In Search of Experimental Evidence for Secondary Antisemitism :
A File Drawer Report

Roland Imhoff

Johannes Gutenberg University Mainz, Germany; Social Cognition Center Cologne, Germany

Mario Messer

Social Cognition Center Cologne, Germany

In 1955, Adorno attributed antisemitic sentiments voiced by Germans to a paradox projection: The only latently experienced feelings of guilt were warded off by antisemitic defense mechanisms. Similar predictions of increases in antisemitic prejudice in response to increased Holocaust salience follow from other theoretical apparatuses (e.g., social identity theory as well as just-world theory). Based on the – to the best of our knowledge – only experimental evidence for such an effect (published in Psychological Science in 2009), the present research reports a series of studies originally conducted to better understand the contribution of the different assumed mechanisms. In light of a failure to replicate the basic effect, however, the studies shifted to an effort to demonstrate the basic process. We report all studies our lab has conducted on the issue. Overall, the data did not provide any evidence for the original effect. In addition to the obvious possibility of an original false positive, we speculate what might be responsible for this conceptual replication failure.

Keywords: file drawer; secondary antisemitism; victim blaming; guilt defense; replication

Back in 2007, we conducted an experimental study to test the widespread notion that ongoing reminders of Jewish suffering due to Nazi crimes will evoke some kind of prejudicial reaction in Germans, a defensive "secondary" antisemitism. The (in hindsight severely underpowered) study "worked" perfectly: Reminding German participants of ongoing Jewish suffering led to an increase in antisemitism (compared to baseline), but only if they felt that untruthful (but socially desirable)

responding was futile as we would detect such lies. All built-in validity checks almost made perfect sense. We had never seen such a pretty data pattern before (and never thereafter) and were very happy when others agreed and the paper got accepted for publication in Psychological Science (Imhoff & Banse, 2009). Fueled by this success, we applied for and received a grant to explore this fascinating effect in more detail. The original plan to infer the underlying theoretical process by identifying moderators and mediators failed, however, as we could not even replicate the basic effect. The following is the tale of a long series of (mostly conceptual) non-replications. We will summarize the theoretical background of our original study, explain the goals we had with an expansion of the line of research, and describe a total of eight studies intended to replicate and

expand the original findings (Studies 1 and 2) or empirically address the failure to replicate the basic finding (Studies 3a to 5).

The notion of secondary antisemitism is a highly popular concept across several disciplines. Although there are nuances in how exactly it was conceptualized, most definitions encapsulate the idea of an antisemitism not despite but because of the Holocaust. Briefly after World War II (WWII) and the Nazi's efforts to literally annihilate Jews all over Europe, Peter Schönbach (1961) observed remarkable levels of antisemitism in German youths. This seemed to be puzzling as the now widespread awareness of the antisemitic atrocities committed only a few years earlier should have served as a potent warning sign against all forms of antisemitism. He thus proposed that the adolescents knew about their parents' complicity (guilt by either action or omission) in the actions of the Nazi regime and had to somehow cope with this knowledge or – psychologically speaking – the experienced dissonance of loving their parents but associating them with such horrific actions. To do so, they – according to Schönbach – were more or less forced to rewarm the Nazi regime's antisemitic propaganda to generate justifications for their parents' demeanor. Adorno (1955) made similar observations in his interpretation of group discussions organized by the Frankfurt Institute for Social Research and his explanation was also similar: The participating adults, so he argued, had feelings of latent guilt for what happened during the Holocaust and had to – psycho-dynamically speaking – project this guilt onto the victims (Jews) to alleviate these feelings. Although this version of antisemitism as a defense mechanism is the most common interpretation of Adorno's reasoning (as also reflected in synonyms like "Schuldabwehrantisemitismus", a defense-against-guilt-antisemitism; Bergmann, 2006), Adorno's writing also point to another explanation (that he never explicates as an alternative mechanism): an identity management account. Over the years, these identity concerns moved to the core of current understanding of secondary antisemitism as an antisemitism borne out of the outrage that Jews' insistence on remembering what happened spoils the positive identity of being German. This has been most famously coined in a quip (ascribed to Israeli psychoanalyst Zvi Rex): "The Germans will never forgive the Jews for Auschwitz" (Buruma, 2003).

Of course, this mechanism is not only a well-established figure in the political arena but makes a lot of sense against the background of a plethora of psychological theories. Blaming innocent victims is a central aspect of just-world theory (Lerner, 1980), whereby construing victims as negative and undeserving helps to uphold the illusion that the world is a just place (Correia & Vala, 2003; Friedman & Austin, 1978). Likewise, from a social identity perspective, derogating outgroup victims is functional to attenuate threats to the moral value of the ingroup (Branscombe, Schmitt, & Schiffhauer, 2007; Castano & Giner-Sorolla, 2006). System Justification Theory (Jost, Banaji, & Nosek, 2004) integrates many of these tenets to postulate that rationalizing the status quo (e.g., the ongoing suffering of Holocaust victims by finding fault in their character) may help reduce guilt, dissonance, and discomfort (Jost & Hunyady, 2002).

Despite these many theoretical lines allowing the same prediction, the very core idea of secondary antisemitism had never been experimentally tested. Existing work on the issue was predominantly non-psychological and based on secondary antisemitism as a rhetoric rather than a process. These studies invited respondents to indicate their agreement with statements that encapsulated what researchers understood as secondary antisemitism. Prominent examples are items like "Jews should stop complaining about what happened to them in Nazi Germany" (Selznick & Steinberg, 1969), "The Jews exploit remembrance of the Holocaust for their own benefit" (Heitmeyer, 2006), or "I am tired of continuously hearing about German crimes against Jews" (Bergmann, 2006). Although such utterances may well reflect what has been conceptualized as secondary antisemitism, agreement with them is not indicative of the underlying process. It is, for instance, conceivable that a respondent just dislikes Jews in general, without any specific emphasis on the Holocaust. This respondent will certainly agree with these statements as they communicate the negativity he or she sees in Jews, but this agreement will not be the result of the need to alleviate guilt or defend one's ingroup's moral value. In fact, the very same argument could be made regarding the original participants in the studies by the Frankfurt Institute for Social Research. Maybe they were antisemitic during WWII and continued to be antisemitic thereafter without any indirect mediation via latent guilt or the need to justify their parents. The fact that subscales tapping into agreement with traditional forms of antisemitism (e.g., "Jews have too much power and influence in this world"; Weil, 1985) and secondary antisemitism correlate up to $r = .84$ at a latent level (Imhoff, 2010) adds further fuel to this fire.

We thus aimed to provide experimental evidence for secondary antisemitism as a process rather than a rhetoric. As a way to induce feelings of (collective) guilt or uneasiness about German atrocities, we aimed to make

Holocaust victims' ongoing suffering salient with the expectation that the salience should increase antisemitism as a form of victim derogation (to alleviate guilt or see the world as just or one's group as moral). Something about this prediction, however, did not feel right. Clearly, telling people how much a certain group suffers should somehow increase the threshold to devalue the group, as suffering is expected to evoke sympathy (Heider, 1958) rather than derogation. We aimed to resolve this by reaching into the bag of tricks of social psychologists: Maybe people did have this sentiment but did not express it because social norms prevented them from doing so. So all we needed was a way to block the influence of such norms. If people had a feeling that we could know what they actually felt, then socially desirable (but dishonest) responding was futile since we would not only find out about their prejudice anyway, but would also see that they are liars (a double norm violation). This sums up the logic of bogus pipeline procedures, which allegedly detect dishonest responding and thus lead participants to respond truthfully to avoid the double norm violation described above.

So, this was how we proceeded: We asked as many of our undergraduate psychology students as we could find (a whopping 70 participants) to indicate their agreement with 29 statements of antisemitism as part of a larger paper-and-pencil test (your infamous "mass testing"). Three months later, they were invited to participate in individual testing sessions and 63 of them agreed and showed up for an experiment involving two independent variables manipulated between subjects: Was the suffering of Holocaust victims described as having ongoing negative consequences for them and their descendants (Ongoing Suffering: yes/no)? Were participants hooked up to (slightly outdated) EEG machinery and a hand palm electrode with the information that this would help us detect untruthful responding (Bogus Pipeline: yes/no)? Afterwards, participants wrote down all the thoughts they had while reading the text, then completed a measure of implicit antisemitism, the same antisemitism scale as three months earlier, and a manipulation check item to make sure that they had indeed read the initial text ("Please briefly recall the introductory text. Did it mention ongoing consequences for the victims?").

When we finally looked at the results, they were beautiful – everything looked exactly as it "should". We had an unexpectedly large number of failed manipulation checks (15 people), but the pattern made perfect sense (in hindsight): Almost all of these wrong responses came from the ongoing suffering conditions (13

people). Thus, instead of derogating the victims to alleviate guilt, they just refused to even take note of the ongoing suffering. The remaining 48 participants, however, showed exactly the pattern we expected (Figure 2, left panel). Without mention of ongoing suffering, the level of antisemitism stayed more or less the same (operationalized as standardized residuals of predicting Time 2 antisemitism from Time 1 antisemitism; $r = .89$). Mentioning ongoing suffering, however, decreased the expression of antisemitic prejudice in the control condition but led to an increase when attached to a bogus pipeline. The results were even significant despite the small sample, but clearly, the strategy of controlling for baseline antisemitism made our measure very sensitive. There were more details in the data that added to the picture of a perfect study: The correlation between implicit and explicit antisemitism was independently moderated by the bogus pipeline condition and a Time 1 measurement of the motivation to control prejudiced reactions (Banse & Gawronski, 2003), further validating the experimental procedure and the data in general.

Presenting this study at conferences in the following months was awarded with a lot of positive feedback that boosted our confidence to reach high with this one: We submitted to Psychological Science and received the happy news roughly 11 weeks later: "In both its subject matter as in its empirical approach, your paper is (in my humble opinion) a prototypical Psychological Science paper: It reports on a phenomenon that many people think or have heard about but does so in a way that makes this phenomenon more worthwhile, more important, and much more consequential than lay psychology would have predicted." Sure, the reviewers still had critical comments; None, however, referred to sample size. We resubmitted the manuscript within 10 days and it was accepted shortly thereafter.

