

**'Terrorist' or 'Mentally Ill': Motivated Biases Rooted in Partisanship Shape
Attributions about Violent Actors**

ACCEPTED IN:

SOCIAL PSYCHOLOGY AND PERSONALITY SCIENCE

Running header: 'Terrorist' or 'Mentally Ill'

Masi Noor, Keele University, School of Psychology, UK

Nour Kteily, Northwestern University, Kellogg School of Management

Birte Siem, FernUniversitaet in Hagen, Institute for Psychology

Agostino Mazziotta, FernUniversitaet in Hagen, Institute for Psychology

Please direct correspondence to Masi Noor, Keele University, School of Psychology,
Staffordshire, ST5 5BG, m.noor@keele.ac.uk or masi.noor1@gmail.com

Abstract

We investigated whether motivated reasoning rooted in partisanship affects the attributions individuals make about violent attackers' underlying motives and group memberships. Study 1 demonstrated that on the day of the Brexit referendum pro-leavers (*vs.* pro-remainers) attributed an exculpatory (*i.e.*, mental health) versus condemnatory (*i.e.*, terrorism) motive to the killing of a pro-remain politician. Study 2 demonstrated that pro- (*vs.* anti-) immigration perceivers in Germany ascribed a mental health (*vs.* terrorism) motive to a suicide attack by a Syrian refugee, predicting lower endorsement of punitiveness against his group (*i.e.*, refugees) as a whole. Study 3 experimentally manipulated target motives, showing that Americans distanced a politically-motivated (*vs.* mentally ill) violent individual from their ingroup and assigned him harsher punishment— patterns most pronounced amongst high group identifiers.

Keywords: terrorism, mental illness, attributions, punitiveness, motivated reasoning

'Terrorist' or 'Mentally Ill': Motivated Biases Rooted in Partisanship Shape Attributions about Violent Actors

It is often said that one person's terrorist is another's freedom fighter. Just as perceivers' motivations might lead them to disagree about the righteousness of a violent actor's behaviour, so too might perceivers disagree about the roots of that behaviour in ways that either incriminate or exonerate that actor and align with perceivers' worldviews. One dimension frequently debated when attributing causes to actors' violent behaviour is whether it is rooted in mental illness versus terrorism. For example, following Omar Mateen's massacre at the Pulse gay club in Orlando, a local imam emphasized Mateen's mental health problems (Fox, 2016), whereas others focused on his potential terrorist motivations (Kassam, 2016; Pilkington & Roberts 2016). Elsewhere, mainstream media has been accused of being reticent to describe White perpetrators (e.g., Dylann Roof) as terrorists (Butler, 2015; Greenwald, 2016). Sometimes the motive behind these violent acts is obvious. Often, however, it is ambiguous, allowing room for different motivated interpretations.

These distinctions matter: an aggressor labelled a terrorist is likely to be condemned and punished much more than the same actor designated mentally ill, because the latter is attributed with reduced control and diminished understanding of their actions' consequences (see M'Naghten Rules, Dalby, 2006). Indeed, perceivers' negative affect and behaviour towards a stigmatised target or group depends on how much they are viewed as having control over the factors at the root of their stigmatisation behaviour (Schwarzer & Weiner, 1991; Weiner, Osborne, & Rudolph, 2011; Weiner, Perry, & Magnusson, 1988).

Relatedly, how much an ingroup member's behaviour negatively affects the reputation of one's group depends in part on whether that member is viewed as in control of—and therefore morally responsible for—his/her harmful actions. To protect its status, group members may be motivated to distance their group from a violent individual. This can

occur via mechanisms such as *denying* the violent individual membership in the group (Kteily, Cotterill, Sidanius, Sheehy-Skeffinton, & Bergh, 2014), or— if membership is hard to deny— by framing the target as a *black sheep* (Marques, Yzerbyt, & Leyens, 1988). Alternatively, as we propose here, individuals may seek to *excuse* the behaviour by making *exculpatory* attributions— for example, framing it as caused by mental illness. Similar processes could underlie attributions about the causes of an *outgroup*’s behaviour: If individuals dislike or feel threatened by another group (Riek, Mania, & Gaertner, 2006; Stephan, Ybarra, & Morrison, 2011), they may be motivated to impugn its reputation by making *condemnatory* attributions holding its members responsible for reprehensible acts.

Taken together, attributions of motives to violent acts may be coloured by social bias embedded in the motivation to protect (impugn) one’s ingroup (an outgroup) or a favoured (resented) partisan position. Drawing on motivated reasoning accounts (Kruglanski & Freund, 1983; Kunda, 1990; Munro, Weih, & Tsai, 2010), we investigate whether perceivers’ partisanship predicts their attributions of motives (terrorism *vs.* mental illness) to violent actors, and consider the downstream consequences of these attributions for the violent actors and their ingroups. We tested these novel predictions across different geo-political contexts, involving two consequential and psychologically salient real-life events (Studies 1 & 2), and a controlled experimental study (Study 3).

