

The Micro-foundations of Naming and Shaming: *Evidence from the Proposed Annexation of the West Bank*

Lotem Bassan-Nygate *

Abstract

‘Naming and shaming’, a strategy in which a particular actor is singled out and publicly criticized for a negative behavior, is commonly used in international relations. Evidence from large-N and case studies are inconclusive regarding the impact of shaming on compliance with international norms. While some suggest that social pressure leads to change, others find a backlash effect. In addition, although both strands of work conjecture that shaming can impact policies through domestic publics, little work evaluated whether shaming impacts public opinion. There are even fewer studies that investigate potential mechanisms that may mobilize domestic support or opposition. To address these issues, I conducted a unique survey-experiment on a nationally representative sample of Jews in Israel, a country frequently shamed on international platforms. I leveraged the ambiguity regarding Israel’s plans to annex territories from the West Bank in order to shape respondents’ perceptions of global criticism. I find that shaming had a backlash effect: individuals treated with condemnation were more supportive of annexation and felt less obligated to international norms and laws in general. This result is likely driven by respondents’ need to maintain a positive image of their in-group.

*Department of Political Science, UW – Madison. ✉: lbassan@wisc.edu.

I am grateful for generous support from the Wisconsin Project on IR and the Center for Jewish Studies at the University of Wisconsin Madison. I thank Mark Copelovitch, Yoi Herrera, Andy Kydd, Michael Masterson, Jon Pevehouse, Jonathan Renshon, Nadav Shelef, Marcy Shieh, Jessica Weeks, Chagai Weiss and X. Zhang for helpful comments and suggestions.

State and non-state actors frequently rely on ‘naming and shaming’, the process of criticizing a government’s transgressions, to advocate domestic policy change. In 2018 alone, Amnesty International, a non-governmental organization, published over 1,000 reports shaming governments for human-rights abuses.¹ Although the international relations (IR) literature has long been interested in this phenomenon,² it provides conflicting evidence of the effects of global criticism. While some works suggest that shaming is a useful strategy to pressure decision-makers to conform with international standards (Franklin 2008; Murdie and Davis 2012; Krain 2012; Cole 2012; DeMeritt 2012; Hendrix and Wong 2013; Kelley and Simmons 2015), others argue that shaming hardly ever works and may backfire (Hafner-Burton and Tsutsui 2005; Bailey 2008; Hafner-Burton 2008; Lebovic and Voeten 2009; Terman 2017; Carnegie and Carson 2018).

Scholarly work does not only diverge when considering the impact of shaming, but also when debating how and why it works. Although most scholars seem to agree that global condemnations operate, to some extent, through the mobilization of domestic public opinion (Finnemore and Sikkink 1998; Risse-Kappen et al. 1999; Davis, Murdie and Steinmetz 2012; Ausderan 2014; Keck and Sikkink 2014; Kelley and Simmons 2015), they tend to black-box the micro-level mechanisms that motivate individual concern for shaming. As a result, there are very little theories on the micro-foundations of naming and shaming and even fewer empirical studies that examine whether public opinion is susceptible to global criticism, and why.

In this paper, I address the following questions: (1) does naming and shaming affect public opinion regarding international norms and transgressive behavior? and if so – (2) what are the mechanisms through which shaming works? I delineate two sets of testable hypotheses, one that predicts change and the other backlash following international criticism. The first set of hypotheses builds on general intuitions laid out by functionalist and constructivist approaches. The second is concerned with social psychology. I explore the role of reputational and moral costs as mobilizers

¹Amnesty International Research.

²Discussions date back to over two decades ago with the seminal work of Keck and Sikkink (1999).

of public opinion following shaming.

To test these theories, I fielded a unique survey experiment on a nationally representative sample of Jews in Israel, a country frequently shamed on international platforms. I leveraged a momentary uncertainty regarding Israel's plans to annex territories from the West Bank, which would constitute a violation of international laws,³ and norms regarding territorial integrity (Zacher 2001), in order to credibly shape respondents' perceptions of the likelihood of global criticism. Doing so, I was able to identify the impact of shaming on public opinion.

My findings are threefold. First, shaming had a backlash effect, leading Israelis to become more supportive of annexation and feel less obligated to observe international norms and laws. Second, this effect appears to be driven by motivated reasoning, as the treated respondents believe annexation is *more* moral, and that Israel faces *less* reputational damage, following condemnations. Third, this result was slightly more pronounced in individuals who reported stronger national attachments, further suggesting that individuals were motivated to maintain a positive image of their in-group. Thus, this research contributes to ongoing debates regarding the impact of social pressure, as well as to recent work on micro-foundations in international politics.

1 The Logic of Naming and Shaming

'Naming and shaming' is a strategy commonly used by states, IOs, NGOs, media outlets, etc. in which an actor is targeted for acting in an undesired way. It is a communicative tool that shines a negative spotlight on a behavior that violates a norm held by the international community (Keck and Sikkink 2014). Shaming usually takes the form of official reports, but recent work has also focused on numerical indicators, "watch lists" (Kelley and Simmons 2015) and peer-review mechanisms in IOs (Terman and Voeten 2018). Issuing globally condemning announcements, should, in theory, promote behavioral change and compliance with international norms. Interestingly, however, the literature provides inconsistent findings when measuring the effects of naming and shaming.

³United Nations Human Rights Report.

A variety of scholarly work argues that shaming has a positive effect on domestic policy change. For example, in the context of human rights, [Krain \(2012\)](#) finds that naming and shaming by IOs, NGOs and the media all have significant effects on the severity genocide. [Kelley and Simons \(2015\)](#) show that governments are more likely to criminalize human trafficking when they have been placed on a “watch list” or included in the US annual trafficking persons report. Conversely, some research suggests that shaming has a conditional, or more nuanced effect. [Franklin \(2008\)](#), for example, finds that the impact of human rights criticism on political repression in Latin America is conditional on foreign capital dependence. [Murdie and Davis \(2012\)](#) argue that shaming done by human rights organizations is effective only if the organization is physically present in the targeted state. More recently, [Terman and Voeten \(2018\)](#) argue that shaming in the UN universal periodic review is more likely to be taken seriously when done by allies.

On the other hand, a growing literature argues that shaming has null effects, and can even backfire. [Bailey \(2008\)](#) shows that shaming did not socialize states into ending commercial whaling, but rather provoked a domestic backlash as a result of a counter-network of pro-whaling activist groups. [Hafner-Burton \(2008\)](#) finds that governments publicly criticized for human rights violation rarely decrease these practices. She conjectures that leaders may even increase terror after publicity in order to offset improvements they make in other political rights (such as voting), following international pressure.

Research that falls in either of these lines of work tends to assume that domestic publics play some role within this process. The idea that shaming works through domestic pressure was first introduced by constructivist scholars ([Finnemore and Sikkink 1998](#); [Risse-Kappen et al. 1999](#); [Katzenstein 2013](#); [Keck and Sikkink 2014](#)). These early works suggested that shaming is part of norm socialization process, or a “boomerang effect”. According to this notion, domestic activists form networks with international actors, who engage in global criticism. In turn, these condemnations shapes domestic public opinion which pressures decision makers to conform with an international standard. The idea that change is ultimately mobilized through domestic public opinion has generally been adopted by the next waves of scholars. [Krain \(2012\)](#), for example, concludes

that shaming is effective due to the domestic opposition that follows. [Kelley and Simmons \(2015\)](#) conjecture that numeric indicators influence policymakers only to the extent that they influence domestic publics.