In light of the positive feedback we got, it seemed only logical to follow up on this line of research. The many theoretical lines that converged in predicting the effect we found were a plus in making a convincing argument. On the flipside, however, this also meant that we had not one but several candidates for the underlying psychological process responsible for this mechanism. Our project sought to tackle this. Specifically, we expected three distinct, not necessarily mutually exclusive, processes to be potentially involved (Figure 1). Building on originally psycho-dynamic reasoning, we reasoned that the mediating mechanism rested on the process that (latent) feelings of guilt that were fought off by derogating the victims and or interpreting their suffering as deserved. The implication would be that

this mechanism should be restricted to victims of one's own group (as feeling guilty for atrocities committed by another seemed unlikely), should be moderated by propensity to feel guilty, mediated via feelings of guilt, and should be reduced if this guilt was alleviated in any other way.

The second alternative was built on the notion of social identity and individuals' motivation to see their own group as moral (Branscombe, Ellemers, Spears, & Doosje, 1999) and defend its positive identity (Branscombe, Schmitt, & Schiffhauer, 2007). Here also the effect should be restricted to victims of the ingroup (as there exists no motivation to see outgroups as moral) and should be particularly prominent among people who identify (defensively) with their ingroup. The mediating mechanism would be the perceived threat to the ingroup's moral image and any alternative means to repair this image might reduce the effect.

The final distinct possibility was that victim derogation here was a means to restore one's illusion of the world as a just place (e.g., Correia & Vala, 2003; Friedman & Austin, 1978; Godfrey & Lowe, 1975; Lerner & Simmons, 1966; Miller, 1977; Simmons & Piliavin, 1972). The strong need to see the world as a place where everyone gets what they deserve and deserves what they get (Lerner, 1980) should prompt the desire to generate reasons why Jewish suffering was actually deserved, likely leading to victim blaming. Importantly, this mechanism is not exclusive to one's own victim but should be a general process independent of who brought about the suffering. People with a greater need to see the world as just should be more prone to show the effect and re-establishing a sense of the world as just by alternative means should reduce the effect.
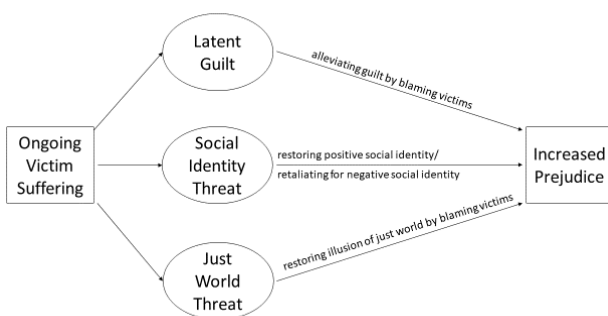

Figure 1. Potential pathways from perception of ongoing victim suffering to increased prejudice.

The present research.

We planned a research program that sought to replicate the basic finding of secondary antisemitism and address the plausibility of each of the three theoretical possibilities outlined above by three strategies. First, all three accounts propose different moderators for the effect: guilt proneness, defensive national identification, just-world beliefs. Second, the boundary conditions of the effect should also be informative. Whereas the first two accounts would predict the effect to be limited to victims of the ingroup, the last would make a general prediction for any (innocent) victim. Third, all three theories allow predictions of the specific kind of alternative means that could serve as an alternative means to alleviate the discomforting feelings of guilt, ingroup threat, or just-world threat. Washing one's hands, we reasoned, should alleviate guilt; re-affirming the morality of one's nation should alleviate concerns about one's group's morality; and providing examples of fair and just procedures should re-establish a sense of justice in this world. As an additional possibility, we planned to explore indirect effects via measured mediators (e.g., latent guilt). Below we describe the first two studies from that line of research, which could not even establish the basic effect let alone a moderation. In light of this, we refrained from conducting additional studies with experimental moderators (e.g., washing hands). Instead, all other reported studies describe efforts to find evidence for the basic process of an increase in antisemitic prejudice by making the history of the Holocaust salient (not necessarily ongoing victim suffering). We employed more subtle measures of prejudice (Studies 3a-3c), less egalitarian samples (Studies 4a and 4b), or more modest forms of negativity, like reduced empathy (Study 5). None of these succeeded in providing such evidence.

Study 1

In the first study, we aimed to replicate Imhoff and Banse's (2009) study and to test the role of latent guilt as a potential mediating process. We utilized an adaptation of the Implicit Positive and Negative Affect Test (IPANAT; Quirin, Kazén, & Kuhl, 2009), which served as an indirect measure of guilt. We examined whether a) ongoing Jewish suffering increases implicit guilt, whether b) implicit guilt is positively correlated with antisemitism under bogus-pipeline conditions, and whether c) implicit guilt mediates the effect of ongoing Jewish suffering on antisemitism. To maximize our chances of finding subtle effects, we took an earlier baseline measurement of our central dependent variable.

Method

Participants. An a priori power analysis suggested a required sample of $N = 120$ to find an interaction of a size of $f = .30$ (effect size was $f = 0.36$ in Imhoff & Banse, 2009) with 90% power. As we expected substantial dropout, we sought to oversample at *t1*. Specifically, we circulated an invitation to participate in a study consisting of two parts (45-minute online study, 15-minute lab experiment) via an e-mail to individuals who had signed up as interested in study participation. To enhance participation at both measurement times, we offered 12 EUR that would be given in cash after completion of the second, lab-based experiment. Despite this incentive and three invitation e-mails, only 109 individuals (34 men, 74 women, 1 missing; mean age: 27.05, $SD = 6.70$) participated in the online study. Upon completion (roughly 3 months after the first invitation to the online study), participants were contacted individually to make appointments for the lab study. A total of 83 participants (29 men, 54 women; mean age: 27.71, $SD = 7.22$; drop-out: 23.9%) were successfully recruited to show up for the lab study. This equipped us with 77% power to detect the estimated effect of $f = .30$. The data of one additional participant in the lab study had to be excluded because he or she provided a participant code not included in the dataset of the pretest.

Online testing.

The purpose of the online test was twofold. First, we needed a baseline measure of antisemitism to control for a t2. This would reduce the noise due to stable individual differences and thus isolate the proportion of the variance that was not due to such individual differences and was therefore in principle susceptible to experimental manipulation. Second, we included a long list of moderators predicted by the different theoretical models outlined above. The overarching goal was to identify systematic patterns across a series of studies to bolster the robustness of one specific theoretical approach. Specifically, we included measures of guilt proneness, national identification, and just-world beliefs. Some additional measures were added on a purely exploratory base.

Antisemitism. Explicit antisemitism was assessed using Imhoff's (2010) scale for the measurement of primary and secondary antisemitism on seven-point scales ranging from 1 (totally disagree) to 7 (totally agree). In order to attenuate reactance and as in the original study, these items were preceded by a filler item ("I think the relationship between Germans and Jews is still influenced by the past."). Additionally, among the clearly negative items we included items that indicated more positive attitudes (e.g., 9 items tapping into collective guilt and regret; Imhoff, Bilewicz, & Erb, 2012; 5 items on contact and contact intention, 5 items on reparation intentions). The actual antisemitism scale consisted of 29 items measuring modern antisemitism (e.g., "Jews have too much influence on public opinion"; 4 reverse-coded; Cronbach's $\alpha = .91$).

As a second measurement approach, participants indicated how warm (5 items, e.g. "good-natured", Cronbach's $\alpha = .92$) and competent (4 items, e.g. "competent", Cronbach's $\alpha = .77$; Fiske, Cuddy, Glick, & Xu, 2002) they perceived Jews to be using a list of 20 adjectives (including 11 filler items) on the same scale.

Guilt proneness. We assessed disposition to experience strong feelings of guilt using two instruments: the Test of Self-Conscious Affect-3 (TOSCA-3; German version by Rüsch & Brück, 2003; 5-point scale) and the Guilt and Shame Proneness Scale (GASP; German translation by Cohen, Wolf, Panter, & Insko, 2011; 7-point scale). Both measures ask participants to imagine various scenarios and to indicate how likely it is for them to experience guilt (among other possible reactions) in these situations. Cronbach's α was .47 for the TOSCA-3 guilt scale and .60 for the guilt – negative behavior evaluation scale of the GASP.

National identification. National identification was measured in two ways so that the impact of the defense form of national identification (i.e., glorification controlled for attachment, collective narcissism) could be isolated. We measured attachment to the national group (8 items; e.g., "Being a German is an important part of my identity"; Cronbach's $\alpha = .90$) and glorification of this group (8 items; e.g., "Germany is better than other nations in all respects"; Cronbach's $\alpha = .82$) on seven-point scales ranging from 1 (totally disagree) to 7 (totally agree) with items by Roccas, Sagiv, Halevy, and Eidelson (2008) that were adapted and translated to German. As an additional measure of defensive national identification, we included a measure of collective narcissism, the exaggerated belief that one's own national group is superior to other groups, on the same scale. To this end we used the German translation of nine items (Cronbach's $\alpha = .85$) of the Collective Narcissism Scale (e.g., "I wish other groups would more quickly recognize authority of the Germans"; Golec de Zavala, Cichocka, Eidelson, & Jayawickreme, 2009).

Belief in a just world. We used Dalbert's (2001) General Belief in a Just World Scale that consists of six items (e.g., "I think basically the world is a just place"; Cronbach's $\alpha = .72$). The items of this scale were answered on a six-point scale ranging from 0 (totally disagree) to 5 (totally agree).

Additional variables. We measured right-wing authoritarianism (RWA; Funke, 2005), social dominance orientation (SDO; von Collani, 2002), the Big Five (BFI-10; Rammstedt & John, 2007), conspiracy mentality (Imhoff & Bruder, 2014), and the coping modes vigilance and cognitive avoidance (Mainz Coping Inventory, ABI; Egloff & Krohne, 1998) using German versions of the scales.

Procedure. After giving informed consent, participants completed all scales in a fixed order (TOSCA-3, Belief in a Just World, Collective Narcissism, Glorification and Attachment, Antisemitism, Conspiracy Mentality, Right-Wing Authoritarian, Social Dominance Orientation, Mainz Coping Inventory, GASP, BFI-10, demographics) before generating the individual code needed to match their pretest data with the lab study data.

Lab Study.

All participants who participated in the online study and left contact details were invited via e-mail to participate in the lab study. Upon arriving at individually arranged sessions they were randomly assigned to one of the four conditions resulting from a 2 (ongoing consequence: yes vs. no) by 2 (bogus pipeline: yes vs. no) design.