Motivated Reasoning in Ascribing Terrorism *vs.* Mental Illness

Motivated reasoning perspectives suggest that perceivers’ investment in their social contexts leads them to select and filter information in a biased, self- or ingroup-serving manner to arrive at conclusions that favour their valued positions (Kruglanski & Freund, 1983; Kunda, 1990; Munro et al., 2010). One illustrative study demonstrated that liberals and conservatives watching the same video of a protest saw protestors as *aggressive* when their

political stance purportedly opposed perceivers' own but as *peaceful* when their stance aligned with perceivers' valued position (Kahan, Hoffman, Braman, & Evans, 2012; see also Hulsizer, Munro, Fagerlin, & Taylor, 2004). Recent work has examined motivated bias in the context of broader ideological stances. For example, Kteily and colleagues (2014) established that White Americans and British participants were more likely to perceive racially and ethnically ambiguous low-status targets (e.g., Dzhokhar and Tamerlan Tsarnaev) in outgroup (*vs.* ingroup) terms when participants were higher in social dominance orientation and right-wing authoritarianism (and thus particularly motivated to maintain the standing or norms of their high-status ingroup). Consistent with the idea that attributions about actors' violence can have downstream consequences, perceiving these targets in more exclusionary ways predicted support for punishing them more severely.

The Current Research

We examine whether motivated reasoning extends to predict the (exculpatory *vs.* condemnatory) attributions that individuals make for targets' violent behaviour — here, attributions of terrorism versus mental health made for violent perpetrators' actions. We argue that when the violent behaviour could shed negative light on one's valued partisan position, perceivers' motivations will lead them to distance the perpetrator from these positions by ascribing the perpetrator an exculpatory motive (i.e., mental illness). Conversely, opponents of the same partisan position might be motivated to magnify the association between the violent individual and the position they are opposing by attributing to the perpetrator a more condemnatory motive (i.e., terrorism). Because terrorism represents an action on behalf of a political collective, this would also enable motivated perceivers to link a violent actor's behaviour to his/her group membership, rendering the collective 'guilty by association' (Doosje, Branscombe, Spears, & Manstead, 1998) and facilitating vicarious retribution (Lickel, Miller, Stenstrom, Denson, & Schmader, 2006).

Study 1

Study 1 examined violence immediately preceding Britain's heated national referendum vote regarding exit from the European Union. A week before the referendum, 52-year-old Thomas Mair murdered Jo Cox, a British Member of Parliament known for her support of the 'Remain' campaign seeking to keep Britain in the E.U. Immediately after (and just preceding the vote), Mair's motive remained ambiguous, with police investigating both reports of him shouting "Britain first" during the act – a reference to the ultra-right political party by the same name – and the possibility of his suffering mental health problems (BBC, 2016; Pidd, 2016). To capture participants' responses while events remained salient, we conducted Study 1 on the actual referendum day.

We predicted that motivated reasoning would lead individuals to make biased ascriptions of Mair's motives. Specifically, we expected Leave (*vs.* Remain) supporters to be more likely to attribute him a mental illness (*vs.* terrorism) motive. We further investigated consequences of these attributions for participants' aggressive reactions towards Mair, focusing on punitive measures. A mentally ill aggressor is viewed as less in control of and culpable for the consequences of his/her actions (Dalby, 2006; see also Weiner et al., 2011). Thus, we reasoned that the more Leave (*vs.* Remain) supporters attributed Mair's violence to mental illness (*vs.* terrorism), the less punitive they would be towards him.

Interestingly, this context contained potential countervailing motives beyond partisanship. For example, due to their shared national identification with the (British) victim, Leave supporters might be motivated to some extent to condemn Mair (by labelling him a terrorist) so as to disassociate him from the *national* ingroup, a potential countervailing force to Leave supporters' motivation to protect their partisan ingroup (i.e., the Leave campaign) by attributing his actions to mental illness. However, given the salience and divisiveness of

the referendum, we predicted that our participants would be more likely to process Mair's action through the lens of political partisanship (*vs.* national identity; see Musgrove & McGarty, 2008; Thomas & McGarty, 2009).

Method

Participants and Procedure

Our sample comprised 234 British adults, participating online via Qualtrics Panels. Data collection commenced at 11:41 AM and completed at 16:33 PM on referendum day (June 23, 2016). We excluded 44 participants missing data on key variables ($n = 28$), being of mixed ($n = 2$) or non-British nationality ($n = 9$), or failing attention checks ($n = 5$). The final sample consisted of 190 adults (46% male; $M_{\text{age}} = 44.96$; $SD_{\text{age}} = 16.70$). Referendum position consisted of two levels: Leave supporters ($n = 97$) *vs.* Remain supporters ($n = 93$). Post-hoc power analysis revealed this study was sufficiently powered (power = .93; details in SOM).

Measures

Unless noted otherwise, all scales across studies ranged from 0 ('*not much*') to 100 ('*very much*'). Full measures, including those for exploratory analyses not reported here, are available in the SOM.

Partisanship. After providing demographics, participants indicated whether or not they favoured Britain leaving the E.U. (1 = *yes*; -1 = *no*).

Attributions for violence. Participants were asked to fill the gap in the sentence 'Thomas Mair is a _____ who killed Jo Cox'. Data were coded for references to terrorism (e.g., 'extremist') *vs.* mental illness ('psycho') (details in SOM). Beyond this subtle metric, two items measured participants' explicit motive attributions (e.g., "I have no doubts

that Thomas Mair killed Jo Cox because he is: ‘*a terrorist*’ — ‘*mentally ill*’). The z -standardized subtle and explicit measures were correlated, $r(190) = .48, p < .001$; we therefore created a composite whereby higher scores indicated more attribution of mental illness.

Punitiveness. Punitiveness was measured with three items, modelled on Kteily et al. (2014) (e.g., “Thomas Mair should be placed in solitary confinement for the duration of any time he spends in jail”; $\alpha = .66$).