Studies that highlight the ineffectiveness of shaming describe similar micro-level dynamics. For example, elites may further violations ([Hafner-Burton 2008](#)) or use cues to frame shaming in a negative light ([Terman 2017](#)), in order to discourage domestic opposition or resistance. Similarly, works on stigma management in IR ([Adler-Nissen 2014](#); [Zarakol 2014](#)), suggest that political elites frame condemnations as celebratory or rebellious, which shapes public opinion and contributes to de-stigmatization, often creating a backlash effect.

Despite the importance of public opinion in theories of shaming, there have only been a handful of works that directly identified the impact of shaming on the general public, and to the best of my knowledge, no work has examined the individual-level mechanisms that drive this effect. [Davis, Murdie and Steinmetz \(2012\)](#) analyze the impact of shaming on the public's opinion of human rights within their country. They use a dataset of shaming reports issued by human rights organizations, together with data from the World Value Survey. [Ausderan \(2014\)](#) takes an experimental approach to study the impact of shaming on perceptions of human rights conditions within one's country. Both studies essentially show that shaming can shape the public's perceptions regarding the severity of human rights violations. In other words, domestic audiences update their beliefs over how much misconducts take place within their state, when they learn of global criticism.

However, these works do not investigate whether said publics are also more likely to support a change in these policies. As previously mentioned, IR theories argue that shaming works through domestic *pressure*. While the findings of [Davis, Murdie and Steinmetz \(2012\)](#) and [Ausderan \(2014\)](#) suggest that shaming provides individuals with credible information about current state behavior, it does not indicate that they are persuaded into supporting alternative policies. When we consider behaviors that are typically directed towards an enemy, rather than the public itself (such as torture or occupation), it may be quite difficult to mobilize public support for policy change.

Certain works in the context of international law utilize experimental designs in order to assess

such support for domestic policies (Chilton 2015; Chilton and Versteeg 2016; Kreps and Wallace 2016; Lupu and Wallace 2019a,b; Strezhnev, Simmons and Kim 2019). These experiments often present respondents with a domestic policy and randomize information on whether it is prohibited by international laws or norms. However, while shaming is often linked to international law, it presents a distinct phenomenon. Not only can shaming exist outside the realm of international law, as a communicative tool between states, shaming is also a deliberate act which includes the process of singling out and condemning an actor for their action. Such criticism is absent from these works and may yield different findings.

Another related issue concerns individual level characteristics of the public. The literature on naming and shaming tends to treat all actors that comprise the “public” as the same. However, a growing literature in IR demonstrated the importance of beliefs and dispositions (Kertzer, Renshon and Yarhi-Milo 2015; Kertzer and Brutger 2016; Kertzer 2017; Yarhi-Milo, Kertzer and Renshon 2018; Mattes and Weeks 2019). As further discussed below, systematic differences between individuals may affect the ways in which they process, interpret and react to international shaming. As domestic publics may vary internationally, learning about these characteristics could contribute to our understanding of the conditions under which shaming is more likely to succeed, and generate more hypotheses-testing work on a cross-national level.

2 Backlash and Change at the Grass-root Level

Clearly, empirical evidence from case studies and large-N analyses produce inconclusive findings regarding the impact of naming and shaming on compliance with international norms. While public opinion plays a role in both strands of work, literature on change and literature on backlash provide opposing expectations regarding the impact of shaming, as well as the mechanisms that drive this effect. Although such mechanisms are not always made explicit, I build on existing works in IR and social psychology to delineate testable hypotheses. The following section presents two competing theories regarding the impact of shaming. The first set of hypotheses (H1) presents an optimistic view towards shaming, and attempts to theorize why individuals may be mobilized by global criticism. The alternative set of hypotheses (H2), theorizes why shaming may backfire at

the individual level.

2.1 Optimistic position: Shaming positively impacts public opinion

Much of the work on international norms predicts that global criticism works through domestic pressure. Thus, to the extent that change is ultimately mobilized through domestic public opinion, we may hypothesize that learning that one's nation has been shamed for its behavior should generally increase support for policy change. Policy change refers to adherence with the international norm, for which the country has been shamed for violating. It thus follows that:

H1a: *Naming and shaming will increase individual-level support for policy change.*

Generating hypotheses regarding the mechanisms that motivate such support, however, requires a deeper burrow into the literature. I focus on functionalist and constructivist approaches to international relations, as these tend to view shaming as a useful tactic to either ensure compliance with international agreements, or socialize norms (respectively). Realist traditions, on the other hand, yield few predictions with respect to shaming, as they often view this tactic cynically as an ineffective political tool (Lebovic and Voeten 2006). In other words, to realists, behavioral change is contingent on security interests or changes in the balance of power, but never a result of normative pressure.

Conversely, functionalist traditions suggest states comply with international law and informal norms as these may foster beneficial cooperation in the future (Keohane 1984; Simmons 2009). Shaming is a form of monitoring state behavior, necessary to ensure commitments are being held. Actors therefore comply with international standards to maintain their reputation as reliable partners. Simmons (2000), for example, asserts that actors comply with international law to signal to others that they are suitable for cooperation in a broader range of issue areas. Moore (2002) argues that violation of human rights agreements may impact investors who view poorly the state's inability to restrain power for long-term benefits, thus providing empirical evidence of reputation linkage. While these studies often only deal with shaming indirectly, they provide a general intuition that targeting specific states for their behavior should raise concerns about reputation costs

(particularly, reputation for future cooperation). To the extent that shaming works through public opinion, we may conjecture that it highlights incongruence with international standards that appear to inhibit future cooperation, thus raising reputational concerns. It therefore follows that reputation is a potential mechanism of shaming:

H1b: *Naming and shaming increases concerns about state reputation, thus enhancing individual-level support for policy change.*

Constructivist theories tend to view shaming as integral to the process of the diffusion of international norms. As a result, works within this tradition are more explicit about how shaming is mobilized. Although some of these works similarly identify reputation as a potential mediator, it is often considered a mechanism that operates at the elite level, rather than one that drives public opinion (Finnemore and Sikkink 1998). Domestic publics, on the other hand, are seen as primarily motivated by normative standards of "appropriateness" and ideational concerns (March and Olsen 2004; Keck and Sikkink 2014). In their seminal work on the power of human rights, Risse-Kappen et al. (1999) argue that shaming is a process that raises moral consciousness by constructing categories of "us" and "them". As actors strive to aspire to belong to a "civilized community", they are convinced that the behavior of their nation is inconsistent with an identity to which they aspire. In their article on transnational advocacy groups, Keck and Sikkink (1999) similarly argue that activist groups can persuade people and stimulate them to take action by providing information on government behavior, which raises moral considerations. Shaming should therefore highlight the immorality and inappropriateness of one's nation, spurring support for change. It therefore follows that another potential mechanism of shaming is morality:

H1c: *Naming and shaming increases moral concerns, thus enhancing individual-level support for policy change.*

Finally, we may consider whether foreign policy dispositions may affect the ways in which individuals process international shaming, thus moderating its effect. As empirical works on naming and shaming have largely used cross-national observational data, there has been no previous

attempt to explore these microfoundations. I therefore turn to recent advances in experimental research in IR to theorize how individual characteristics can moderate the effects of shaming. Military assertiveness refers to the division between hawks and doves, which has been found to have important implications in shaping foreign policy attitudes. In particular, I acknowledge the recent advances in the study of reputation (Dafoe, Renshon and Huth 2014; Kertzer, Renshon and Yarhi-Milo 2015; Yarhi-Milo 2016; Brutger and Kertzer 2018), and consider how these dispositions can impact reputational concerns, which may be a mediator of shaming.