Information on ongoing consequences. Participants read a text, ostensibly taken from a history book, which described the German atrocities committed against Jews in the Auschwitz concentration camp. This text was identical to that used by Imhoff and Banse (2009). The last paragraph contained the manipulation of ongoing consequences. Participants either read that the suffering of the Jewish victims was part of a terrible history that has no direct implications for Jews today (no ongoing consequences) or that even today Jews are suffering either as Auschwitz survivors or as their descendants because of "secondary traumatization" (ongoing consequences).

Bogus Pipeline. The implementation of the bogus pipeline differed from the original study (Imhoff & Banse, 2009) because we initially intended to explore physiological reactions to both versions of the text about the Holocaust. In the bogus pipeline condition, the electrode belt of a heart rate monitor watch was applied to participants' chests. In addition, electrodes were attached to the palmar surfaces of the participants' index and middle fingers and to the back of their hands, supposedly to measure galvanic skin response. Participants were informed that physiological data were measured because "previous research has shown that we can detect quite well whether someone answers truthfully or with a lie". Participants in the control condition underwent measurement of heart rate as well but did not have electrodes attached to their hands. Importantly, participants in this condition were informed that physiological measures were obtained merely in order to explore whether physiological parameters correlate with information processing in reading.

Measures.

Implicit guilt. We used an adaptation of the Implicit Positive and Negative Affect Test (IPANAT; Quirin, Kazén, & Kuhl, 2009) that assesses anger, fear, happiness, and guilt (IPANAT-4-EM) to measure implicit guilt. Participants were asked to judge the extent to which artificial words (e.g., "VIKES") express each of three emotional qualities per emotion cluster. Guilt was represented by the emotion words "guilt", "regret", and "shame", Cronbach's α = .88.

Explicit guilt. The same emotions that were measured with the IPANAT-4-EM in an indirect way were also assessed using a self-report measure. Participants indicated to what extent they felt anger, fear, happiness, and guilt ("guilty", "regretful", and "ashamed") at that moment, Cronbach's α = .81.

Antisemitism. Participants completed the same scale as in the online study, α = .93.

Heart rate variability. We collected heart rate variability data for exploratory purposes using heart rate monitor watches by Polar.

Procedure. After an interval ranging between seven days and three months between the online survey and participation in the lab study (Time 2), participants were randomly assigned to one of four experimental conditions in a 2 (ongoing consequences vs. no ongoing consequences) × 2 (bogus pipeline vs. control) factorial design. The session started with the bogus pipeline manipulation and the physiological device set up. After a two-minute baseline measurement of heart rate variability, participants read a text about the German atrocities in the Auschwitz concentration camp, which included the manipulation of consequences for present-day Jews. The individual paragraphs of the text moved across the screen over a period of 140 seconds to allow for a mapping of physiological reactions to specific parts of the text. After the reading task, participants were asked to write down on a piece of paper the thoughts they had had while reading the text. Subsequently, they completed the IPANAT-4-EM, which included our measure of implicit guilt, and filled in the measure of explicit guilt. Finally, they again answered the same antisemitism questionnaire that they had completed at Time 1

and indicated whether the text presented to them before contained information about ongoing negative consequences for Jews today as a manipulation check ("yes" or "no").

Results

Antisemitism showed high stability between both measurements, $r(83) = .89$, $p < .001$. We followed the strategy of the original study (Imhoff & Banse, 2009) in analyzing the effect of the information on ongoing consequences on antisemitism. Time 1 antisemitism scores were entered as a predictor of Time 2 antisemitism scores in a regression analysis and standardized residual change scores were used as an index of change in antisemitism. The resulting residual change scores were subjected to a 2 (ongoing consequences vs. no ongoing consequences) × 2 (bogus pipeline vs. control) analysis of variance (ANOVA). In contrast to our hypothesis and the results of the original study, no evidence was found for an interaction between the information on ongoing negative consequences for Jews and the bogus pipeline manipulation, $F(1, 79) = 0.28$, $p = .602$, $\eta_p^2 = 0.003$ (Figure 2, right panel). Likewise, none of the experimental factors showed a main effect, $Fs < 1$. Confronting German participants with ongoing negative consequences for present-day Jews did not result in increased antisemitism, even when participants thought that untruthful responses could be detected by the experimenter.
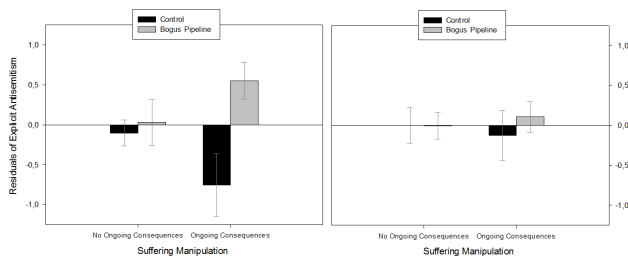

Figure 2. Change in explicit antisemitism (standardized residuals) from Time 1 to Time 2 as a function of the information on ongoing consequences and bogus pipeline manipulations in the original study (Imhoff & Banse, 2009; left panel) and in Study 1 of the current research. Error bars represent standard errors of the mean.

Despite this lack of support for the basic effect, we analyzed whether ongoing Jewish suffering increases implicit guilt. A t-test for independent samples revealed no significant difference in implicit guilt between the ongoing consequences condition ($M = 3.25$, $SD = 0.95$) and the no ongoing consequences condition ($M = 3.06$, $SD = 0.90$), $t(81) = 0.93$, $p = .354$, Hedges's $g_s = 0.20$, 95% CI [-0.23, 0.64]. In contrast to our hypothesis, implicit guilt was not positively correlated with antisemitism under bogus pipeline conditions, $r(44) = .12$, $p = .451$. In order to test the moderator hypotheses, we performed separate hierarchical multiple regression analyses using the standardized residual change scores in antisemitism as a dependent variable. Product terms representing the three-way interactions among both experimental factors and the potential moderator variables were entered as predictors in a third step after the simple predictors and all possible two-way products. None of these regression analyses revealed evidence for a moderation effect of collective narcissism (see Table Osm.1 on our OSF project page), national glorification (see Table Osm.2), just-world beliefs (see Table Osm.3), or guilt proneness (see Table Osm.4).

Discussion

Study 1 provided no lead on the research question of which psychological processes are plausibly responsible for increased prejudice in light of ongoing suffering, predominantly because it failed to replicate this finding. Although descriptively the mean scores were in the predicted direction, this trend was far from significant. Several reasons appeared conceivable for this. As always, the non-significant findings could be a false-negative and due to too little power. We failed to collect data from 120 participants as planned based on a priori power analyses and these analyses might already have been biased by an effect size estimate that was too optimistic, taken from the original study. Alternatively, the bogus pipeline manipulation might not have worked as it did in the original study. We had used different equipment (a heart rate monitor plus hand electrodes instead of forehead electrodes plus hand electrodes) in a different setting (neutral, almost empty room instead of a slightly messy laboratory with many cables lying around) and sampled from a different population (via a volunteer participant e-mail list instead of first-year undergraduates) with different incentives (cash payment instead of course credit). Potentially, any of these factors or their combination undermined the credibility of our bogus pipeline manipulation. In fact, unlike the previous study, we have no evidence for the validity of the procedure. In our original study, we had included an Affective Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005) as a measure of implicit antisemitism. As we expected, this measure correlated substantially with the explicit measure under bogus pipeline conditions (i.e., participants really self-report what

they "feel"), but not under control condition (where they corrected their responses in a socially desirable way). We had eliminated the indirect measure between the ongoing suffering manipulation and the dependent variable in an effort to streamline the procedure. Nevertheless, we continued as planned with Study 2.

Study 2

In Study 2, we aimed to test the just-world theory as an explanation of the effect of ongoing Jewish suffering against the hypotheses of guilt-defense and the protection of a positive social identity. We did so by introducing a condition in which just-world theory would make a different prediction than guilt-defense or social identity theory. Just-world beliefs should be threatened by unjustly suffering victims in any case, irrespective of who the perpetrator is. In contrast, not every case of injustice should result in increased guilt or in a threatened positive social identity. Only if the perpetrators are members of the in-group (in this case Germans), one should be motivated to derogate the victims. Accordingly, we manipulated group membership of the perpetrators.

Method

Participants. We again aimed for a final sample of 120 participants. One hundred and eighty-five first-year psychology students (27 men, 158 women; mean age: 22.29, SD = 4.88) from the University of Cologne, Germany, participated in an online study at the first measurement. Seventy-eight participants dropped out between the first and the second measurement occasion (42%). The post-test data of two participants had to be excluded because they provided participant codes not included in the dataset of the pretest. We excluded nine participants before running the analyses because they did not remember that the historical text they had read contained information about ongoing negative consequences for the victims or because they did not remember who the perpetrators had been. The remaining sample of N = 96 (86 women, 10 men) ranged from 17 to 39 in age (M = 21.55, SD = 4.11). Participants received 12 EUR for their participation (approx. 7.50 EUR per hour).

Measures. The main dependent variable in Study 2 was explicit prejudice against the victim group participants read about during the experiment. Depending on experimental condition, participants responded to items measuring prejudice against Jews or Chinese. We chose ten items from the antisemitism scale (Imhoff, 2010;

Cronbach's α = .72) that could be modified to assess prejudice against Chinese (e.g., "Chinese have too much influence on public opinion"; Cronbach's α = .80). The prejudice items were supplemented by four items on collective guilt (e.g., "I can easily feel guilty for the negative consequences that were brought about by Germans [Japanese]", Cronbach's α = .83 and .62, respectively) and, for participants that read about the Holocaust, by two items on primary and five items on secondary antisemitism for exploratory purposes. Study 2 included the same measures of potential moderators and additional variables as in Study 1 except that we excluded the TOSCA-3 (but kept the GASP as a measure of guilt proneness), the in-group attachment and glorification scales (but kept the measure of collective narcissism), and the ABI (measuring anxiety coping styles that might be related to the tendency to avoid – and therefore misremember – threatening information). In addition, we included the following measures on a purely exploratory basis: a response latency-based measure of prejudice (adapted from Vala, Pereira, Eugênio, Lima, & Leyens, 2012), a rating of Jews and Chinese on eight warmth-related traits, and a feeling thermometer assessing feelings towards these groups (among other groups). The BIOPAC system that had the main purpose of serving as a bogus pipeline setup (see below) was also used to record electrodermal activity data for exploratory purposes. We did not analyze physiological data, but the raw data can be obtained from the authors.

