.253*** (0.075)	0.245 (0.163)	0.285* (0.171)	-0.031 (0.038)	-0.015 (0.039)	-0.084* (0.048)	-0.075 (0.051)	0.073 (0.173)	0.107 (0.180)
Age		-0.002 (0.002)		0.006 (0.005)		0.004*** (0.001)		0.005*** (0.001)		0.003 (0.005)
Female		0.094 (0.072)		0.280* (0.162)		-0.032 (0.037)		0.012 (0.048)		-0.261 (0.172)
Education Years		-0.007 (0.011)		-0.007 (0.024)		0.006 (0.006)		0.009 (0.007)		-0.037 (0.026)
log(Income)		0.080 (0.060)		0.287** (0.133)		-0.025 (0.031)		0.039 (0.040)		-0.076 (0.140)
Ethnic Minority		-0.165 (0.216)		0.405 (0.424)		-0.016 (0.096)		0.175 (0.132)		-0.122 (0.456)
Constant	2.516*** (0.046)		6.769*** (0.105)		2.946*** (0.024)		3.914*** (0.031)		6.569*** (0.111)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	687	687	728	728	638	638	709	709	712	712
R ²	0.019	0.878	0.003	0.914	0.001	0.977	0.004	0.975	0.0003	0.900
Adjusted R ²	0.017	0.875	0.002	0.912	-0.001	0.976	0.003	0.975	-0.001	0.897

Note: Design weight is applied, two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used to restrict observations to respondents with complete information available for all variables included in each pair of models for a given dependent variable (e.g. Models 1 and 2 for Political Impartiality).

Table A7. Full Regression Models for the +/-15 days bandwidth with post stratification weights applied.

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.260*** (0.072)	-0.252*** (0.075)	0.248 (0.164)	0.253 (0.172)	-0.012 (0.038)	0.001 (0.039)	-0.064 (0.049)	-0.060 (0.051)	0.094 (0.175)	0.123 (0.183)
Age		-0.001 (0.002)		0.007 (0.005)		0.004*** (0.001)		0.004*** (0.001)		0.005 (0.005)
Female		0.128* (0.072)		0.272* (0.162)		-0.041 (0.037)		0.007 (0.049)		-0.199 (0.174)
Education Years		-0.003 (0.011)		-0.001 (0.025)		0.008 (0.006)		0.012* (0.008)		-0.034 (0.027)
log(Income)		0.078 (0.060)		0.321** (0.132)		-0.029 (0.031)		0.031 (0.040)		-0.084 (0.140)
Ethnic Minority		-0.284 (0.215)		0.205 (0.423)		0.0002 (0.094)		0.186 (0.133)		-0.191 (0.458)
Constant	2.508*** (0.046)		6.786*** (0.104)		2.943*** (0.024)		3.904*** (0.031)		6.543*** (0.111)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	687	687	728	728	638	638	709	709	712	712
R ²	0.019	0.879	0.003	0.915	0.0002	0.978	0.002	0.975	0.0004	0.898
Adjusted R ²	0.017	0.875	0.002	0.913	-0.001	0.977	0.001	0.975	-0.001	0.895

Note: Post stratification weight is applied, two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used restrict to observations to respondents with complete information available for all variables included in each pair of models for a given dependent variable (e.g. Models 1 and 2 for Political Impartiality).

Table A8. Full Regression Models for the complete bandwidth with design weights applied