Specifically, I follow Brutger and Kertzer (2018) who suggest that hawks and doves view reputation costs through inherently different lenses. They find that hawks are concerned with negative reputational consequences of inconsistency, while doves are more concerned with belligerence. Hence, hawks are more likely to punish leaders who back down from a threat, while doves are more likely to punish those who pursue it. If we adapt this notion to the context of shaming, hawks and doves may interpret global condemnations in different ways. Since hawks tend to focus on reputation for resolve, they may view compliance with international pressure (or, changing policies) as a form of yielding which may invite further criticism and external pressure. Doves, on the other hand, may be more concerned with reputation for future cooperation, thus leading to stronger support for changing the current policy. Hence, while doves may be susceptible to naming and shaming, hawks may be more likely to reject it. It thus follows that:

H1d: *Naming and shaming increases support for policy change amongst doves, but not hawks.*

2.2 Pessimistic position: Shaming negatively impacts public opinion

Alternative theories, however, predict that shaming is likely to be met with a backlash effect. Recent theoretical work on stigma management (Adler-Nissen 2014; Zarakol 2014; Terman 2017) and from political psychology (Ilgit and Prakash 2019; Snyder 2020), suggest that tactics of shame may be counterproductive. These theories often point to empirical evidence from large-N analyses and case studies that suggest that global criticism may backfire and lead to more violations of international norms (Hafner-Burton and Tsutsui 2005; Bailey 2008; Hafner-Burton 2008; Lebovic

and Voeten 2009; Carnegie and Carson 2018). Thus, an alternative expectation to H1a is:

H2a: *Naming and shaming will decrease individual-level support for policy change (or, increase support for the current policy, violating the international norm).*

Different intuitions may be provided to derive expectations regarding the mechanisms that drive this negative effect. As recently underscored by Ilgit and Prakash (2019) and Snyder (2020), psychological literature often emphasizes the defensive reaction produced by shaming. Like other moral emotions, shame arises when there is a discrepancy between how one is expected to behave and how one actually behaves (Doosje et al. 1998; Tangney, Stuewig and Mashek 2007). However, as opposed to guilt, in which one puts an emphasis on the negative behavior, in shame individuals interpret the emotion-eliciting event as a failure of the self, that is – *I did that horrible thing* as opposed to *I did that horrible thing* (Lewis 1971). Shame is thus considered incredibly painful, and may provoke hiding, avoidance and aggression (Rees, Allpress and Brown 2013).

Importantly, psychological work on shaming focus on transgressions at the individual level. Thus, one may ask whether a translational error has been made when importing such theories into IR. However, works in social psychology show that individuals often experience shame ‘on behalf’ of their groups. Social identity theory (SIT), suggests that people derive a sense of identity from social associations and are thus motivated to maintain a positive image of their social groups (Tajfel 2010). As a result, when faced with transgressions of their in-group, individuals fear that the actions of their group reflect on the very nature of who they are, and are likely to defend their in-group, and by proxy their self-esteem (Lickel, Schmader and Barquissau 2004).

Recent work in social psychology finds that regulating painful emotions such as collective guilt and shame involves the process of motivated reasoning (Sharvit et al. 2015). Since harmful information about an in-group can challenge group members’ social identity, when faced with such information people are motivated to maintain their group’s positive image. Thus, cognitive processes of evaluating new information are biased by one’s own desires (Kunda 1990). Indeed, individuals may justify their group’s harmful behavior (Miron, Branscombe and Schmitt 2006), or deny their group’s responsibility (McGarty et al. 2005).

Importantly, this type of motivated reasoning is characterized as directional motivation (as opposed to accuracy motivation), in which individuals pursue a directional goal, such as identity protection (Druckman and McGrath 2019). Directional motivation often results in a backlash effect, since when exposed to new information that is inconsistent with one's prior beliefs, individuals may generate counter-arguments, creating disconfirmation bias. In the context of international shaming, when individuals are presented with harmful information about the in group – their nation – they may seek counter-arguments to suggest that their nation's actions are in fact moral, or non-transgressive. It is thus hypothesized that:

H2b: *Naming and shaming increases individuals' belief that their nation's actions are moral, thus decreasing support for policy change.*

I further consider how individual dispositions may moderate this effect. SIT as well as the literature on motivated reasoning suggests that the more attached one feels to their in-group, the more pronounced their need to defend it. Indeed, work from social psychology (Branscombe and Wann 1994; Doosje, Ellemers and Spears 1995; Ellemers, Spears and Doosje 2002), as well as recent work in IR (Herrmann 2017), finds that the greater one's identification with their group, the stronger their motivation to uphold its positive view. It therefore follows that:

H2c: *National attachments moderate the impact of naming and shaming on support for policy change, such that more nationalist individuals are less likely to support policy change.*

An alternative, though somewhat related strand of work similarly predicts that shaming may result in a backlash effect. Work on stigma management in IR suggests that international pressure in the form of shaming can create hierarchies between countries (Adler-Nissen 2014; Zarakol 2014). In parallel to the functionalist theories presented in section 2.1, concerns about status loss drive actors to improve their position vis à vis the international community. However, contra to the functionalist approach, improvement of status is not done by adopting the international standard. Rather, state actors strive to transform the dominant moral discourse to frame transgressions as rebellious or celebratory.

Since this theory is largely concerned with leaders, deriving testable hypotheses for the public introduces some challenges. Crucially, it is unclear whether the public is initially supportive of policy change, prior to elites' framing of the violation in positive terms. However, since recent work suggests that the elite-public gaps in decision-making are often overstated (Kertzer 2020), I hypothesize that status loss is a driving motivator for ordinary individuals as well:⁴

H2d: *Naming and shaming increases individuals' belief that their nation's status has been compromised, decreasing support for policy change.*

3 Research Design

Identifying the impact of international condemnations on public opinion is challenging since public opinion may be both a cause and an effect of shaming. In other words, countries in which the general public is not supportive of certain international norms may be more frequently subject to scrutiny by the international community. Alternatively, condemnations themselves may trigger more hostility towards the international community and lead to less support of internationally normative policies.

To overcome this inferential hurdle, I adapt a unique experimental design. Specifically, I follow Tankard and Paluck (2017) who leveraged political uncertainty around US Supreme Court rulings to exert stronger external validity in their experiment. Similarly, I exploited a limited time window in which it was unclear whether Israel will annex territories from the West Bank, to examine whether global criticism affects public attitudes. Doing so allowed me to examine a realistic scenario for which Israel was, indeed, condemned by the international community.

⁴Another observable implication of stigma management theories may be connected to partisanship as a moderator. In Israel, leaders from the political right often frame condemnations as a threat to national security, thus trying to reshape the moral discourse produced by international actors. We may thus expect right-wing supporters, who are more likely to take these elite cues at face value to have a stronger backlash effect to shaming. I further explore this point in the appendix, in section E.

During the 2019 legislative elections campaign in Israel, Prime Minister Benjamin Netanyahu vowed to annex large parts of the occupied West Bank by July 1st 2020.⁵ Israeli administration over Palestinian territories would be illegal under international law,⁶ and would be considered a violation of international norms regarding territorial integrity (Zacher 2001). Indeed, shortly after Netanyahu's announcement there were some reports of international condemnation.⁷ Nonetheless, reports were relatively new which allowed me to credibly shape respondents' perceptions about criticism of Israel's intentions and identify its effect on a host of outcomes, detailed below.