Independent variables. We manipulated group membership of the perpetrators by presenting participants with either a text about the Holocaust (which was the same as in Study 1) or about the ongoing suffering of Chinese victims of the Nanking massacre committed by Japanese troops. In both conditions, the last paragraph stressed the ongoing negative consequences for present-day Jews or Chinese, respectively. Presentation of the text differed from Study 1 in that the whole text was shown on the screen at once, whereas the individual paragraphs moved across the screen in Study 1.

In contrast to Study 1 and more similar to the original study (Imhoff & Banse, 2009), we operationalized the bogus pipeline manipulation as measuring electrodermal activity under the pretext of lie detection vs. no physiological measurement at all. Participants in the bogus pipeline condition were informed that "specific parameters of electrodermal activity allow us to detect whether someone answers truthfully or with a lie". Subsequently, the experimenter attached the electrodes of a BIOPAC system to the palmar surfaces of the participants' index and middle fingers. In order to increase

credibility of the bogus pipeline, the experimenter continued with an alleged calibration that required participants to follow some instructions while the experimenter was monitoring the physiological parameters at another computer behind a room divider. Specifically, participants were instructed to take a deep breath and hold the breath for a moment. After that, the experimenter asked participants to memorize a number between one and six printed on a card (which was a 4 in every case). Analogous to a concealed information test, the experimenter then read a series of numbers that could have been on the card, and participants were instructed to answer "yes" to every number, whether accurate or not. After a few seconds, participants were informed that the apparatus was working properly and that they were ready to start with the study. Participants in the control condition received no treatment at all.

Procedure. The first measurement of explicit prejudice against the victim group alongside assessment of the potential moderator variables was obtained in a classroom testing session (Time 1). After an interval of five or six months, participants were invited to the laboratory for an individual session (Time 2). Participants were randomly assigned to one of four groups in a 2 (group membership of the perpetrators: in-group vs. out-group) × 2 (bogus pipeline vs. control) design. After the bogus pipeline manipulation had been administered, participants gave demographic information and read a neutral text about the history of an abandoned town, which served as a control task for the assessment of electrodermal activity, involving reading but without injustice-related content. In the bogus pipeline condition, this reading task was preceded by a three-minute baseline measurement of electrodermal activity. After this initial reading task, participants were given two minutes to write down on a piece of paper their thoughts about the text. After a one-minute rest period, participants were presented with the critical text that contained the manipulation of the perpetrators' group membership and again wrote down their thoughts. Subsequently, participants completed 48 trials of the response latency-based measure of prejudice and answered the prejudice questionnaire. Finally, they were asked whether the text contained information on ongoing negative consequences for the victims ("yes" or "no") and who the perpetrators had been as a manipulation check ("the Red Army", "Japanese troops", "SS officers", or "American soldiers").

Results

The stability of antisemitism was lower than in Study 1, $r(44) = .57$, $p < .001$. The stability of prejudice against Chinese was $r(52) = .72$, $p < .001$. Prejudice against the victim group was analyzed as change in prejudice between both measurement occasions exactly as in Study 1. The standardized residual change scores were subjected to a 2 (group membership of the perpetrators: in-group vs. out-group) × 2 (bogus pipeline vs. control) ANOVA. Results neither revealed a significant main effect of the bogus pipeline manipulation, which would have been predicted by just-world theory, $F(1, 92) = 0.05$, $p = .830$, $\eta_p^2 = 0.00$, nor an interaction effect, which would have been predicted by guilt-defense and social identity theory, $F(1,92) = 0.01$, $p = .919$, $\eta_p^2 = 0.00$. The only significant experimental effect was a (hard-to-explain) main effect of victim group, $F(1, 92) = 15.74$, $p < .001$, $\eta_p^2 = 0.17$: Whereas antisemitic prejudice showed a relative decrease compared to $t1$, the opposite was true for anti-Chinese prejudice (Figure 3). Separate moderator analyses confirmed this result for participants high in collective narcissism (see Table Osm.5), just-world beliefs (see Table Osm.6), and guilt proneness (see Table Osm.7).
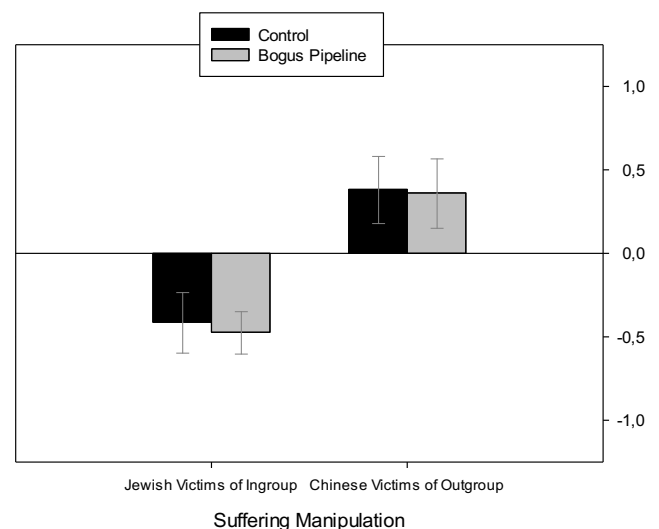

Figure 3. Change in explicit prejudice against the victims (standardized residuals) from Time 1 to Time 2 as a function of the group membership of the perpetrators and the bogus pipeline manipulation in Study 2. Error bars represent standard errors of the mean.

Discussion

Studies 1 and 2 failed to replicate the basic effect of an increase in antisemitism in response to the ongoing

suffering of Jewish victims, which had been reported in the original study (Imhoff & Banse, 2009). In light of this repeated failure to replicate the interaction of bogus pipeline and ongoing suffering, we decided to switch gears and focus on establishing the basic effect. The bogus pipeline manipulation appeared to us as the most plausible candidate for this failure. Clearly, participants needed a lot of trust in the researchers to believe that the researchers could indeed detect untruthful responding. In contrast to the time when bogus pipelines were originally proposed in the early 1970s (e.g., Sigall & Page, 1971), current students are very likely aware of the fact that a simple "lie detector" is a gadget from fictional literature, not a real thing. Based on the working hypothesis that lie detection machines have been too thoroughly debunked in public discourse to affect participants' responding, we turned to another popular approach to circumvent social desirable responding: more subtle measures.

Study 3a – 3c

In Studies 3a to 3c, we investigated whether the very basic effect shown in the original study (Imhoff & Banse, 2009) – Germans show increased antisemitism when confronted with the Holocaust – is detectable. As we were not confident in the effectiveness of the bogus pipeline manipulation given the results of Studies 1 and 2, we employed an alternative approach in addressing the problem of measuring antisemitic attitudes, which are socially very undesirable to express. Instead of a bogus pipeline setup, we adopted a reverse-correlation paradigm as a subtle, indirect measure of prejudice. If confronting Germans with the crimes their ancestors committed against Jews results in them becoming more antisemitic, we expected Germans to remember the face of a Jewish person as more negative when the Holocaust is mentioned at the initial confrontation with this person. To test this hypothesis, we asked participants to form a first impression of a person that was either Jewish or Christian. In addition, we manipulated whether the text about this person contained information about the Holocaust or not. Participants then completed a reverse-correlation image-classification task based on the memory they had of the target person's face, which allowed us to visualize the remembered facial appearance of that person. We replicated this study twice (Studies 3b and 3c) with minor changes regarding the materials, as explained below.

Method

Participants. Seventy-eight psychology students from the University of Cologne, Germany, were recruited via mailing lists, flyers, social networks, or by being personally approached on the university campus to take part in Study 3a. Based on a priori set criteria (see below), we excluded 17 participants before running the analyses because they did not remember correctly that the target person was Jewish [vs. Christian] or that he was volunteering in an organization that supports Holocaust survivors [vs. an organization working to protect forests] or both. The remaining sample of $N = 61$ (47 women and 14 men) ranged from 20 to 49 years in age ($M = 24.67$, $SD = 5.82$). Participants received course credit for their participation.

Roughly 120 students from different fields of study participated in exchange for 4 EUR in Study 3b ($N = 121$) and Study 3c ($N = 120$), respectively. The effective sample size after exclusions based on the same criteria as in Study 3a was $N = 94$ (50 women and 44 men; age 18 to 38 years, $M = 22.71$, $SD = 3.42$) in Study 3b, and $N = 89$ (59 women, 29 men, one participant did not indicate; age 18 to 40 years, $M = 23.22$, $SD = 4.23$) in Study 3c.

Independent Variables. The session started with an impression formation task that contained the manipulation of both independent variables. Participants read a short text about a person containing irrelevant information about that person's job, residence, and leisure time, and, critically, cues to the person's religious affiliation and a sentence mentioning the Holocaust or a control issue. Participants were told that the person was active in his synagogue [vs. church] and volunteered with an organization that helps Holocaust survivors because his grandfather had been murdered in the Auschwitz concentration camp [vs. an organization working to protect forests]. In Studies 3b and 3c, we introduced minor changes in the manipulations. Specifically, we reasoned that volunteer work in any religious group might be seen as a cue to morality or other positive traits. In Studies 3b and 3c, religious affiliation was thus made salient without implying volunteer work: The sentence containing the manipulation of group membership was changed so that the target person was not active in a synagogue or church but had been asked whether he wanted to become active in his father's synagogue [vs. church]. Participants in the Holocaust condition read that the target person was involved in an organization demanding reparation payments for Holocaust survivors (whereas he was working for another charity not related to the Holocaust in the other condi-

tion). Contrary to Study 3a, the text contained no information about any victims among his family members to eliminate potential effect of direct sympathy. In each of the three versions of Study 3, the text about the target person was accompanied by a picture showing the face of a young man. In Studies 3a and 3b, the face image was the neutral male face of the Averaged Karolinska Directed Emotional Faces database (Lundqvist & Litton, 1998), whereas we used a morph of sixteen emotionally neutral faces in frontal view taken from the Radboud Faces Database (Langner et al., 2010) in Study 3c. Both images have been used in previous reverse-correlation research (e.g., Dotsch et al, 2008, and Imhoff et al., 2013, respectively).

Central Dependent Variable: Reverse-Correlation Image-Classification Task. We relied on reverse correlations to assess whether participants' memory of a person's face is biased by information on that person's group membership and mention of the Holocaust. Reverse correlation is a data-driven approach that enables researchers to visualize an idealized decision criterion. By tracking which kind of subtle (and random) alterations in the appearance of face correlates with a classification decision (e.g., which of two faces look more female; Mangini & Biederman, 2004), one can estimate what a face that fulfills all criteria in an ideal way looks like (classification image). Beyond very basic decisions (e.g., male vs. female), and more relevant to this study, reverse-correlation techniques can be used to construct images that reflect the expected or remembered facial appearance of a target person without making any a priori assumptions about relevant features.