Results

As predicted, results revealed that Leave supporters ($M_{\text{Leave}} = 0.24, SD = 0.78$) were significantly more likely than Remain supporters to attribute Mair’s killing to mental illness relative to terrorism ($M_{\text{Remain}} = -0.25, SD = 1.14, t(188) = -3.43, p = .001, \text{Cohen’s } d = .50, \text{mean difference} = -0.48, 95\% \text{ CI } [-0.76, -0.20]$). Although results for punitiveness were in the expected direction—Leave supporters ($M_{\text{Leave}} = 60.27, SD = 23.59$) indicated less punitiveness than Remain supporters ($M_{\text{Remain}} = 64.55, SD = 23.11$)—this difference was not significant, $t(185) = -1.25, p = .21, \text{Cohen’s } d = .003, \text{mean difference} = -4.28, 95\% \text{ CI } [-2.46, 11.02]$.¹ We nevertheless examined evidence for an indirect effect on punitiveness via attribution of terrorism vs. mental illness. Using Hayes’ (2013) PROCESS macro (Model 4), we observed support for this relationship (see Figure 1), suggesting that Leave supporters’

¹ Research suggests a link between right-wing attitudes and punitiveness (e.g., Kteily et al., 2014). Given that the Leave campaign was associated with the political right (Moore, 2016), we considered, in an exploratory analysis, whether right-wing authoritarianism might be suppressing the link between referendum position and punitiveness. Leave supporters were significantly higher on RWA ($p = .001$); after controlling for this difference, we observed a significant direct effect from referendum position to punitiveness ($p = .02, \eta_p^2 = .03$). See SOM for details.

lower likelihood of attributing Mair’s actions to terrorism predicted less punitiveness towards him, $IE = -0.05$, $SE = 0.03$; 95% CI [-0.12, -0.02].

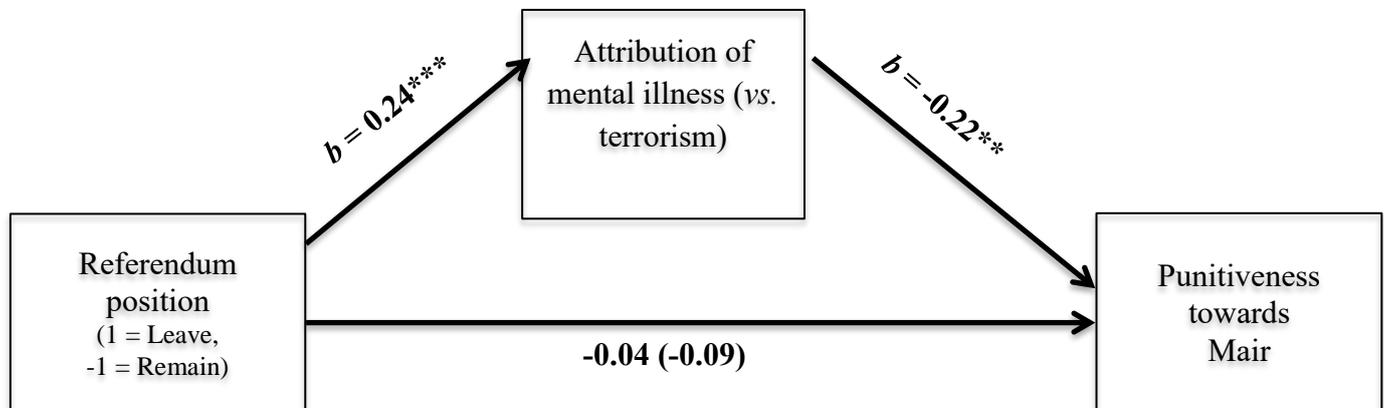


Figure 1. Indirect effect ($IE = -0.50$, 95% CI [-0.12, -0.02]) of Brexit referendum position on punitiveness towards Thomas Mair through motive attribution (mental illness vs. terrorism).

Coefficients are unstandardized regression weights. $**p < .01$; $*** p < .001$

Discussion

Study 1 provides novel evidence that attributions of terrorism vs. mental illness to a real-life violent individual can vary as a function of perceivers’ partisan positions. Moreover, we observed an indirect (but not total) effect between participants’ partisanship and their punitiveness towards Mair via participants’ motive attributions. Notably, this occurred despite potential countervailing motives rooted in Leave and Remain supporters’ shared national identity as British.

In a second real-life violent context, we examined these findings’ generalizability and explored consequences of the terrorism vs. mental illness attribution extending beyond the violent individual.

Study 2

Study 2 examined host-immigrant intergroup relations in Germany, who recently admitted almost one million immigrants (Weaver & Farrer, 2016). German public opinion on immigration has been divided, fuelled by reports implicating immigrants in several violent incidents. We examined our hypotheses against the backdrop of an incident in the Bavarian city of Ansbach in July 2016, in which Mohammad Daleel, a 27-year-old Syrian refugee, set off an explosive device outside a music festival, killing himself and injuring fifteen German civilians. As with Mair's killing, the motives behind Daleel's attack (i.e., terrorism *vs.* mental illness) were subject of debate (see SOM). Tapping participants' immediate responses, we launched Study 2 one day after the incident and completed it within a week.