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.193*** (0.053)	-0.226*** (0.054)	0.147 (0.120)	0.169 (0.122)	0.022 (0.028)	0.040 (0.029)	-0.048 (0.035)	-0.020 (0.036)	0.033 (0.126)	0.094 (0.129)
Age		-0.004*** (0.001)		0.008*** (0.003)		0.005*** (0.001)		0.008*** (0.001)		0.005* (0.003)
Female		0.036 (0.040)		0.064 (0.091)		-0.011 (0.022)		0.020 (0.027)		0.038 (0.096)
Education Years		0.008 (0.006)		-0.011 (0.014)		-0.0004 (0.003)		-0.0004 (0.004)		-0.046*** (0.015)
log(Income)		0.028 (0.035)		0.343*** (0.077)		0.023 (0.019)		0.079*** (0.023)		-0.028 (0.081)
Ethnic Minority		-0.148 (0.104)		-0.111 (0.220)		0.008 (0.055)		0.0001 (0.067)		-0.670*** (0.239)
Constant	2.516*** (0.048)		6.769*** (0.109)		2.946*** (0.026)		3.914*** (0.032)		6.569*** (0.115)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,246	2,246	2,375	2,375	2,060	2,060	2,293	2,293	2,323	2,323
R ²	0.006	0.865	0.001	0.910	0.0003	0.974	0.001	0.975	0.00003	0.894
Adjusted R ²	0.005	0.863	0.0002	0.909	-0.0002	0.974	0.0004	0.974	-0.0004	0.893

Note: Design weight is applied, Two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used to restrict observations to respondents with complete information available for all variables included in each pair of models for a given dependent variable (e.g. Models 1 and 2 for Political Impartiality).

Table A8. Full Regression Models for the complete bandwidth with post stratification applied

	<i>Dependent variable:</i>									
	Political Impartiality		Trust in the Police		Procedural Fairness		Moral Alignment		Obligation to Obey	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
After 'Black Thursday'	-0.174*** (0.054)	-0.208*** (0.055)	0.145 (0.120)	0.171 (0.122)	0.035 (0.029)	0.054* (0.029)	-0.026 (0.035)	0.002 (0.035)	0.082 (0.127)	0.145 (0.129)
Age		-0.003** (0.001)		0.010*** (0.003)		0.005*** (0.001)		0.008*** (0.001)		0.008*** (0.003)
Female		0.066 (0.041)		0.076 (0.091)		-0.010 (0.022)		0.024 (0.027)		0.077 (0.097)
Education Years		0.004 (0.006)		-0.009 (0.014)		0.001 (0.003)		0.001 (0.004)		-0.043*** (0.015)
log(Income)		0.011 (0.034)		0.351*** (0.075)		0.023 (0.019)		0.073*** (0.022)		-0.057 (0.080)
Ethnic Minority		-0.217** (0.101)		-0.253 (0.213)		0.006 (0.054)		-0.015 (0.064)		-0.750*** (0.233)
Constant	2.508*** (0.049)		6.786*** (0.109)		2.943*** (0.026)		3.904*** (0.032)		6.543*** (0.115)	
Region fixed effects?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,246	2,246	2,375	2,375	2,060	2,060	2,293	2,293	2,323	2,323
R ²	0.005	0.864	0.001	0.911	0.001	0.974	0.0002	0.975	0.0002	0.894
Adjusted R ²	0.004	0.862	0.0002	0.910	0.0002	0.974	-0.0002	0.975	-0.0003	0.893

Note: Post stratification weight is applied, two tailed tests, Standard Errors in parentheses, *** p < .01, ** p < .05, * p < .1. Listwise deletion is used restrict observations to respondents with complete information available for all variables included in each pair of models for a given dependent variable (e.g. Models 1 and 2 for Political Impartiality).

A.2. TESTING UESD ASSUMPTIONS

A.2.1 EXCLUDABILITY ASSUMPTION

Placebo outcomes

Recall that the excludability assumption states that the interview date is an instrumental variable which should affect the outcome(s) only through the event. A convenient way to test this assumption is therefore to run ‘placebo’ regressions on outcomes that are very close to the actual outcome but should, theoretically, not be affected by the interview timing. For the outcome ‘political impartiality’, the ESS contains an item which is very suitable for such a test. Respondents are asked whether they agree that ‘The courts’ decisions are unduly influenced by political pressure’. Other suitable outcomes are all institutional trust variables within the ESS.¹ The results which are summarized in Figure A1 show that none of these outcomes appears to be affected in any way by the event.

¹ We focus on placebo outcomes for political impartiality and trust in the police. Since these are the only outcomes that appear to be affected by the treatment in any way.

Figure A1. Placebo outcomes from the German subsample (+/-15 days bandwidth, covariates included).