Previous studies applied this approach to investigate biased expected facial appearance of out-group members (Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Dotsch, Wigboldus, & van Knippenberg, 2013; Imhoff & Dotsch, 2013; Imhoff, Dotsch, Bianchi, Banse, & Wigboldus, 2011) and previously encountered individuals (Karremans, Dotsch, & Corneille, 2011). For instance, Karremans et al. (2011) found that people involved in a romantic relationship held a less attractive memory of an attractive alternative's face than uninvolved individuals. When asked to select a face that best represents a typical member of a certain social group (e.g., manager, nursery teacher), stereotypical beliefs about these groups' warmth as well as competence are encoded in the face and can be decoded from the classification image by independent perceivers (Imhoff, Woelki, Hanke, & Dotsch, 2013).

Image Creation. Subsequently, participants worked through the reverse-correlation task, which allowed us to obtain visualizations of the participants' memories of the target face. We used a two-images, forced-choice variant of the reverse-correlation paradigm (e.g., Dotsch et al., 2008; Imhoff et al., 2011), in which each participant completed 400 trials of selecting one of two presented faces. In each of these trials, they selected the face that they thought looked more like the target person they had seen before (i.e., during the impression formation task). The stimuli used in the picture classification task were all based on the face they had seen on the page about the target person. To generate the stimuli, this base image had been converted to grayscale and superimposed with random noise resulting in random variations of the facial appearance between the stimuli (for noise generation, see Dotsch & Todorov, 2011). Every trial employed a different noise pattern displaying the original pattern on the left and the negative of that pattern on the right side of the screen. Participants selected pictures by pressing a left or right button on the keyboard.

By averaging all noise patterns participants had selected separately for each experimental condition and superimposing these classification patterns on the base image, we obtained a classification image for every condition (see Figure 4). Trials with a response time lower than 200ms were excluded before constructing the classification images (<5% of the trials). The resulting classification images visualized how participants in each of the four experimental groups remembered the target face on average. In addition to the classification images aggregated on a group level, we also analyzed classification images of individual participants in Studies 3b and 3c in order to explore the possibility that derogation of victims could occur on inter-individually different dimensions and hence be reflected in different facial features.

Holocaust Mentioned	Control

Jewish Target

Christian Target

Figure 4. Classification Image as a function of information about the Holocaust (Holocaust is mentioned vs. control) and group membership of the target person (Jewish vs. Christian) in Study 3a.

Image Rating. In the second phase of Study 3, the classification images created by every experimental group in the first phase were rated on warmth (Cronbach's α between .84 and .92) and competence (Cronbach's α between .72 and .90) by 56 independent participants recruited via Amazon MechanicalTurk (MTurk; 30 women and 26 men, age 18 to 75 years, $M = 38.57$, $SD = 14.59$; Study 3b: $N = 43$, 20 women and 23 men, age 20 to 66 years, $M = 37.93$, $SD = 13.04$; Study 3c: $N = 64$, 40 women, and 23 men, one person did not indicate, age 18 to 71 years, $M = 35.56$, $SD = 12.68$). Five other participants were excluded because they indicated that they had answered randomly or purposely false, or that they would exclude their data if they were the researcher (six exclusions in Study 3b and four in Study 3c). The warmth and competence items were the same as in the first phase of the study. Responses were made using a five-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Every rater in this second phase of the study rated each of the four group-wise classification images. Accordingly, ratings were analyzed using within-subjects tests. The warmth ratings of the classification images constituted the main dependent variable. The individual classification images from the first phases of Studies 3b and 3c were rated by independent participants by indicating "how likable" they found each of the persons. Participants were paid 25 cents in Study 3a and 50 cents in Studies 3b and 3c.

Additional measures. After completing the reverse-correlation task, participants were probed for suspicion using a funneled debriefing procedure (cf. Chartrand & Bargh, 1996) and were then asked to indicate their first impression of the target person by a) describing the person in their own words and b) rating the person's warmth and competence. For the warmth and competence ratings, participants indicated to what extent each of 20 adjectives representing warmth (5 items, e.g. "good-natured", Cronbach's α = .82) and competence

(4 items, e.g. "competent", Cronbach's α = .69; Fiske et al., 2002) characterized the target person on a five-point scale (1 = not at all to 5 = very much). In Studies 3b and 3c, we excluded the question asking participants to describe the target person and replaced the warmth and competence items with ten items assessing likability of the target person (e.g., "How likable do you find David S.?"), which also included five reverse-coded items representing common negative stereotypes about Jews (e.g., "How stingy do you find David S.?"). These ten items were combined into a single explicit likability scale, Cronbach's α = .84 in Study 3b and .88 in Study 3c. Next, participants answered ten (in Studies 3b and 3c, six) questions about the target person of which three served as a manipulation check and gave demographic information. Finally, they completed an antisemitism questionnaire (only in Study 3a) consisting of 14 items taken from the scale used in Study 1 (Imhoff, 2010; Cronbach's α = .86).

In Studies 3b and 3c, we included a word stem completion task to explore whether representations of the Holocaust were successfully activated in the Holocaust condition. This task was administered after the warmth and competence ratings and asked participants to complete 30 word stems of which ten could be completed to form a word related to the Holocaust (e.g., "Endl____" could be completed to "Endlösung" [final solution]). Answers on the ten critical items were coded as Holocaust-related or not by a single rater and aggregated to a sum score. Furthermore, the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) was added in between the impression formation task and the reverse-correlation image-classification task in Studies 3b and 3c for exploratory purposes.

Materials and Procedure. Participants were seated at a computer in individual cubicles and were randomly assigned to one of four experimental conditions following a 2 (group membership of the target person: Jewish vs. Christian) × 2 (Holocaust is mentioned vs. control information) design. Secondary antisemitism, we reasoned, would be exhibited in a face that independent others would perceive as less warm if the person was introduced as Jewish and the Holocaust was mentioned.

Results

Based on the idea of secondary antisemitism, we expected the classification images created by participants who were both presented with a Jewish target person and reminded of the Holocaust to be rated as less warm or likable than those from the other conditions. Warmth ratings of the group-wise classification images were

subjected to a 2 (group membership of the target person: Jewish vs. Christian) × 2 (Holocaust is mentioned vs. control information) repeated measures ANOVA. Contrary to the hypothesis, in Studies 3a and 3b results did not show a significant interaction effect, $F(1, 55) = 0.03$, $p = .872$, $\eta_p^2 = .00$, and $F(1, 42) = 0.02$, $p = .897$, $\eta_p^2 = .00$, respectively. In Study 3c, a significant interaction effect emerged, $F(1, 63) = 10.06$, $p = .002$, $\eta_p^2 = .14$. However, the pattern of means was in contrast to expectations, as the classification image from the Jewish condition was rated as warmer when participants were reminded of the Holocaust (vs. control information).

For the analysis of individual classification images, likability ratings were averaged across raters yielding a mean likability rating for every individual classification image. The likability scores were then submitted to a 2 (group membership of the target person: Jewish vs. Christian) × 2 (Holocaust is mentioned vs. control information) between subjects ANOVA. Neither for Study 3b nor for Study 3c the ANOVAs revealed any differences between experimental conditions, test of interaction effects, $F(1,90) = 0.03$, $p = .871$, $\eta_p^2 = .00$ and $F(1,85) = 0.16$, $p = .693$, $\eta_p^2 = .00$, respectively.

In addition to the primary analyses looking at the classification images reported above, we explored the explicit ratings of the target person's warmth (Study 3a) and likability (Studies 3b and 3c). Between-subjects ANOVAs did not yield an interaction effect in any of the studies, $F(1, 57) = 0.02$, $p = .879$, $\eta_p^2 = .00$ in Study 3a, $F(1, 90) = 2.97$, $p = .088$, $\eta_p^2 = .03$ in Study 3b, and $F(1, 85) = 0.38$, $p = .537$, $\eta_p^2 = .00$ in Study 3c. To explore whether representations of the Holocaust were activated to a higher degree in the conditions mentioning the Holocaust than in the control conditions, we compared the number of Holocaust-related answers in the word stem completion task. In contrast to our expectations, the sum of Holocaust-related answers was not significantly higher in the Holocaust conditions ($M = 2.22$, $SD = 1.74$ in Study 3b and $M = 1.58$, $SD = 1.18$ in Study 3c) than in the control conditions ($M = 1.66$, $SD = 1.58$ and $M = 1.43$, $SD = 1.17$), $t(92) = 1.63$, $p = .108$, Hedges's $g_s = 0.33$, 95% CI [-0.08, 0.74] and $t(87) = 0.59$, $p = .559$, Hedges's $g_s = 0.12$, 95% CI [-0.29, 0.54], respectively.

Discussion

Studies 3a to 3c failed to provide any evidence for the notion that making the Holocaust salient increases participants' need to derogate the victim group. If anything, the effect was in the opposite direction in one study, but not reliably in the other studies. This invites

speculation as to whether the chosen measure is indeed immune to social desirability concerns. Although it is not explicitly an evaluation task, participants are of course free to take all the time they need to select images according to whatever impression they want to convey of themselves (e.g., as particularly unprejudiced). It may thus be that the measure taps into participants' very explicit and elaborate evaluation as much as typical prejudice scales do. The unexpected effect (somewhat reminiscent of the pattern in the no bogus pipeline condition in the original paper) is compatible with this interpretation, but the lack of any effect in the following studies does not corroborate this speculation. At present, there is no consistent effect (in any direction) of making the atrocities of the Holocaust salient.

As perhaps a side effect rather than the focus of the current interest, we also were not able to produce consistent effects on what we perceived as a simple manipulation check: a word stem completion task. The logic was that making the history of the Holocaust salient should increase participants' tendency to complete ambiguous word stems in a semantically consistent way. Such tasks are highly popular instruments in the field of social cognition to tap into the semantic accessibility of certain constructs (or concept activation). While our failure to find any effect in such measures may raise doubts about their validity, it should be noted that the employed task was constructed ad hoc without proper pilot testing of base rates of word completion tendencies. In our own lab, we have gathered experiences with such tasks in other domains (i.e., to what extent pictures of or real pregnant women make baby-related word completions more likely) with more success (Marhenke & Imhoff, 2018). We would thus caution against throwing the baby out with the bath water based on the failure presented here. At the same time, we caution that it is bad practice how naïvely we and other colleagues construct such measures ad hoc and interpreted them as valid as long as they produced the desired effects, but discard them as unreliable and invalid if they do not.