We predicted immigration supporters (*vs.* opponents) would be more likely to deem the attacker mentally ill (*vs.* terrorist). We further tested whether attributing mental illness motives to Daleel might be associated with punitiveness towards his whole ingroup. We based this prediction on research suggesting that perceivers often fail to distinguish between guilty and innocent outgroup members (Doosje et al., 1998), and sometimes engage in vicarious retribution towards the entire group (Lickel et al., 2006). The present context again contained potential countervailing motives: immigration advocates more focused on their shared national identification with the (German) victims might feel motivated to impugn Daleel (a foreigner) by attributing him a terrorist motive.

Method

Participants and Procedure

Data collection commenced on July 25, 2016, at 20:05 PM, one day after the Ansbach attack, and completed on August 1, 2016 at 20:46 PM. We recruited 316 German participants from a distance teaching university and German media outlets' social media pages. We excluded 38 participants due to being non-German ($n = 24$), failing attention checks ($n = 13$),

or admitting to not taking the study seriously ($n = 1$) (final sample: $N = 278$ participants; 47.5% male; $M_{\text{age}} = 37.12$; $SD_{\text{age}} = 13.90$). Post-hoc power analysis revealed this study was sufficiently powered (power = .98) (see SOM).

Measures

Partisanship. After providing demographics, participants indicated whether they supported immigration (1 = yes; -1 = no) to Germany. Overall, 235 participants indicated being pro-immigration supporters whereas 43 were opposed, comparable to national proportions (see Zick, Küpper, & Krause, 2016).

Attributions for violence. Participants' attributions were assessed as in Study 1, adapted to the current context. Measures of subtle and explicit attributions were again correlated, $r(276) = .57$, $p < .001$, and thus z -standardized to create a composite (higher scores indicating greater attribution of mental illness).

Punitiveness towards immigrants. As Daleel died in his attack, we did not assess punitiveness towards him as an individual (as we had with Mair in Study 1). To tap punitive consequences for Daleel's ingroup (i.e., immigrants) as a whole, participants indicated their agreement with five statements (Armbrost, 2014) endorsing punitive measures against migrants and refugees in Germany (e.g., "I generally find the punishments handed out to criminal migrants and refugees too light"; reverse-coded; $\alpha = .88$).

Results

Because the sample sizes differed between the two groups, we followed Delacre, Lakens, and Leys' (2017) recommendation to use Welch's t -test when testing for mean differences. Welch's t -test offers a stronger control for Type 1 error rates when homogeneity of variance is not met, while being similarly robust compared to Student's t -test when

homogeneity assumptions are met.² Conceptually replicating Study 1, immigration supporters ($M_{\text{Supporters}} = 0.09$, $SD = 0.85$) were significantly more likely than opponents ($M_{\text{Opponents}} = -0.49$, $SD = 0.91$) to attribute Daleel’s attack to mental illness relative to terrorism, Welch’s $t(56.21) = -3.85$, $p < .001$, Cohen’s $d = 0.66$, mean difference = -0.58 , 95% CI $[-0.88, -0.28]$. Moreover, as expected, immigration supporters ($M_{\text{Supporters}} = 35.78$, $SD = 25.42$) were less punitive towards immigrants following the Daleel incident than immigration opponents ($M_{\text{Opponents}} = 72.65$, $SD = 21.11$), Welch’s $t(66.39) = 10.18$, $p < .001$, Cohen’s $d = 1.58$, mean difference = 36.87 , 95% CI $[29.64, 44.10]$. Again consistent with Study 1, results from an indirect effects analysis suggested that attributions of Daleel’s actions to mental illness (vs. terrorism) helped to explain a part of the link between supporting immigration and lower punitiveness towards immigrants in the aftermath of the attack, IE = -0.11 , $SE = 0.03$, 95% CI $[-0.18, -0.05]$ (see Figure 2).

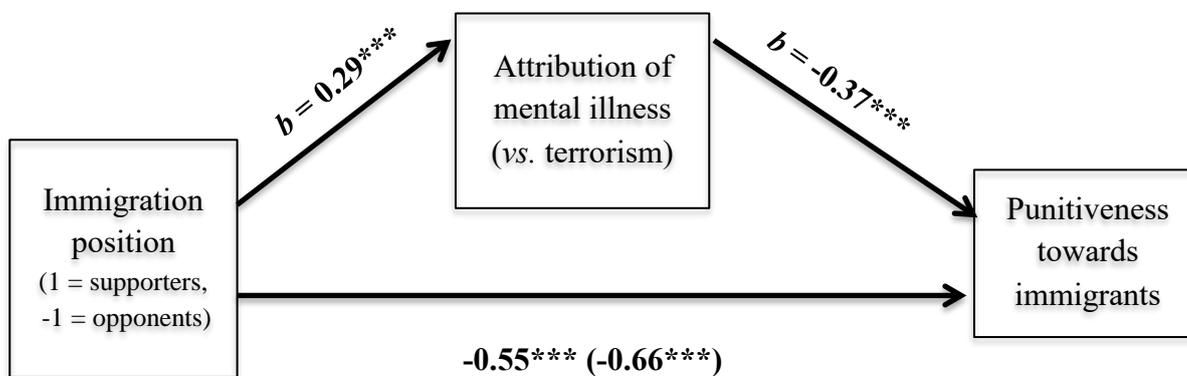


Figure 2. Indirect effect (IE = -0.11 , 95% CI $[-0.18, -0.05]$) of immigration position on punitiveness towards immigrants through attribution of Mohammad Daleel’s motives (mental illness vs. terrorism). Coefficients are unstandardized regression weights. $^{***} p < .001$

² Indirect effect analyses used regular regression/ t -test in the absence, to our knowledge, of an equivalent for Welch’s t -test.