Empirically evaluating the theoretical expectations laid out in the previous sections in the Israeli context involves certain benefits, as well as challenges. Since Israel is one of the most publicly condemned countries across the world (Lebovic and Voeten 2006), testing my propositions on the general public in Israel simulates a real-world scenario. Arguably, scholars of international pressure as well as actors who rely on tactics such as naming and shaming particularly care about the impact of condemnations in places like Israel, which are frequently shined with a global spotlight.

Furthermore, the international community's routine criticism of Israel makes it a particularly intriguing case to study shaming. On the one hand, Israelis may have adopted a cynical view of shaming altogether, or a general disdain of the international community, a result of excessive criticism. Indeed, a 2013 Pew global poll found that the Israeli public holds the most hostile attitudes towards the UN of the 39 nations polled, with 65% of respondents expressing unfavorable views (Fagan and Huang 2019).

Thus, it may be particularly difficult to identify an effect of shaming in the Israeli context, making it a least likely case for theories related to change. On the other hand, frequent shaming may have made Israelis more attentive to international criticism and more sensitive to status loss. Thus, Israel may serve as a most likely case for theories concerned to reputation and status, such

⁵“Netanyahu vows to annex large parts of occupied West Bank”, *The Guardian*.

⁶United Nations Human Rights Report.

⁷“Devastating: Global condemnation after Netanyahu pledges to annex Jordan Valley, in occupied West Bank”, *Independent*.

as functionalism and stigma management.

3.1 Experimental sample

I fielded the study during June 2020, recruiting 1,500 respondents using iPanel, Israel's largest opt-in online survey company. The sample is representative of the Israeli Jewish population based on gender, age, ethnicity, living area, and religiosity. Descriptive statistics, and a comparison of the Israeli public and the sample on a variety of covariates, can be found in the appendix.

3.2 The experiment

The survey was programmed online using Qualtrics and presented to subjects in Hebrew. Respondents were exposed to the experiment – the structure of which is depicted in Figure A6 – following a battery of questions on demographics, along with a military assertiveness scale⁸ and a nationalism indicator.⁹ The experiment presented all subjects with a baseline vignette in which respondents were told that Israel is likely to annex part of the West Bank in the coming weeks. Respondents were then randomly assigned into control and treatment groups. Subjects in the treatment condition were told that many countries condemned Israel for its policy and said that Israel should be ashamed of itself. The control group was simply provided with the baseline vignette.¹⁰

I deployed two measures to capture the outcome variable, support for the international norm. First, respondents were asked whether they support annexation of territories from the West Bank. Subjects reported their support for the violation of the international norm on a scale ranging from 1 to 5, where higher values indicate stronger support for annexation. I also provide a more general

⁸This scale was used to determine hawks and doves in the sample, I borrowed it from [Brutger and Kertzer \(2018\)](#) who modify a scale typically used by political psychologists, developed by [Herrmann, Tetlock and Visser \(1999\)](#).

⁹Which I borrow from [Levendusky \(2018\)](#).

¹⁰Studies suggest that control and treatment groups should have equal length text. However, in this case the treatment includes one short sentence, I therefore do not provide unrelated information to the control group which may confuse respondents and interrupt their reading flow.

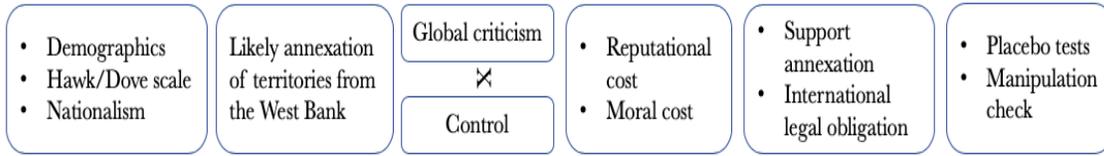


Figure 1: Experimental design

measure of international legal obligation, developed by [Bayram \(2017\)](#), in which respondents were asked about the extent to which they believe Israel should respect international law and norms (see appendix, section A for full measure). This variable ranges from 1 to 5 where high values indicate stronger obligation to international law and norms.

All respondents were also provided with questions intended to capture potential mediators: reputation and morality. Specifically, they were asked whether they believe Israel’s reputation has been damaged, and whether annexation of territories from the West Bank is moral. Subjects reported their answers on scales ranging from 1 to 5, where higher values indicated more reputational damage and less moral, respectively. Finally, respondents were required to complete manipulation checks and placebo tests in order to ensure that the treatment was successful and that “information leakage” about shaming has not confounded the treatment ([Tomz and Weeks 2013](#)).¹¹

4 Results

I begin by evaluating the effect of the shaming treatment on my main outcomes of interest. To do so, I look at the average treatment effects (ATEs) of the shaming treatment on support for annexation of territories from the West Bank and international legal obligation, as depicted in [Figure A2](#). The results suggest that shaming has a backlash effect, thus providing support for the second set of hypotheses, laid out in [Section 2.2](#). Respondents who received information about international shaming were more likely to support annexation of territories from the West Bank by 18% of a standard deviation, and reported lower levels of international legal obligation by 11% of a standard deviation. Both results are statistically significant ($p < 0.001$ and $p < 0.03$, respectively).

Clearly, the shaming treatment has made respondents more supportive of the transgressive

¹¹See appendix, section F, for a discussion of these tests.

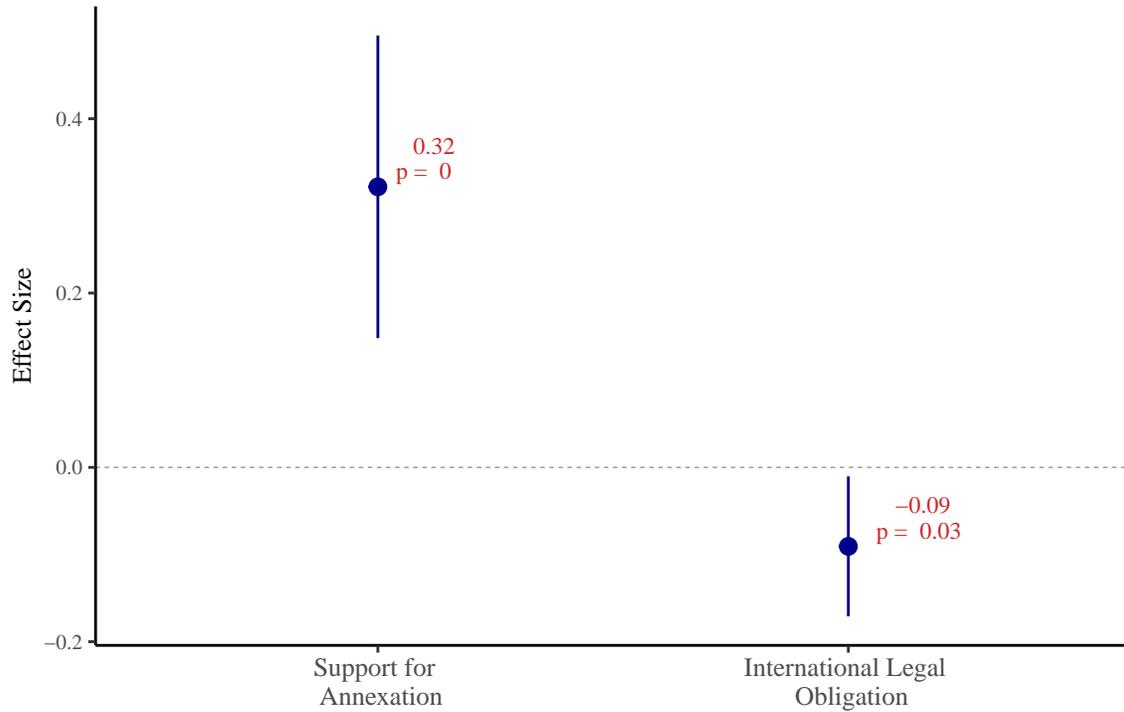


Figure 2: Average treatment effect plot. The effect of shaming treatment on support for annexation of territories from the West Bank and international legal obligation.

behavior, and less obligated to international norms and laws in general. Next, I explore different mechanisms that may drive this effect. It is important to remember that opposing theories, presented in sections 2.1 and 2.2, offer alternative expectations with respect to morality and reputation. Figure 3 provides additional support for the psychological dynamics laid out in Section 2.2. The shaming treatment seems to increase respondents' belief that Israel's policy is moral by approximately 13% of a standard deviation ($p = 0.01$). This finding suggests that when respondents were provided with information that is incongruent with their national identity they generated counter-arguments, suggesting their nation's actions are non-transgressive and moral, thus providing evidence in support of motivated reasoning.