Pre-existing time trends

A crucial issue regarding the UESD is that the focal event must not be foreseeable to the respondents since otherwise they might change their attitudes even before it has occurred, thus biasing results. This point requires special consideration in the present investigation, as the police had already been deployed in the Schlossgarten in Stuttgart before the 'Black Thursday', even though the use of brutal force was only reported afterwards.

Figure A2. Pre-existing time trends depicted by locally weighted regression lines (LOESS).

Figure A2 thus summarizes the outcomes' temporal stability in the control group. The plots show the trajectories of the five outcomes before the focal event. Simple regressions on the interview date provide additional formal tests. As can be seen in figure A2, the only outcome to have a significant time trend is 'Moral Alignment'. Adding the covariates and fixed effects to the model reduces the effect to non-significance, which is why we are quite sure that this is a statistical artifact. Overall, there is little indication that there are robust pre-existing time-trends which could have biased the results.

UK sample as a placebo group

Another good test for the excludability assumption is to use samples of the same survey program (a) at a different time in the same country and (b) at the same time in a different country. Since the inclusion of police specific items in the ESS was a unique research program which has not yet been reproduced in different survey waves, option (a) is not possible. We can, however, test whether the UK sample of the ESS was in any way affected by the event since the survey period coincided with the fieldwork period in Germany. Theoretically, this should of course not be the case. Any placebo effects could therefore be an indication that something else happened which might have driven our results. This would then violate the excludability assumption.

Figure A3. Weekly bandwidth choice models in the UK as a placebo comparison.

We use the weekly sampling approach we already applied in the paper (Figure 4), because this way we can see short-term and long-term effects as well as their trajectories over time. The picture that emerges from Figure A3 is quite clear: there is not a single model with a significant effect on any outcome, bolstering the robustness of the excludability assumption.

Placebo treatment at median of control group

A final test for the excludability assumption that comes from the regression discontinuity literature is to split the control group at its empirical median and to test for (the absence of) a treatment effect. Figure A4 summarizes simple regressions (due to the small number of observations we do not include covariates in order to avoid losing too many degrees of freedom) that show only a significant effect at the 10% level for procedural fairness. Looking at Figure A2, which depicts the trajectories of outcomes in the control period, makes it quite clear that the depicted non-linearity leads to this 'placebo effect', which is a common problem in regression discontinuity analysis.

Figure A4. Coefficient plot on all outcomes. Cut-off is the empirical median of control group.

A.2. IGNORABILITY ASSUMPTION

One general test that can assess the ignorability assumption is the use of balance checks (see mean differences of controls and outcomes in Table 1 in the paper) to test whether assignment to pre- or post-event group is statistically independent on the covariates. A good strategy is then to use a sample with a maximized balance between those groups and to compare those results with models using the whole fieldwork period (as we do in the paper). Additional analyses with different time-windows like those presented in Figure 4 in the paper also give further confidence that the results are not due to arbitrary bandwidth selection. To see why a narrow bandwidth can support the ‘as good as random’ assumption through reducing bias from both omitted unbalanced covariates as well as potential collateral events, Muñoz et al. (2019) formalize the problem by:

$$Y_i(0), Y_i(1) \perp\!\!\!\perp T_i \mid h_i \rightarrow 0$$

where h_i is i 's absolute distance in days from the day when the event happened. Bandwidth selection is thus a trade-off (since a smaller time windows reduces statistical power), which we have addressed accordingly by using both a series of different time-windows, as well as reporting models with and without covariate adjustment, including federal state level fixed effects.

Another test for the ignorability assumption is to analyze non-response patterns. Assume that, on average, respondents sharing certain characteristics are more likely to participate early in the survey, since they are easier to reach than other respondents. This would be a danger for the

validity of the whole design. Unfortunately, the ESS5 does not allow one to specifically analyze non-response patterns. However, as has been demonstrated in Figure 2 in the paper, there should be no such pattern for the ± 15 days period because non-response in that period is relatively stable. The prolonged period might be more biased (as evidenced by the imbalance checks in Table 1 in the paper), which is we rely on both time periods for our analyses. Figure A5 reproduces a less compressed histogram depicting the weekly number of interviews.

Figure A5. Weekly number of interviews during fieldwork period.