Discussion

Extending results to a second real-life, violent event, German participants' attributions of terrorism versus mental illness motives to Daleel's suicide killing varied depending on their immigration policy positions. Study 2 further highlighted that these attributions can bear consequences not only for the target individual, but his or her whole ingroup (Lickel et al., 2006): the ascription of a mental illness (*vs.* terrorism) motive predicted the link between immigration position and lower punitiveness towards Daleel's ingroup members.

Study 3

Studies 1 and 2 provided (correlational) evidence suggesting that, when the motivations of a violent actor associated with their partisan position remain uncertain, individuals make exculpatory attributions (*i.e.*, mental illness). If (as we reason) this behaviour is rooted in a desire to protect the ingroup's standing, partisans should seek to distance the violent actor from the ingroup in other ways when his condemnable motivations cannot be denied. In our pre-registered Study 3, we directly manipulated the violent target's likely motives (mental illness *vs.* terrorism) and then assessed the impact of the manipulation on the extent to which participants distanced him from their ingroup. We expected Americans would be motivated to distance their ingroup from a target when they were given more condemnatory (*i.e.*, politically-motivated; terrorist) versus exculpatory (mental illness) motivations for the target's actions. We examined distancing by assessing how much Americans deemed the target to be foreign, indexed by perceptions of his prototypicality and beliefs about his religious background. If distancing of blameworthy targets indeed reflects motivated reasoning to protect the ingroup, it should be particularly likely among strongly identified American perceivers, for whom the ingroup's standing is most important (Lewis & Sherman, 2010; Yzerbyt, Castano, Leyens, & Paladino, 2000). In a secondary analysis, we explored this possibility here.

Method

Participants and Procedure

This study was pre-registered with AsPredicted.org (<http://bit.ly/2mppN8b>). We specified *a priori* a desired sample size of 500 valid participants. We recruited 620 American residents from mTurk, of whom 509 successfully passed both attention checks (52.1% male; $M_{\text{age}} = 35.44$; $SD_{\text{age}} = 11.17$).

Materials

All items were assessed on a 1 ('*strongly disagree*') to 7 ('*strongly agree*') scale. See SOM for all measures (including exploratory ones).

American identification. American identification was assessed using a three-item scale adapted from Doosje, Ellemers, and Spears (1995; e.g., "I feel a bond with Americans"; $\alpha = .90$).

Experimental materials. The experiment was presented as a study of how people use intuitions to make inferences about others (see SOM). Participants read a summary about a student ('Mr A.') arrested and charged for acting violently towards members of the public. Participants in the 'mental health' condition ($n = 163$) read that the violent individual's medical records revealed bipolar disorder and anti-depressant usage. In the 'politics' condition ($n = 177$), participants learned that police records revealed his heavy involvement in campus political gatherings. Our design also included a control condition ($n = 169$) where participants were given no information about the target's motives. As stated in our pre-registration, comparisons between our critical experimental conditions and the control condition were secondary. For brevity, we therefore do not include these comparisons in the main text (available in full in SOM). Importantly, all analyses below comparing the terrorism

and mental health conditions are taken from overall regressions or ANOVAs using the *total* sample (i.e., including the data from the control condition).

Measures of distancing

Prototypicality. From among filler questions, participants reported their perception of whether Mr A.’s parents were probably born in the U.S. and about the extent of his love for America. As pre-registered, we formed a composite from these items, $r(509) = .38, p < .001$, with higher scores indicating greater prototypicality with respect to the American ingroup.³

Muslim-ness. Given Islam’s status as a minority religion in the U.S. and its strong association with foreignness and distance from the U.S. ingroup (Selod, 2015), attributing Muslim-ness to a target also implies a form of distancing. Participants indicated whether Mr A. was “Jewish”, “Christian”, “Muslim”, or “Atheist”. To capture perceived Muslim-ness, we calculated a difference score between perceived likelihood of the target being Muslim and the perceived average likelihood of the target having one of the other three possible religious group memberships.⁴

Punitiveness. Participants indicated their endorsement of measures to punish or rehabilitate the target. Two items assessed support for offering the target voluntary or compulsory counselling, $r(509) = .47, p < .001$. Three items assessed participants’ support for jailing the target across time periods ranging from one year of house arrest to a minimum of 20 years’ imprisonment ($\alpha = .80$). We took the difference score between these two scales as our index of punitiveness (higher scores = more punitive).

Results

³ Results were very similar—albeit marginally significant ($ps < .06$)—for each of these items considered separately.

⁴ Excluding Muslims in our sample ($n = 4$) did not affect conclusions.

As pre-registered, we conducted planned comparisons between the mentally-ill and politically-motivated conditions (omnibus ANOVAs including control condition can be found in SOM). As predicted, individuals were significantly more likely to distance the target (i.e., deeming him non-prototypical and more likely to be Muslim) and to express greater punitiveness when he was said to be politically-motivated versus mentally ill (see Table 1).