On the other hand, contra to the expectations provided by functionalist theories and theories related to stigma management, it appears that shaming does not have a statistically significant effect on reputational concerns. Even more intriguing is the direction of the relationship. It appears that respondents who received the shaming treatment were less likely to perceive annexation as harming

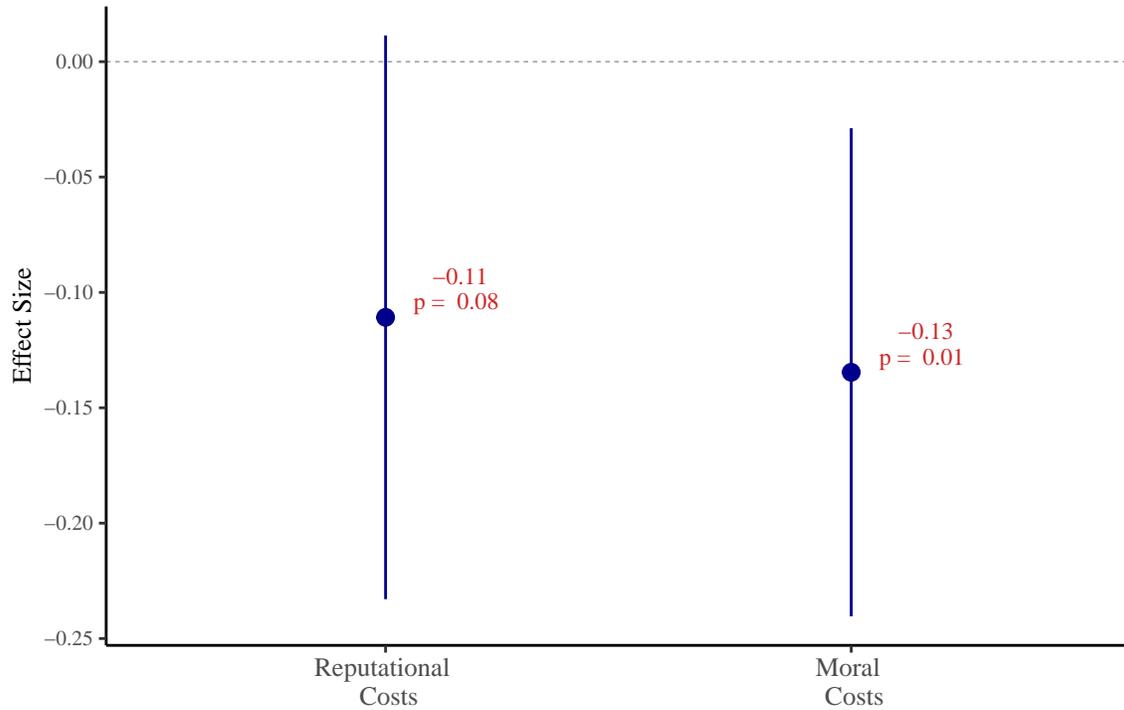


Figure 3: Exploring potential Mediators. The effect of shaming treatment on moral and reputational costs.

Israel’s reputation. One may thus interpret this result as further suggestive evidence of motivated reasoning. That is, respondents are biased by their own desires to maintain their national identity in positive terms, and are thus motivated to deny the potential implications of their group’s actions.

While the effect reported in Figure 3 is promising, it does not directly evaluate H2b. To learn whether moral costs mediate the relationship between shaming and support for annexation, I conduct a mediation analysis. The results of this analysis are depicted in Figure 4. The average causal mediation effect (ACME), or the indirect effect of moral cost, is positive and statistically significant at $p = 0.02$. The direct effect and the total effect are also positive and statistically significant at $p < 0.002$.

A potential concern with this finding is that an omitted variable might account for both moral costs and support for annexation, thus violating the “sequential ignorability” assumption necessary for mediation analysis (Imai et al. 2011). To address this issue, I conducted a sensitivity analysis, calculating the mediation effect for different magnitudes of a potential confounder’s effect on the

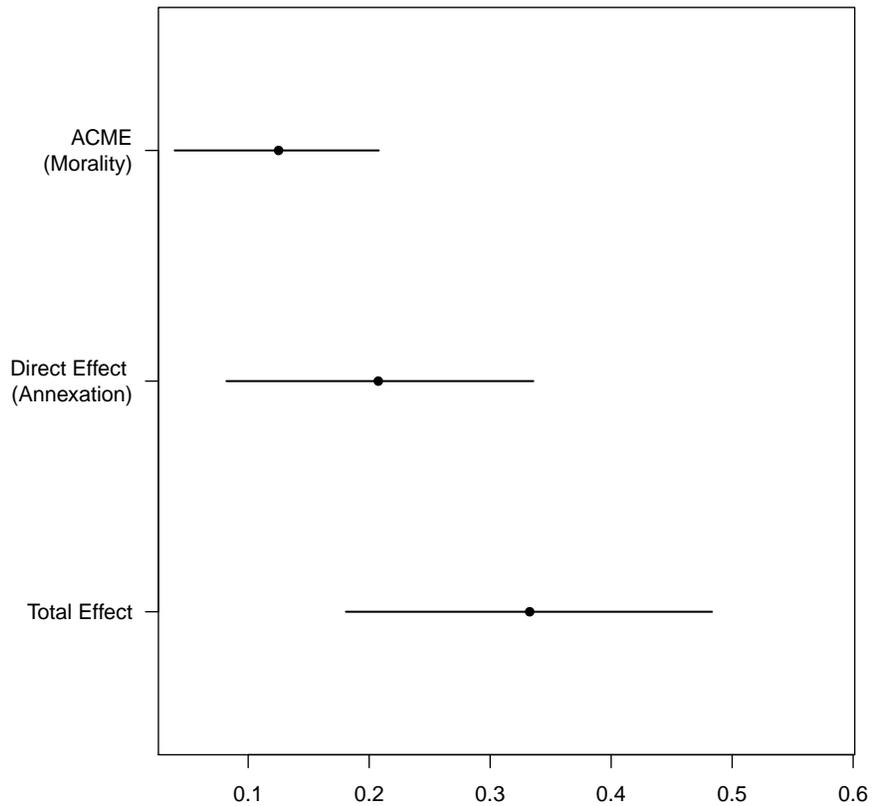


Figure 4: Causal mediation plot. Treatment is shaming manipulation, Mediator is morality cost (post-treatment), Outcome is support for support for annexation. Horizontal lines represent 90% confidence intervals for estimates.

mediator and the outcome. The results (in the appendix, section D), present an analysis for two cases: (a) where the omitted variable influences policy support and morality in the same direction, and (b) where it influences them in opposite directions.