Another reason for the failure to replicate the effect could be the population we sampled from in Studies 1-3. Most were student samples from the University of Cologne, more specifically the School of Humanities with a specialization in special needs education. Students from this school have a reputation to be particularly liberal (and their average self-reported political orientation was left of the scale midpoint in both Studies 1 and 2), whereas students in the original study were psychology students who do not necessarily have the same reputation. To increase our chances of finding support for

the mechanism of secondary antisemitism, we thus changed the research setting to a less restricted sample that might not have egalitarian norms to the same extent. We thus conducted two studies in the city center of Cologne with pedestrians from the general population as participants.

Studies 4a and 4b

To include more politically diverse participants, we recruited individuals walking in front of the main station in Cologne, Germany, to fill in a "short survey on opinions on violent conflicts". Instead of open antisemitic expressions, we used agreement to criticism of Israel as a dependent variable. We assumed that criticism of Israel would be perceived as less taboo and thus be reported openly in a questionnaire so that we would not need a bogus pipeline setup. This approach was built on the notion that not only are anti-Israeli sentiment and antisemitism highly correlated in Europe (Kaplan & Small, 2006), but certain forms of criticism of Israeli politics are construed as a substitute communication. Demonizing Israel is socially more accepted than demonizing Jews (Steinberg, 2004), but – in the context of secondary antisemitism – serves the same purpose: By portraying the (Jewish) state of Israel as ruthless perpetrator of human right violations, the (German) crimes against Jews become less salient (i.e., victim-perpetrator reversal; Imhoff, 2010). In line with the hypothesis of secondary antisemitism, we expected participants to show higher agreement (relative to a control condition) to statements criticizing Israel after being reminded of the Holocaust. This effect might be greater for individuals high in national glorification.

Method

Participants. One hundred passers-by approached in front of the main station of Cologne, Germany, participated in Study 4a (57 women and 43 men). Participants ranged from 17 to 76 years in age ($M = 33.87$, $SD = 14.59$). For Study 4b we recruited 196 passers-by (119 women, 73 men, four did not indicate their gender) ranging from 14 to 63 in age ($M = 27.55$, $SD = 12.03$). Another four participants were excluded before running the analyses because of missing responses on more than 50% of the items of the main dependent variable. In both studies, we included an attention check by asking participants in the last sentence of the instruction to write an X on the page margin. As a very high proportion of participants failed this attention check (45% in Study 4a and 27% in Study 4b), we decided to

keep these participants in the sample. The results reported below do not change when these participants are excluded. Participants received no compensation.

In Study 4b, participants scored higher on national glorification ($M = 2.32$, $SD = 1.22$) than our student sample in Study 1 ($M = 1.72$, $SD = 0.88$), $t(276) = 4.02$, $p < .001$, Hedges's $g_s = 0.52$, 95% CI [0.26, 0.79] using the same three items. Although, a mean of 2.32 is still relatively low on a seven-point scale, our goal to acquire a less liberal sample was achieved.

Materials and Procedure. The study was conducted in summer 2014 during the 2014 Israel-Gaza conflict. Participants were approached by the experimenter and asked to participate in a "short survey on opinions on wars and violent conflicts" (in Study 4b, "on the Israeli-Palestinian conflict"). They were then handed a two-page paper-and-pencil questionnaire and were randomly assigned to one of two experimental conditions in which they were either reminded of the Holocaust or not. The manipulation of being reminded of the Holocaust vs. a control condition was embedded in the instructions of the questionnaire. In Study 4a, participants in the Holocaust condition read that "70 years have passed since the monstrous German crime, the Holocaust. Within 4 years, the Germans systematically killed 6 million Jews in extermination camps like Auschwitz." In the control condition, that first part of the instructions read, "History of humanity is a sequence of wars" making no reference to the Holocaust. Participants then indicated their agreement to 13 statements criticizing Israel (Cronbach's $\alpha = .77$), which had been taken from the existing literature (e.g., "Israel is a state that stops at nothing"; Kempf, 2014). To make the cover story (a survey on wars and violent conflicts) more credible, the questionnaire also included ten items on two other wars, five on the war in Ukraine and five on the war in Syria, which were not analyzed.

In Study 4b, we included a more detailed description of the Holocaust in the Holocaust condition emphasizing that a) most Germans from all parts of German society participated in the genocide or willfully ignored the crimes and that b) Jews are still suffering today as a result of the Holocaust. Besides the text, the Holocaust condition included a picture showing corpses of prisoners of the Buchenwald concentration camp. The information about the Holocaust was introduced by referring to the public discussion in Germany about the role of the Nazi past for the contemporary relations to Israel. The control condition in Study 4b was a baseline measure that simply asked participants to report their opinion on the Israeli-Palestinian conflict. The main dependent variable, criticism of Israel, was assessed using an

18-item scale (as compared to 13 items in Study 4a; Cronbach's α = .84). This scale comprised the same 13 items that had been used in Study 4a and five more items that were newly created on the basis of actual comments on the 2014 Israel-Gaza conflict from the media (e.g., "The war that Israel initiated against Gaza is completely unjustifiable. It is a crime, whether from the air or on the ground.")

To test whether Holocaust reminders increase criticism of Israel especially among Germans who glorify their national group, we included three of the items by Roccas et al. (2008) on national glorification in Study 4b (Cronbach's α = .62). In Study 4b, we also included five items from the antisemitism scale (Imhoff, 2010), four items on group-based shame, three items on national attachment, and, for exploratory purposes, an open-ended question on participants' understanding of German identity.

Results

Neither Study 4a nor Study 4b provided evidence for an increase in criticism of Israel as a reaction to being reminded of the Holocaust, $t(98)$ = -0.29, p = .776, Hedges's g_s = -0.06, 95% CI [-0.45, 0.33] and $t(194)$ = 0.14, p = .890, Hedges's g_s = 0.02, 95% CI [-0.26, 0.30], respectively. Participants who read about the Holocaust did not criticize Israel more (Study 4a: M = 3.02, SD = 0.72; Study 4b: M = 3.78, SD = 0.85) than those in the control condition (Study 4a: M = 3.06, SD = 0.91; Study 4b: M = 3.76, SD = 0.88). This result also held for participants high in national glorification as revealed by a hierarchical multiple regression analysis. In a first step, the effect-coded group variable (Holocaust = 1; control = -1), attachment to Germany, and glorification of Germany were entered as predictors of criticism of Israel, followed by three product terms representing all possible two-way interactions between the predictor variables in a second step, and the three-way interaction in a third step. The interaction between experimental condition and glorification of Germany did not significantly predict criticism of Israel in the full model, β = .05, $t(187)$ = 0.60, p = .550.

Study 5

In Study 4, we tried to assess antisemitism in a less blatant form, harsh criticism of Israel. In Study 5, we accounted for the possibility that Germans might be equally sensitized to criticism of Israel as to open antisemitic statements. It might be the case that when confronted with statements criticizing Israel in a questionnaire, Germans retrieve learned answers, leaving little room for situational influences. However, emotional reactions and behavior towards individual persons might be elicited more spontaneously and thus be more responsive to situational influences. Therefore we turned away from antisemitism and investigated a more subtle reaction in the area of intergroup-emotions and intergroup-behavior: Does reminding Germans of the Holocaust result in decreased empathy and support towards Israeli victims of rocket attacks? Again, we also aimed at testing whether this presumably defensive reaction is dependent on the level of national glorification.

Method

Participants. Ninety-eight students (53 women and 45 men) of different fields of study from the University of Cologne, Germany, participated in exchange for a bar of chocolate and the opportunity to win 50 EUR in a raffle. Participants ranged from 17 to 56 years in age (M = 22.97, SD = 5.68). Another two participants dropped out during the experiment and were deleted from the data set.

Materials and Procedure. Participants were seated at computers in individual cubicles and were randomly assigned to one of two experimental conditions (Holocaust reminder vs. control). After reading the instructions in which they were informed that the study was about German perceptions of Israel, participants started by reporting their age, gender, educational status, citizenship, and whether their family had a history of migration. Next, they were either reminded of the Holocaust or not. In the Holocaust reminder condition, participants read the same text that already had been used in Study 4b. The control condition was a baseline condition in which participants were simply told that "we are interested in your perception of different aspects of Israel." Subsequently, participants were presented with a short (47s) television news video about the Gaza-based rocket attacks on a village in Israel. After presentation of the video, empathy towards the Israeli victims was assessed using six adjectives from the empathy literature (e.g., "compassionate"; Cronbach's α = .73; Batson, Fultz, & Schoenrade, 1987). Using a seven-point scale (1 = not at all to 7 = extremely), participants indicated for each of the items to what extent they had felt the given emotion while watching the video about the situation of the Israeli victims. The empathy items were presented among ten filler items – eight distress adjectives (e.g., "worried") and two guilt adjectives (e.g., "guilty") for exploratory purposes.

In order to increase credibility of the cover story that the study was about German perceptions of different aspects of Israel, participants were presented with another short video, which was a report about young Israelis protesting against housing shortage and high rent prices, and indicated their agreement to six statements about the social protests (e.g., "I can easily identify with the requests of the protesters."). Subsequently, participants answered the eight-item national attachment (Cronbach's α = .86) and the eight-item national glorification scale (Cronbach's α = .79). Finally, they were presented with a screen saying that the study was over and that they could participate in a raffle offering the opportunity to win 50 EUR. Participation in the raffle included a measure of financial support of the Israeli victims as a second dependent variable (in addition to empathy). Participants were invited to pledge to donate a portion of their choice of these 50 EUR (between 0 and 50 EUR) to an organization supporting the Israeli victims in case of them winning the raffle. The text included a description of the charity organization.