Next, we examined whether American identification moderated effects of experimental condition on our key variables, using PROCESS 2.16 (Hayes, 2013; Model 1) and specifying a multi-categorical independent variable (i.e., condition). To capture the three experimental conditions, this analysis employs dummy coding with two vectors (D1: control condition = 1, other two conditions = 0; D2: politically-motivated condition = 1, other two conditions = 0). As predictors of the outcome measure, we entered American identification (mean-centered), D1, D2, and the two-way interaction between American identification and each of these two vectors. We compared the effect of experimental condition on the outcome variables at one standard deviation above and below the mean on American identification (note that the values reported below reflect *projected* means). As noted above, although results are taken from the full model, we focus here on the results relevant to comparing effects of the mental health versus the politically-motivated condition.

Table 1

Means and standard deviations of dependent variables by key experimental conditions

| Outcomes | Experimental Condition | | <i>p</i> -value | Cohen’s <i>d</i> |
|-----------------|--|---|-----------------|------------------|
| | Mentally Ill <i>M</i> (<i>SD</i>) | Politically- Motivated <i>M</i> (<i>SD</i>) | | |
| Prototypicality | 4.34 (1.03) | 3.97 (1.26) | .003 | 0.32 |
| Muslim-ness | 0.23 (1.49) | 0.61 (1.97) | .036 | 0.22 |
| Punitiveness | -0.60 (2.50) | 0.04 (2.54) | .023 | 0.25 |

As theorized, American identification significantly interacted with experimental condition (i.e., mentally-ill vs. politically-motivated) to predict perceived target prototypicality, $b = -0.33$, $p < .001$, 95% CI [-0.49, -0.16]. There were no significant differences between the mental health and politically-motivated conditions among low American identifiers ($-1SD$) ($F < 1$). In contrast, high American identifiers ($+1SD$) in the politically-motivated condition ($M_{projected} = 3.78$) saw the target as significantly less prototypical of the U.S. than those in the mentally-ill condition ($M_{projected} = 4.66$), mean difference = 0.87, $SE = 0.18$, $p < .001$, 95% CI [0.53, 1.22].

American identification also interacted with experimental condition to predict perceived Muslim-ness, $b = 0.41$, $p < .001$, 95% CI [0.17, 0.65]. There was no difference across conditions among low identifiers ($p = .38$). In contrast, high American identifiers in the politically-motivated condition were significantly more likely to see the target as Muslim ($M_{projected} = 1.16$) compared to the mentally-ill condition ($M_{projected} = 0.19$), mean difference = 0.97, $SE = 0.26$, $p < .001$, 95% CI [0.46, 1.48].

Although there was a significant main effect of condition on punitiveness (see Table 1), this effect was not moderated by American identification ($p = .44$).

Moderated mediation. American identification moderated the effects of experimental condition on our distancing measures (i.e., perceived prototypicality to the U.S., and Muslim-ness) but not on punitiveness. We nevertheless examined (see Muller, Judd, & Yzerbyt, 2005, and the definition of prototypical moderated mediation), in an exploratory analysis, whether there was evidence of significant moderation in the strength of the indirect effects linking experimental condition to punitiveness via perceived prototypicality to the U.S. and Muslim-ness (entered simultaneously into the model; using PROCESS, Model 7, which

tests moderation of the a paths from the independent variable to the mediators). We observed evidence of significant moderation of condition's indirect effects on punitiveness via each index of distancing. Specifically, the indirect pathway from condition to punitiveness via perceived prototypicality was significantly stronger among high American identifiers, $IE = 0.44$, 95% CI [0.19, 0.81], relative to low American identifiers $IE = -0.05$, 95% CI [-0.23, 0.09], index of moderated mediation = 0.17, 95% CI [0.07, 0.32]. The same was independently true for the indirect pathway through Muslim-ness: high American identifiers $IE = 0.15$, 95% CI [0.01, 0.38]; low American identifiers $IE = -0.03$, 95% CI [-0.17, 0.03]; index of moderated mediation = 0.06, 95% CI [0.006, 0.16].

Discussion

Study 3 broadened our findings' scope, highlighting that perceivers' motivations predict not only their attributions of a target's motives as a function of that target's partisan position, but also their attributions of a target's group membership as a function of his likely motives. Just as individuals protect their partisan position by attributing exculpatory motives for violent acts by ingroup members, they achieve the same by distancing targets from their group when the motivations behind targets' actions threaten the standing of the group. We thus highlight another mechanism—alongside work on the “black sheep effect” (Marques et al., 1988)—by which individuals can protect their group's reputation. Notably, that this pattern was accentuated among strong group identifiers is highly consistent with the idea that these patterns reflect motivated reasoning. It also counteracts potential concerns that participants might deem the violent target foreign merely because of stereotypic associations Americans have between violent actors and foreigners: Whereas high American identifiers might well care more about protecting the standing of their group, there is no clear reason why they would be more likely than lower identifiers to be *knowledgeable* about this stereotypic association.

General Discussion

Building on motivated reasoning research, two studies revealed that individuals varied in attributing exculpatory (i.e., mentally ill) versus condemnatory (i.e., terrorist) motives to the same violent target as a function of individuals' partisan beliefs. A third study provided evidence for the reverse process, showing that individuals—and especially, high ingroup identifiers—were more likely to distance targets from the ingroup when their violent acts could not be excused as a result of mental illness. The attribution of terrorist (*vs.* mental illness) motives predicted harsher treatment of the perpetrator, and—consistent with work on vicarious retribution (Lickel et al., 2006)—their fellow ingroup members.