Under the assumption that the confounder influences the mediator and the outcome in the same direction, my analysis suggests that for the positive mediation effect to be negative, the confounder must explain 50% of the variation in morality and 50% of the variation in annexation support. However, under the assumption that the confounder influences the mediator and the outcome in opposite directions, the sensitivity analysis suggests that the confounder only strengthens my results. This is indicated by the mediation effects attached to each contour line in figure D1.(a) in appendix D, which are all positive and grow stronger as more variance is explained.

Finally, I test hypotheses H1d and H2c, to examine whether my main effect is moderated by hawkishness or national attachments, respectively. To do so, I interact my shaming treatment with the hawk-dove scale, as well as the nationalism indicator. Models 1 and 2 in Table 1 suggest that Hawkishness is positively associated with support for annexation. This result is statistically significant, which suggests that hawks are generally more likely to support annexation. However, the interaction of hawkishness and shaming yields a weak and statistically insignificant result. Since H1d builds on the work of [Brutger and Kertzer \(2018\)](#), and is essentially about the interpretation of reputational costs, I also evaluate the impact of the interaction term `hawkishness*shaming` on reputational concerns in the appendix (see table E.1 in the appendix). While hawks appear to have minimal concerns for reputational costs in general, the interaction does not appear to have a statistically significant association with reputation.

Table 1: Heterogeneous Treatment Effects on Annexation

	Support Annexation			
	(1)	(2)	(3)	(4)
Shaming	-0.267 (0.370)	-0.411 (0.363)	-1.126 (0.424)	-0.715 (0.424)
Hawkishness	0.397 (0.051)	0.210 (0.055)		
Shaming*Hawkishness	0.110 (0.072)	0.133 (0.070)		
Nationalism			0.287 (0.049)	0.248 (0.049)
Shaming*Nationalism			0.259 (0.068)	0.181 (0.069)
Demographic Controls	No	Yes	No	Yes
<i>N</i>	1,529	1,529	1,521	1,521

On the other hand, national attachments appear to moderate the impact of shaming on support for annexation. As presented in models 3 and 4 in Table 1, the interaction of nationalism and shaming yields statistically significant results, and accounts for 14% of a standard deviation. Results remain statistically significant even when controlling for a host of demographic controls. These

findings provide further support for theories from social psychology presented in Section 2.2. In essence, they suggest that as one's attachments to their in-group (nation) grows, the need to defend it intensifies.

5 Conclusion

In this paper I draw on existing works in IR and social psychology to theorize about the effect of 'naming and shaming' on public opinion. I delineate two sets of testable hypotheses regarding this relationship, which point to backlash and change at the grass-root level. Since causally identifying the impact of international condemnation on public opinion is challenging, I utilized a unique survey experiment in Israel, which leveraged uncertainty regarding annexation of territories from the West Bank. I find that shaming has a backlash effect, causing individuals to further support the transgressive policy. Respondents treated with shaming are also less likely to support compliance with and respect for international laws and norms in general.

This study contributes to the existing literature on three fronts. First, it contributes to longstanding debates regarding the effect of international condemnations. While some works emphasize the ability of international pressure to enhance compliance with international norms (Franklin 2008; Murdie and Davis 2012; Krain 2012; Cole 2012; DeMeritt 2012; Hendrix and Wong 2013; Kelley and Simmons 2015), others find it may backfire, leading to more violations (Hafner-Burton and Tsutsui 2005; Bailey 2008; Hafner-Burton 2008; Lebovic and Voeten 2009; Carnegie and Carson 2018). I provide causally identified evidence for the latter, suggesting that shaming leads to more support of the transgressive behavior.

Second, while both strands of work conjecture that public opinion plays a role in the process of shaming, little to no empirical evidence is provided to test how the public reacts to condemnations, and why. In this work, I provide evidence to suggest that backlash effect occurs through a process of motivated reasoning, in which individuals are motivated to maintain a positive image of their in-group, denigrating arguments that challenge their identity through disconfirmation bias.

Specifically, I find that individuals are more likely to view the transgression as moral, and less likely to view their country's reputation as compromised, when they are told of international con-

demnations. In addition, I find that these backlash dynamics are more pronounced in individuals who have stronger national attachments, further supporting theories of social psychology regarding motivated reasoning and collective shame, which theorize regarding the heterogeneous treatment effects amongst stronger in-group identifiers. Hence, by focusing on these mechanisms and sub-group effects, I contribute to the growing literature on the micro-foundations of international politics (Kertzer 2017; Hafner-Burton et al. 2017).

Third, from practitioners' perspective, to the extent that actors such as NGOs who frequently rely on shaming as a strategy want to change the public's hearts and minds, it appears that global criticism is counterproductive. My finding on international legal obligations is even more alarming, as it suggests that shaming does not only effect individuals' support for a single policy or issue, but rather for international norms in general. Such effect may have devastating consequences for the international legal order. Recent work suggests that public opinion influences decision makers' foreign policies (Tomz, Weeks and Yarhi-Milo 2020). Thus, political elites may view the general public's hostility towards international norms as a green-light to engage in more non-compliant behavior.

Despite these contributions, this paper faces some limitations. Conducting the research in Israel, a country frequently shamed on international platforms, introduces some challenges to inference. In particular, Israelis may have adopted a cynical view of shaming altogether, a result of excessive criticism, and thus developed a general disdain for the international community. Perhaps shaming is thus more likely to backfire in countries that are frequently shamed, such as Israel. It is important to remember, however, that the question of whether Israel remains an easy or hard case for shaming is an empirical one. Alternative logics may suggest that Israelis are rather more attentive to shaming and sensitive to status loss, and should thus be more likely to care about international criticism. Future research should thus strive to develop similar designs to identify the impact of shaming in places where international criticism is rare.

References

- Adler-Nissen, Rebecca. 2014. "Stigma management in international relations: Transgressive identities, norms, and order in international society." *International Organization* pp. 143–176.
- Ausderan, Jacob. 2014. "How naming and shaming affects human rights perceptions in the shamed country." *Journal of Peace Research* 51(1):81–95.
- Bailey, Jennifer L. 2008. "Arrested development: The fight to end commercial whaling as a case of failed norm change." *European Journal of International Relations* 14(2):289–318.
- Bayram, A Burcu. 2017. "Due Deference: Cosmopolitan Social Identity and the Psychology of Legal Obligation in International Politics." *International Organization* 71.
- Branscombe, Nyla R and Daniel L Wann. 1994. "Collective self-esteem consequences of outgroup derogation when a valued social identity is on trial." *European Journal of Social Psychology* 24(6):641–657.
- Brutger, Ryan and Joshua D Kertzer. 2018. "A dispositional theory of reputation costs." *International Organization* 72(3):693–724.
- Carnegie, Allison and Austin Carson. 2018. "The Spotlight's Harsh Glare: Rethinking Publicity and International Order." *International Organization* pp. 1–31.
- Chilton, Adam S. 2015. "The laws of war and public opinion: An experimental study." *Journal of Institutional and Theoretical Economics: JITE* pp. 181–201.
- Chilton, Adam S and Mila Versteeg. 2016. "International law, constitutional law, and public support for torture." *Research & Politics* 3(1):2053168016636413.
- Cole, Wade M. 2012. "Institutionalizing shame: The effect of Human Rights Committee rulings on abuse, 1981–2007." *Social science research* 41(3):539–554.