Results

Empathy scores and donation pledge amounts were subjected to independent samples t-tests. Empathy scores did not significantly differ between participants who were reminded of the Holocaust ($M = 4.03$, $SD = 1.07$) and those in the control group ($M = 3.72$, $SD = 0.94$), $t(96) = -1.53$, $p = .129$, Hedges's $g_s = -0.31$, 95% CI[-0.70, 0.09]. Likewise, results revealed no significant group difference in the amounts participants pledged to donate in support of the Israeli victims (Holocaust reminder: $M = 18.48$, $SD = 13.41$; control: $M = 21.52$, $SD = 15.26$), $t(46) = 0.74$, $p = .466$, Hedges's $g_s = -0.21$, 95% CI[-0.78, 0.36]. To test whether level of national glorification moderated the hypothesized effect of Holocaust reminders on empathy, we conducted a hierarchical multiple regression analysis with the effect-coded group variable (Holocaust = 1; control = -1), attachment to Germany, glorification of Germany, and all possible two-way and three-way interaction terms as predictors. In contrast to our hypothesis, the product term representing the interaction between glorification of Germany and experimental condition did not significantly predict empathy in the full regression model, β = -.20, $t(90) = -1.29$, $p = .200$.

Table 1
Means, t-tests, and effect sizes for simple effects that would indicate secondary antisemitism across all studies.

Study Measure	N		M (SD)	M (SD)	t	df	p	Hedges's g [95% CI]	SE_g
Study 1 Change in explicit antisemitism	Ongoing consequences 22	No ongoing consequences 22	Ongoing consequences 0.11 (0.90)	No ongoing consequences -0.00 (0.78)	0.44	42	.666	0.13 [-0.45, 0.71]	0.30
Study 2 Change in explicit antisemitism	Bogus Pipeline 21	Control 23	Bogus Pipeline -0.47 (0.58)	Control -0.41 (0.87)	-0.27	42	.791	-0.08 [-0.66, 0.50]	0.30
Study 3a Warmth ratings of Classification Images	56		Jewish/Holocaust 3.02 (0.84)	Jewish/control 3.51 (0.96)	2.89	55	.006	0.54 [0.15, 0.93]	0.20
Study 3b Warmth ratings of Classification Images	43		Jewish/Holocaust 3.00 (0.89)	Jewish/control 3.16 (0.86)	1.11	42	.272	0.17 [-0.14, 0.48]	0.16
Study 3c Warmth ratings of Classification Images	64		Jewish/Holocaust 3.10 (0.84)	Jewish/control 2.61 (0.78)	-4.49	63	< .001	-0.60 [-0.88, -0.31]	0.15
Study 4a Criticism of Israel	Holocaust 50	Control 50	Holocaust 3.02 (0.72)	Control 3.06 (0.91)	-0.29	98	.776	-0.06 [-0.45, 0.33]	0.20

Study 4b Criticism of Israel	Holocaust 100	Control 96	Holocaust 3.78 (0.85)	Control 3.76 (0.88)	0.14	194	.890	0.02 [-0.26, 0.30]	0.14
Study 5 Empathy towards Israeli victims of rocket attacks	Holocaust 50	Control 48	Holocaust 3.72 (0.94)	Control 4.03 (1.07)	-1.53	96	.129	-0.31 [-0.70, 0.09]	0.20

Note: The effect size measure used in Studies 1, 2, 4a, 4b, and 5 is Hedges's g_s. In Studies 3a through 3c, we used Hedges's g_{av} for the difference of two correlated measurements as recommended by Lakens (2013).

Discussion

Study 5 did not provide evidence for the notion that being reminded of the Holocaust makes Germans (who score high on national glorification) less empathic with Israeli victims of Palestinian rocket attacks. We thus, again, failed to observe data patterns in line with secondary antisemitism. Although the study did not have particularly high statistical power according to current standards, the effect on empathy was – if anything – in the unexpected direction (Hedges's g of 0.31).

Meta-analysis across studies

Before discussing the findings, we want to integrate them to give an overall feeling for the obtained evidence or lack thereof. To do so, we calculated simple effects for each study that would indicate secondary antisemitism (Table 1). Although we predicted interactions for the first three studies, we decided to select simple comparisons instead of effect sizes of interaction terms, because simple comparisons are more informative regarding the direction of effects. For Studies 1 and 3 (that have a 2x2 design), we focused on the comparison between conditions with reminders of either ongoing suffering or the Holocaust and conditions with these on the evaluation of Jewish targets. For Study 2, each condition included an ongoing suffering manipulation and thus we compared the degree of (baseline-corrected) antisemitism in the bogus pipeline condition to that in the control condition. We calculated Hedges's g_s, respectively g_{av} (see Lakens, 2013), for each of the studies and conducted a random-effects model using the R metafor package (Viechtbauer, 2010). The data show substantial heterogeneity, $Q(7) = 27.14$, $p = < .001$, $I^2 = 72.26\%$, and meta-analytically again no evidence for any secondary antisemitism (Figure 5). Given the large heterogeneity, an average effect size of almost exactly zero may be taken as an indication that – despite underpowered studies – it was not just too small of samples that prevented us from replicating the effect.

General Discussion

Across a research program spanning two years and eight studies, we did not provide evidence for the notion that reminders of the Holocaust evoke negative responses towards Jews among German participants. None of the studies had particularly strong statistical power and none were
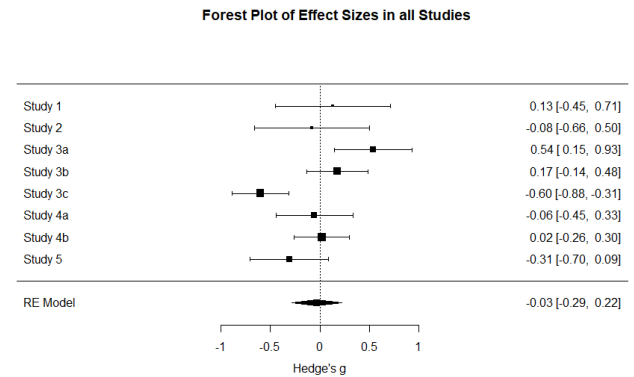

Forest Plot of Effect Sizes in all Studies

Figure 5. Forest plot of all studies reported in the manuscript. Simple effects are coded so that positive effects speak to the hypothesis of secondary antisemitism.

a direct replication in exact detail. Nevertheless, their consistency in not producing any result has shaken our confidence in the very basic effect. In light of this, the original goal to better understand the involved processes has lost some of its relevance.

As a caveat, although all of the studies reported here were conducted after the current debate on how to achieve more reproducible and reliable science had already taken off in 2011, the spirit behind this research program is still rooted in the "old way" of doing research. What we did (without success) was hunt down an effect, desperately seeking a way to "make it significant". What we did not do is systematically plan for compelling evidence – in either direction. In hindsight, many steps and detours we made may seem premature and a single extremely high-powered study might have

been more advisable. That was not the typical procedure for conducting social psychological science in our book. Instead, if the effect was not there, the researcher had the wrong method to show it and therefore needed to change that method. Independent of how we might have obtained more compelling evidence in either direction, all our studies consistently converged in not producing results in line with a very influential theory of post-Holocaust antisemitism.

What can we now make of this pattern? Does it mean that the original study (Imhoff & Banse, 2009) was a false positive? Although virtually everything in that original study fell exactly in its place, luck might have played a trick on us and convinced us of something that was never there to begin with (despite a plethora of qualitative writings and discourse analyses along the same lines). In light of the present research, we would argue that this is very well possible. Assuming that the original study was indeed a false positive might mean that either the very idea of secondary antisemitism is wrong or – as a more modest interpretation – that psychologists' illusions of omnipotence to translate a societal discourse and its dynamics into a pretty, 30-minute experiment are ill-advised. Maybe scholarly interpretations of the effect of continuous Holocaust reminders are indeed to the point but it is naïve to emulate this in a cute little study.

While we can only emphasize again that we are more than open to this possibility, we would – for the sake of the argument – like to entertain alternative explanations. Under the (admittedly speculative) assumption that the original study was a true positive, how could we explain the absence of any effect in that direction in all studies reported here?

One of the most parsimonious (and potentially cheap) explanations could rest on the assertion that – for whatever reasons – the bogus pipeline procedure just never worked as nicely as it did in 2007. As the original study attested, however, this is crucial to find the diverging effects of Holocaust reminders. The dilution of psychological research and knowledge about the (missing) validity of "lie detectors" might have made it increasingly difficult to convince people of the operating principle of the bogus pipeline. In fact, lie detectors are debunked on a regular base not only in undergraduate psychology classes but also popular media outlets. If that was true, the psychological processes of secondary antisemitism indeed happened within our participants as they did in 2007, but our setup was not potent enough to make participants admit antisemitism. Although we have no direct evidence against this plausibility, we would argue that the several other steps we undertook to reduce social desirability should then have produced at least a suggestive pattern in line with the hypothesis.

Another counter point to that argument might rest on the observation that psychologists tend to overestimate the power of social desirability, as most people are quite confident in the validity of their beliefs and do not adapt them according to what they think they ought to think and feel instead. In that case bogus pipelines would fail to produce effects not because they are not working, but because there is no hidden "real" belief to be revealed: People already speak frankly without such an apparatus. This would, however, mean that ongoing victim suffering did not increase our participants' prejudice. If it had, it should have produced a main effect independent of the bogus pipeline manipulation (not an interaction by the – in this logic – obsolete bogus pipeline), which we never observed. The lack of effect in subsequent studies while maintaining the speculation that the original effect was a true positive is addressed by a second potential explanation.

The second explanation is a little more difficult to argue with and in fact a reoccurring argument in the replication debate. Heraclitus' dictum that you cannot step into the same river twice, as it is not the same river and you are not the same person, served as an encapsulation of the notion that the objectively same thing is not the same as time has passed. Meanings change and persons change and potentially the effect we sought is subject to Zeitgeist effects to a much greater extent than we anticipated. In fact, many effects, particularly in social psychology, may be more prone to changing times and norms than most of us are typically willing to concede. The year 2007 not only was the year in which our original study took place, but also the year in which the last public negotiation between the Jewish Claims Conference Against Germany with the German federal government came to an agreement. This was a publicly followed event and for many Germans Martin Walser's (1998) infamous speech, in which he conceded that "I am almost glad when I think I can discover that more often not the remembrance, the not-allowed-to-forget is the motive, but the exploitation of our shame for current goals" still resonated. Very possibly, much like studies on prejudices against African Americans from the 1970s would not necessarily replicate in the 1980s and much less today, maybe also here the discursive context changed and thus made the seemingly highly similar experimental situation a psychologically very different one.