Our work is the first to document how motivated reasoning influences the ascription of two specific motives—terrorism *vs.* mental illness—frequently disputed in many consequential real-world contexts (most recently, the Las Vegas shootings killing 58 people; BBC, 2017). Our results highlight the political importance and ramifications of this distinction: Deeming someone a terrorist says something not only about them, but about their group, and calls for different—more punitive—responses. Notably, we conducted our research amidst two important events as they were occurring, boosting confidence in our results' ecological validity.

We build on research highlighting the versatility of motivated perceptions (e.g., Hartmant & Newmark, 2012; Kundra & Sinclair, 2009). One important contribution we make is demonstrating that the impact of motivated partisan perceptions transcends ethnic boundaries (and associated potential countervailing motives). In Study 1, we found that despite our participants sharing an ethnic group membership with the perpetrator, their position on the Brexit referendum nevertheless predict their attributions of motive (terrorism *vs.* mental illness) behind his murder of an ethnic ingroup (but, for 'Leave' supporters,

political outgroup) member. Study 2 similarly showed that pro-immigration participants ‘protected’ an ethnic outgroup attacker by assigning him a more benign motive (mental illness), despite the harm his actions imposed on ethnic ingroup members.

One limitation of our work is its correlational design. Future work could systematically (and jointly) manipulate the ethnic and partisan group memberships of a perpetrator to more clearly establish their relative contributions to the ascription of motives. Future research could also experimentally manipulate perceiver partisanship in order to infer causality. Relatedly, we note that although we modelled punitiveness as a *consequence* of attributed motives on theoretical grounds (e.g., Weiner, 1980; see also Darley & Pittman, 2003), our correlational data cannot rule out the viability of the reverse causal path.

Future research could also consider the role of motive ambiguity. Here, the perpetrator’s presumed motives (or, in Study 3, group membership) were always ambiguous. It is likely that ambiguity is a necessary (or catalysing) condition for the impact of motivated reasoning to be observed. When an attacker clearly pledges allegiance to a terrorist goal prior to conducting their attack, even staunch partisans should be more constrained from attributing an exculpatory motivation like mental illness. Future work could systematically vary this factor, presenting participants with evidence that restricts (*vs.* invites) ambiguity (e.g., showing or not showing political manifesto at the scene of a mass shooting perpetrated by an immigrant to individuals who favour versus oppose immigration). Interestingly, even when the motive itself is unambiguous, individuals might seek to find other ways to protect their partisan position— for example, framing the attacker as a black sheep (Marques et al., 1988) or otherwise diminishing their centrality to the group.

Conclusion

One person’s terrorist *does* appear to be another person’s mentally-ill: We demonstrate partisan bias in individuals’ ascriptions to violent actors of terrorist versus mental illness motives. These attributions are consequential, predicting punishment of the perpetrators themselves and attitudes towards perpetrators’ ingroup members. As partisan polarization continues to increase, better understanding the nature, flexibility, and limits of partisan bias— and how to combat it — is ever-more important.

References

- Armborst, A. (2014). Kriminalitätsfurcht und punitive Einstellungen: Indikatoren, Skalen und Interaktionen [Fear of crime and punitive attitudes: Survey indicators, measuring scales, and interactions]. *Soziale Probleme*, 25, 105-142. Retrieved from <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-419015>
- BBC. (2017, October 3). Las Vegas shooting: Is the gunman a terrorist? *BBC*. Retrieved from <http://www.bbc.com/news/world-us-canada-41483943>
- BBC. (2016, June 17). Jo Cox death: What we know about the suspect. *BBC*. Retrieved from <http://www.bbc.co.uk/>
- Butler, A. (2015, June 18). Shooters of color are called 'terrorists' and 'thugs.' Why are white shooters called 'mentally ill'? *Washington Post*. Retrieved from <https://www.washingtonpost.com>
- Dalby, J. T. (2006). The case of Daniel McNaughton: Let's get the story straight. *American Journal of Forensic Psychiatry*, 27, 17-32.
- Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, 7, 324-336. https://doi.org/10.1207/s15327957pspr0704_05
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology*, 30, 92–101. <http://doi.org/10.5334/irsp.82>
- Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, 75, 872- 886. <https://doi.org/10.1037//0022-3514.75.4.872>
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, 31,

410-436. <https://doi.org/10.1006/jesp.1995.1018>

Fox, B. (2016, June 13). Orlando shooter's imam blames mental illness: 'This has nothing to do with ISIS'. *Talking Points Memo*. Retrieved from <http://talkingpointsmemo.com>

Hartman, T. K., & Newmark, A. J. (2012). Motivated reasoning, political sophistication, and associations between President Obama and Islam. *Political Science & Politics*, *45*, 449-455. <https://doi.org/10.1017/s1049096512000327>

Hayes, A.F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Hulsizer, M. R., Munro, G. D., Fagerlin, A., & Taylor, S. P. (2004). Molding the past: Biased assimilation of historical information. *Journal of Applied Social Psychology*, *34*, 1048-1074. <https://doi.org/10.1111/j.1559-1816.2004.tb02583.x>

Kahan, D. M., Hoffman, D. A., Braman, D., & Evans, D. (2012). They saw a protest: Cognitive illiberalism and the speech-conduct distinction. *Stanford Law Review*, *64*, 851-906. Retrieved from <https://www.stanfordlawreview.org/print/article/they-saw-a-protest/>

Kassam, R. (2016, September 26). Emotions, revenge, repressed homosexuality' ... How the media covered for Orlando terrorist Omar Mateen. *Breitbart*. Retrieved from <http://www.breitbart.com>