- Dafoe, Allan, Jonathan Renshon and Paul Huth. 2014. "Reputation and status as motives for war." *Annual Review of Political Science* 17:371–393.
- Davis, David R, Amanda Murdie and Coty Garnett Steinmetz. 2012. "' Makers and Shapers': Human Rights INGOs and Public Opinion." *Human Rights Quarterly* pp. 199–224.
- DeMeritt, Jacqueline HR. 2012. "International organizations and government killing: does naming and shaming save lives?" *International Interactions* 38(5):597–621.
- Doosje, Bertjan, Naomi Ellemers and Russell Spears. 1995. "Perceived intragroup variability as a function of group status and identification." *Journal of experimental social psychology* 31(5):410–436.
- Doosje, Bertjan, Nyla R Branscombe, Russell Spears and Antony SR Manstead. 1998. "Guilty by association: When one's group has a negative history." *Journal of personality and social psychology* 75(4):872.
- Druckman, James N and Mary C McGrath. 2019. "The evidence for motivated reasoning in climate change preference formation." *Nature Climate Change* 9(2):111–119.
- Ellemers, Naomi, Russell Spears and Bertjan Doosje. 2002. "Self and social identity." *Annual review of psychology* 53(1):161–186.
- Fagan, Moira and Christine Huang. 2019. "United Nations gets mostly positive marks from people around the world."
URL: <https://www.pewresearch.org/fact-tank/2019/09/23/united-nations-gets-mostly-positive-marks-from-people-around-the-world/>
- Finnemore, Martha and Kathryn Sikkink. 1998. "International norm dynamics and political change." *International organization* 52(4):887–917.
- Franklin, James C. 2008. "Shame on you: the impact of human rights criticism on political repression in Latin America." *International Studies Quarterly* 52(1):187–211.

- Hafner-Burton, Emilie M. 2008. "Sticks and stones: Naming and shaming the human rights enforcement problem." *International Organization* 62(4):689–716.
- Hafner-Burton, Emilie M and Kiyoteru Tsutsui. 2005. "Human rights in a globalizing world: The paradox of empty promises." *American journal of sociology* 110(5):1373–1411.
- Hafner-Burton, Emilie M, Stephan Haggard, David A Lake and David G Victor. 2017. "The behavioral revolution and international relations." *International Organization* 71(S1):S1–S31.
- Hendrix, Cullen S and Wendy H Wong. 2013. "When is the pen truly mighty? Regime type and the efficacy of naming and shaming in curbing human rights abuses." *British Journal of Political Science* 43(3):651–672.
- Herrmann, Richard K. 2017. "How attachments to the nation shape beliefs about the world: A theory of motivated reasoning." *International Organization* 71:S61.
- Herrmann, Richard K, Philip E Tetlock and Penny S Visser. 1999. "Mass public decisions on go to war: A cognitive-interactionist framework." *American Political Science Review* 93(3):553–573.
- Ilgit, Asli and Deepa Prakash. 2019. "Making Human Rights Emotional: A Research Agenda to Recover Shame in "Naming and Shaming"." *Political Psychology* 40(6):1297–1313.
- Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* 105(4):765–789.
- Katzenstein, Suzanne. 2013. "Reverse-Rhetorical Entrapment: Naming and Shaming as a Two-Way Street." *Vand. J. Transnat'l L.* 46:1079.
- Keck, Margaret E and Kathryn Sikkink. 1999. "Transnational advocacy networks in international and regional politics." *International social science journal* 51(159):89–101.
- Keck, Margaret E and Kathryn Sikkink. 2014. *Activists beyond borders: Advocacy networks in international politics*. Cornell University Press.

- Kelley, Judith G and Beth A Simmons. 2015. "Politics by number: Indicators as social pressure in international relations." *American journal of political science* 59(1):55–70.
- Keohane, Robert O. 1984. *After hegemony*. Vol. 54 Princeton: Princeton University Press.
- Kertzer, Joshua D. 2017. "Microfoundations in international relations." *Conflict management and peace science* 34(1):81–97.
- Kertzer, Joshua D. 2020. "Re-Assessing Elite-Public Gaps in Political Behavior." *American Journal of Political Science* .
- Kertzer, Joshua D, Jonathan Renshon and Keren Yarhi-Milo. 2015. "How Do Observers Assess Resolve?" *British Journal of Political Science* pp. 1–23.
- Kertzer, Joshua D and Ryan Brutger. 2016. "Decomposing audience costs: Bringing the audience back into audience cost theory." *American Journal of Political Science* 60(1):234–249.
- Krain, Matthew. 2012. "J'accuse! Does naming and shaming perpetrators reduce the severity of genocides or politicides?" *International Studies Quarterly* 56(3):574–589.
- Kreps, Sarah E and Geoffrey PR Wallace. 2016. "International law, military effectiveness, and public support for drone strikes." *Journal of Peace Research* 53(6):830–844.
- Kunda, Ziva. 1990. "The case for motivated reasoning." *Psychological bulletin* 108(3):480.
- Lebovic, James H and Erik Voeten. 2006. "The politics of shame: the condemnation of country human rights practices in the UNCHR." *International Studies Quarterly* 50(4):861–888.
- Lebovic, James H and Erik Voeten. 2009. "The cost of shame: International organizations and foreign aid in the punishing of human rights violators." *Journal of Peace Research* 46(1):79–97.
- Levendusky, Matthew S. 2018. "Americans, not partisans: Can priming American national identity reduce affective polarization?" *The Journal of Politics* 80(1):59–70.
- Lewis, Helen B. 1971. "Shame and guilt in neurosis." *Psychoanalytic review* 58(3):419.

- Lickel, Brian, Toni Schmader and Marchelle Barquissau. 2004. "The distinction between collective guilt and collective shame." *Collective guilt: International perspectives* pp. 35–55.
- Lupu, Yonatan and Geoffrey PR Wallace. 2019a. "Violence, nonviolence, and the effects of international human rights law." *American Journal of Political Science* 63(2):411–426.
- Lupu, Yonatan and Geoffrey PR Wallace. 2019b. "Violence, nonviolence, and the effects of international human rights law." *American Journal of Political Science* 63(2):411–426.
- March, James G and Johan P Olsen. 2004. The logic of appropriateness. In *The Oxford handbook of political science*.
- Mattes, Michaela and Jessica LP Weeks. 2019. "Hawks, Doves, and Peace: An Experimental Approach." *American Journal of Political Science* 63(1):53–66.
- McGarty, Craig, Anne Pedersen, Colin Wayne Leach, Tamarra Mansell, Julie Waller and Ana-Maria Bliuc. 2005. "Group-based guilt as a predictor of commitment to apology." *British Journal of Social Psychology* 44(4):659–680.
- Miron, Anca M, Nyla R Branscombe and Michael T Schmitt. 2006. "Collective guilt as distress over illegitimate intergroup inequality." *Group Processes & Intergroup Relations* 9(2):163–180.
- Moore, David H. 2002. "A signaling theory of human rights compliance." *Nw. UL Rev.* 97:879.
- Murdie, Amanda M and David R Davis. 2012. "Shaming and blaming: Using events data to assess the impact of human rights INGOs." *International Studies Quarterly* 56(1):1–16.
- Rees, Jonas H, Jesse A Allpress and Rupert Brown. 2013. "Nie wieder: Group-based emotions for in-group wrongdoing affect attitudes toward unrelated minorities." *Political Psychology* 34(3):387–407.
- Risse-Kappen, Thomas, Thomas Risse, Stephen C Ropp and Kathryn Sikkink. 1999. *The power of human rights: International norms and domestic change*. Vol. 66 Cambridge University Press.