This is a specific case of Gergen's (1973) more general argument that psychology is primarily a historical inquiry as it "deals with facts that are largely nonrepeatable and which fluctuate markedly over time" (p. 310). If indeed empirical studies are then mere manifestations of a historically situated effect that neither does nor should be expected to be diachronically robust, what is the use of doing such studies? First, (obviously always under the premise that the original finding was no false positive) it illustrates this historically situated principle. It might be of interest (admittedly primarily for historical reasons) that in the early 21st century a markedly negative reaction towards Jewish suffering was observable among German students. Second, and potentially more interesting, the original study might illustrate a general principle and a methodological approach to it rather than establish a human constant. The principle may be that the emphasis on victim suffering can backfire if respondents are liberated from social desirability concerns. The method might thus provide a potential blueprint of how to tackle such research questions.

Is this enough for a field of science that so desperately seeks to be as hard a science as the physical sciences in which "stable and broad generalizations can be established with a high degree of confidence" (Gergen, 1971, p. 309) and thus allow explanations that can be empirically tested? Does such an understanding of science have a place in the replication era? One positive aspect of the new way of doing psychology is that (hopefully), unlike in the present case, original studies will have been pre-registered and (internally) replicated before publication. The likelihood that these results were true positives is much higher than in the current case of a single underpowered study. Failure to replicate in a different time, location, and/or context might thus inform our theorizing about the situatedness of the effect at hand. Thus, pre-registration will not only increase the trust in published findings, but also allow an informative hint whether it is advised or futile to seek hidden moderators.

In summary, we repeatedly failed to conceptually replicate one of our findings across a relatively large number of studies, with different methodological approaches. Thus, the claim that confrontation with the Holocaust evokes a backlash of antisemitism among Germans is not empirically well supported. Either the initial finding was a false positive or this process needs to be specifically situated in a given time and context.

Open Science Practices

This article earned the Open Data and the Open Materials badge for making the data and materials available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, are published in the online supplement.

References

Adorno, T. W. (1975). Schuld und Abwehr. In T. W. Adorno, *Gesammelte Schriften, Band 9 Soziologische Schriften II.2* (pp. 121-326). Frankfurt: Suhrkamp.

Banse, R., & Gawronski, B. (2003). Die Skala Motivation zu vorurteilsfreiem Verhalten: Psychometrische Eigenschaften und Validität. *Diagnostica, 49*, 4-13.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality, 55*, 19-39.

Bergmann, W. (2006). "Nicht immer als Tätervolk dastehen" - Zum Phänomen des Schuldabwehr-Antisemitismus in Deutschland In D. Ansorge (Ed.), *Antisemitismus in Europa und in der arabischen Welt* (pp. 81-106). Paderborn-Frankfurt: Bonifatius Verlag.

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35-58). Oxford, England: Blackwell Science.

Branscombe, N. R., Schmitt, M. T., & Schiffhauer, K. (2007). Racial attitudes in response to thoughts of White privilege. *European Journal of Social Psychology, 37*, 203-215.

Buruma, I. (2003, August). How to talk about Israel. *New York Times*, Section 6, p. 28.

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infrahumanization in response to collective responsibility for intergroup killing.

Journal of Personality and Social Psychology, 90, 804-818.

Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology, 71,* 464-478.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100,* 947-966.

Correia, I., & Vala, J. (2003). When will a victim be secondarily victimized? The effect of observer's belief in a just world, victim's innocence and persistence of suffering. *Social Justice Research, 16,* 379-400.

Dalbert, C. (1999). The world is more just for me than generally: About the personal belief in a just world scale's validity. *Social Justice Research, 12,* 79-98.

Dotsch, R. & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science, 3,* 562-571.

Dotsch, R., Wigboldus, D. H. J., & van Knippenberg, A. (2013). Behavioral information biases the expected facial appearance of members of novel groups. *European Journal of Social Psychology, 43,* 116-125.

Dotsch, R., Wigboldus, D. H., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science, 19,* 978-980.

Egloff, B. & Krohne, H. W. (1998). Die Messung von Vigilanz und kognitiver Vermeidung: Untersuchungen mit dem Angstbewältigungs-Inventar (ABI). *Diagnostica, 44,* 189-200.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878-902.

Friedman, J. S., & Austin, W. (1978). Observers' reactions to an innocent victim: Effect of characterological information and degree of suffering. *Personality and Social Psychology Bulletin, 4,* 569-574.

Funke, F. (2005). The dimensionality of right-wing authoritarianism: Lessons from the dilemma between theory and measurement. *Political Psychology, 26,* 195-218.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology, 26,* 309-320.

Godfrey, B. W., & Lowe, C. A. (1975). Devaluation of innocent victims: An attribution analysis within the just world paradigm. *Journal of Personality and Social Psychology, 31,* 944-951.

Golec de Zavala, A., Cichocka, A., Eidelson, R., & Jayawickreme, N. (2009). Collective narcissism and its social consequences. *Journal of Personality and Social Psychology, 97,* 1074-1096.

Heider, F. (1958*). The psychology of interpersonal relations.* New York: Wiley.

Heitmeyer, W. (2005). *Deutsche Zustände (Folge 3)* [German circumstances (Vol. 3)]. Frankfurt: Suhrkamp.

Imhoff, R., & Banse, R. (2009). Ongoing victim suffering increases prejudice: The case of secondary anti-semitism. *Psychological Science, 20,* 1443–1447.

Imhoff, R., & Bruder, M. (2014). Speaking (un-)truth to power: Conspiracy mentality as a generalised political attitude. *European Journal of Personality, 28,* 25–43.

Imhoff, R., & Dotsch, R. (2013). Do we look like me or like us? Visual projection as self- or ingroup-projection. *Social Cognition, 31,* 806-816.

Imhoff, R. (2010). Zwei Formen des modernen Antisemitismus? Eine Skala zur Messung primären und sekundären Antisemitismus. *Conflict & communication online, 9.*

Imhoff, R., Bilewicz, M., & Erb, H. (2012). Collective regret versus collective guilt: Different emotional reactions to historical atrocities. *European Journal of Social Psychology, 42,* 729–742.

Imhoff, R., Dotsch, R., Bianchi, M., Banse, R., & Wigboldus, D. H. J. (2011). Facing Europe: Visualizing spontaneous in-group projection. *Psychological Science, 22,* 1583-1590.

Imhoff, R., Woelki, J., Hanke, S., & Dotsch, R. (2013). Warmth and competence in your face! Visual encoding of stereotype content. *Frontiers in Psychology, 4,* 386.

Jost, J. T., & Hunyady, O. (2002). The psychology of system justification and the palliative function of ideology. *European Review of Social Psychology, 13,* 111-153.

Jost, J.T., Banaji, M.R., & Nosek, B.A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25,* 881–919.

Kaplan, E. H., & Small, C. A. (2006). Anti-Israel sentiment predicts anti-Semitism in Europe. Journal of Conflict Resolution, 50, 548-561.

Karremans, J. C., Dotsch, R.,& Corneille, O.(2011). Romantic relationship status biases memory of faces of attractive opposite-sex others: Evidence from a reverse-correlation paradigm. *Cognition, 121*, 422-426.

Kempf, W. (2014). Anti-Semitism and criticism of Israel: Methodology and results of the ASCI survey. *conflict & communication online, 14.*

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*, 863

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S., & van Knippenberg, A.(2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*, 1377–1388.

Lerner, M. J., & Simmons, C. H. (1966). Observer's reaction to the" innocent victim": Compassion or rejection? *Journal of Personality and Social Psychology, 4*, 203-210.

Lerner, M. J. (1980). *Belief in the just world*. New York: Plenum Press.

Lundqvist, D., Flykt, A., & Öhman, A. (1998*). The Karolinska directed emotional faces (KDEF)*. CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, (1998).

Mangini, M. & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science, 28*, 209–226.

Marhenke, T., & Imhoff, R. (2018*). Increased accessibility of semantic concepts after (more or less) subtle activation of related concepts: Support for the basic tenet of priming research.* Unpublished manuscript.

Miller, D. T. (1977). Altruism and threat to a belief in a just world. *Journal of Experimental Social Psychology, 13*, 113-124.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.

Quirin, M., Kazén, M., & Kuhl, J. (2009). When nonsense sounds happy or helpless: The Implicit Positive and Negative Affect Test (IPANAT). *Journal of Personality and Social Psychology, 97*, 500-516.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short

version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*, 203-212.

Roccas, S., Klar, Y., & Liviatan, I. (2006). The paradox of group-based guilt: modes of national identification, conflict vehemence, and reactions to the in-group's moral violations. *Journal of Personality and Social Psychology, 91*, 698-711.

Rüsch, N., Corrigan, P. W., Bohus, M., Jacob, G. A., Brueck, R., & Lieb, K. (2007). Measuring shame and guilt by self-report questionnaires: A validation study. *Psychiatry Research, 150*, 313-325.

Schönbach, P. (1961). *Reaktionen auf die antisemitische Welle im Winter 1959/1960* [Reactions to the anti-Semitic wave in the winter 1959/1960]. Frankfurt: Europäische Verlagsanstalt.

Selznick, G. J., & Steinberg, S. (1969). *The tenacity of prejudice: Anti-Semitism in contemporary America*. Oxford, England: Harper & Row.

Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology, 18*, 247-255.

Simons, C. W., & Piliavin, J. A. (1972). Effect of deception on reactions to a victim. *Journal of Personality and Social Psychology, 21*, 56-60.

Steinberg, G. (2004). Abusing the legacy of the Holocaust: The role of NGOs in exploiting human rights to demonize Israel. *Jewish Political Studies Review, 16*, 59-72.

Vala, J, Pereira, C. P., Eugênio, M., Lima, O., & Leyens, J. (2012). Intergroup Time Bias and Racialized Social Relations. *Personality and Social Psychology Bulletin, 38*, 491-504.

von Collani, G. (2002). Das Konstrukt der Sozialen Dominanzorientierung als generalisierte Einstellung: eine Replikation [The construct of social dominance orientation as a generalized attitude: A replication]. *Zeitschrift für Politische Psychologie, 10*, 263-282.

Walser, M. (1998). Erfahrungen beim Verfassen einer Sonntagsrede [Experiences while composing an oration]. In: Börsenverein des Deutschen Buchhandels (Hg.): *Friedenspreis des Deutschen Buchhandels 1998 - Ansprachen aus Anlaß der Verleihung* [Peace prize of the German booktrade 1998 – Speeches from the award ceremony]. Frankfurt/Main.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect. The PANAS scales.

Journal of Personality and Social Psychology, 54, 1063-1070.

Weil, F. D. (1985). The variable effects of education on liberal attitudes: A comparative historical analysis of anti-Semitism using public opinion survey data. *American Sociological Review, 50*, 458-474.