Kteily, N., Cotterill, S., Sidanius, J., Sheehy-Skeffington, J., & Bergh, R. (2014). "Not one of us": Predictors and consequences of denying ingroup characteristics to ambiguous targets. *Personality and Social Psychology Bulletin*, *40*, 1231-1247. <https://doi.org/10.1177/0146167214539708>

Kundra, Z., & Sinclair, L. (1999). Motivated reasoning with stereotypes: Activation, application, and inhibition. *Psychological Inquiry*, *10*, 12-22. https://doi.org/10.1207/s15327965pli1001_2

- Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impression primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology, 19*, 448-468.
[https://doi.org/10.1016/0022-1031\(83\)90022-7](https://doi.org/10.1016/0022-1031(83)90022-7)
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498.
<https://doi.org/10.1111/j.1651-2227.2007.00343.x>
- Lewis, A. C., & Sherman, S. J. (2010). Perceived entitativity and the black-sheep effect: When will we denigrate negative ingroup members?. *The Journal of Social Psychology, 150*, 211-225. <https://doi.org/10.1080/00224540903366388>
- Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F., & Schmader, T. (2006). Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review, 10*, 372-390. https://doi.org/10.1207/s15327957pspr1004_6
- Marques, J. M., Yzerbyt, V. Y., & Leyens, J. P. (1988). The “black sheep effect”: Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology, 18*, 1-16.
<https://doi.org/10.1002/ejsp.2420180102>
- Moore, P. (2016, June 27). How Britain voted. Retrieved from <https://yougov.co.uk/news/2016/06/27/how-britain-voted/>
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*, 852- 863.
<https://doi.org/10.1037/0022-3514.89.6.852>
- Munro, G. D., Weih, C., & Tsai, J. (2010). Motivated suspicion: Asymmetrical attributions of the behaviour of political ingroup and outgroup members. *Basic and Applied Social Psychology, 32*, 173-184. <https://doi.org/10.1080/01973531003738551>
- Musgrove, L., & McGarty, C. (2008). Opinion-based group membership as a predictor of

- collective emotional responses and support for pro- and anti-war action. *Social Psychology*, 39, 37–47. <https://doi.org/10.1027/1864-9335.39.1.37>
- Pidd, H. (2016, June 17). Suspect in Jo Cox's killing described as quiet, polite and reserved. *The Guardian*. Retrieved from <https://www.theguardian.com/>
- Pilkington, E., & Roberts, D. (2016, June 14). FBI and Obama confirm Omar Mateen was radicalized on the internet. *The Guardian*. Retrieved from <https://www.theguardian.com>
- Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, 10, 336-353. https://doi.org/10.1207/s15327957pspr1004_4a
- Schwarzer, R., & Weiner, B. (1991). Stigma controllability and coping as predictors of emotions and social support. *Journal of Social and Personal Relationships*, 8, 133-140. <https://doi.org/10.1177/0265407591081007>
- Selod, S. (2015). Citizenship denied: The racialization of Muslim American men and women post-9/11. *Critical Sociology*, 41, 77-95. <https://doi.org/10.1177/0896920513516022>
- Stephan, W. G., Ybarra, O., & Rios-Morrison, K. (2009). Intergroup threat theory. In T. Nelson (Ed.), *Handbook of prejudice* (pp. 44–60). Mahwah, NJ: Lawrence Erlbaum.
- Greenwald, G. (2016, June 17). Retrieved from <https://www.theguardian.com/uk-news/2016/jun/16/suspect-in-mp-killing-described-as-quiet-polite-and-reserved>
- The Guardian Why is the killer of British MP Jo Cox not being called a “terrorist”? *The Intercept*. Retrieved from <https://theintercept.com>
- Thomas, E. F., & McGarty, C. A. (2009). The role of efficacy and moral outrage norms in creating the potential for international development activism through group-based interaction. *British Journal of Social Psychology*, 48, 115–134. <https://doi.org/10.1348/014466608x313774>

- Weaver, M., & Farrer, M. (2016, July 25). Ansbach bombing: attacker pledged allegiance to Isis, says official – as it happened. *The Guardian*. Retrieved from <https://www.theguardian.com>
- Weiner, B. (1980). A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving. *Journal of Personality and Social Psychology*, 39, 186-200. <https://doi.org/10.1037//0022-3514.39.2.186>
- Weiner, B., Osborne, D., & Rudolph, U. (2011). An attributional analysis of reactions to poverty: The political ideology of the giver and the perceived morality of the receiver. *Personality and Social Psychology Review*, 15, 199-213. <https://doi.org/10.1177/1088868310387615>
- Weiner, B., Perry, R. P., & Magnusson, J. (1988). An attributional analysis of reactions to stigmas. *Journal of Personality and Social Psychology*, 55, 738-748. <https://doi.org/10.1037//0022-3514.55.5.738>
- Yzerbyt, V., Castano, E., Leyens, J. P., & Paladino, M. P. (2000). The primacy of the ingroup: The interplay of entitativity and identification. *European Review of Social Psychology*, 11, 257-295. <https://doi.org/10.1080/14792772043000059>
- Zick, A., Krause, D., & Küpper, B. (2016). *Gespaltene Mitte - Feindselige Zustände: Rechtspopulistische und rechtsextreme Einstellungen in Deutschland* [Right-wing populist and right-wing extremist attitudes in Germany]. Berlin, Germany: Dietz.