- Sharvit, Keren, Marco Brambilla, Maxim Babush and Francesco Paolo Colucci. 2015. "To feel or not to feel when my group harms others? The regulation of collective guilt as motivated reasoning." *Personality and Social Psychology Bulletin* 41(9):1223–1235.
- Simmons, Beth A. 2000. "International law and state behavior: Commitment and compliance in international monetary affairs." *American Political Science Review* 94(4):819–835.
- Simmons, Beth A. 2009. *Mobilizing for human rights: international law in domestic politics*. Cambridge University Press.
- Snyder, Jack. 2020. "Backlash against naming and shaming: The politics of status and emotion." *The British Journal of Politics and International Relations* p. 1369148120948361.
- Strezhnev, Anton, Beth A Simmons and Matthew D Kim. 2019. "Rulers or Rules? International Law, Elite Cues and Public Opinion." *European Journal of International Law* 30(4):1281–1302.
- Tajfel, Henri. 2010. *Social identity and intergroup relations*. Cambridge University Press.
- Tangney, June Price, Jeff Stuewig and Debra J Mashek. 2007. "Moral emotions and moral behavior." *Annu. Rev. Psychol.* 58:345–372.
- Tankard, Margaret E and Elizabeth Levy Paluck. 2017. "The effect of a Supreme Court decision regarding gay marriage on social norms and personal attitudes." *Psychological science* 28(9):1334–1344.
- Terman, Rochelle. 2017. "Rewarding resistance: Theorizing defiance to international norms." *Unpublished Manuscript* .
- Terman, Rochelle and Erik Voeten. 2018. "The relational politics of shame: Evidence from the universal periodic review." *The Review of International Organizations* 13(1):1–23.
- Tomz, Michael, Jessica LP Weeks and Keren Yarhi-Milo. 2020. "Public opinion and decisions about military force in democracies." *International Organization* 74(1):119–143.

- Tomz, Michael R and Jessica LP Weeks. 2013. "Public opinion and the democratic peace." *American political science review* 107(4):849–865.
- Yarhi-Milo, Keren. 2016. "Who Fights for Reputation in International Politics? Leaders, Resolve and the Use of Force."
- Yarhi-Milo, Keren, Joshua D Kertzer and Jonathan Renshon. 2018. "Tying hands, sinking costs, and leader attributes." *Journal of Conflict Resolution* 62(10):2150–2179.
- Zacher, Mark W. 2001. "The territorial integrity norm: International boundaries and the use of force." *International Organization* pp. 215–250.
- Zarakol, Ayse. 2014. "What made the modern world hang together: socialisation or stigmatisation." *IT* 6:311.

The Micro-foundations of Naming and Shaming

Supplementary Information

A Survey Instrument	SI-1
A.1 Demographics	SI-1
A.2 Dispositional Measures	SI-1
A.3 Vignette	SI-2
A.4 Mechanism questions	SI-2
A.5 Outcome questions	SI-2
B Descriptive statistics	SI-4
C Manipulation check	SI-6
D Sensitivity Analysis	SI-6
E Heterogeneous Treatment Effects	SI-7
F Placebo Tests	SI-7
G Power analysis	SI-9

A Survey Instrument

Scales that are marked with * have been recoded.

A.1 Demographics

As part of the survey, respondents were asked to report the following demographics: Sex, Age, Ethnicity, Religiosity, Locality, Self-reported vote choice, and position on Left-Right ideological scale

A.2 Dispositional Measures

A.2.1 Militant assertiveness

On a scale of 1 through 7, where 1 indicates “strongly agree” and 7 indicates “strongly disagree”, how much do you agree with the following sentence:

- The best way to ensure peace is through the IDF’s strength. [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree] *
- The use of military force only makes problems worse. [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]
- Going to war is unfortunate, but sometimes the only solution. [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree] *

A.2.2 International trust

- Generally speaking, Israel can trust other nations [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree] *

A.2.3 Nationalism

Please indicate to what extent you agree with each of the following statements from 1 [do not agree at all] to 5 [agree to a great extent]:

- I am proud to be an Israeli
- I identify as a Zionist

A.2.4 Self-reported ideology

On a scale of 1-7, where 1 is left and 7 is right, where would you position yourself ideologically?

A.3 Vignette

A.3.1 Baseline Information

In the next couple of days, Israel is expected to annex many territories from the West Bank.

A.3.2 Shaming treatment

Israel was internationally criticized for this plan. Many countries said Israel should be ashamed of itself.

A.4 Mechanism questions

A.4.1 Reputation

How much damage do you think there would be to Israel's reputation as a result of annexing territories from the West Bank? [A great deal, A lot, A moderate amount, A little, None at all]

A.4.2 Morality

How moral is Israel's plan to annex territories from the West Bank? [Extremely moral, moral, don't know, not so moral, very immoral]

A.5 Outcome questions

A.5.1 Support for Annexation

Do you support annexing large territories from the West Bank? [Strongly support, support, neither support nor objects, objects, strongly objects] *

A.5.2 International Legal Obligation

How strongly do you agree with the following statements: [Strongly agree, agree, neither agree nor disagree, disagree, strongly disagree]

- It is important to me personally that Israel will comply with international law.

- Complying with international law is an important value.
- Complying with international law is important, even if it contradicts the national interest.
- I feel uncomfortable when Israel violates international laws and norms.
- If Israel defies international laws and norms, criticism from other countries is justified.

A.5.3 Manipulation check

Was Israel shamed for its planned to annex territories from the West Bank? [Yes, No, I don't know]

A.5.4 Placebo tests

- Will Israel suffer from economic sanctions due to its plan to annex territories from the West Bank? [Yes, No, I don't know]
- Did you think of a specific country when you were told that Israel was shamed? [Yes, No] if Yes: Please write which country [Open ended]

B Descriptive statistics

Table A1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Female	1,540	0.513	0.500	0	0	1	1
Male	1,540	0.487	0.500	0	0	1	1
18-22	1,540	0.121	0.326	0	0	0	1
23-29	1,540	0.175	0.380	0	0	0	1
30-39	1,540	0.223	0.416	0	0	0	1
40-49	1,540	0.179	0.383	0	0	0	1
50-70	1,540	0.303	0.460	0	0	1	1
Ashkenazi	1,499	0.374	0.484	0.000	0.000	1.000	1.000
Mizrahi	1,499	0.379	0.485	0.000	0.000	1.000	1.000
Russian	1,499	0.176	0.381	0.000	0.000	0.000	1.000
Ethiopian	1,499	0.067	0.250	0.000	0.000	0.000	1.000
Other	1,499	0.004	0.063	0.000	0.000	0.000	1.000
Secular	1,540	0.527	0.499	0	0	1	1
Traditional	1,540	0.306	0.461	0	0	1	1
Religious	1,540	0.138	0.345	0	0	0	1
Haredi	1,540	0.029	0.168	0	0	0	1
Jerusalem	1,540	0.108	0.311	0	0	0	1
Tel Aviv	1,540	0.318	0.466	0	0	1	1
North	1,540	0.264	0.441	0	0	1	1
South	1,540	0.221	0.415	0	0	0	1
Sharon	1,540	0.089	0.285	0	0	0	1
Nationalism	1,525	6.071	1.243	1.000	5.000	7.000	7.000
Reputation	1,521	3.118	1.215	1.000	2.000	4.000	5.000
Moral	1,521	2.913	1.054	1.000	2.000	4.000	5.000
Legal Obligation	1,509	2.892	0.796	1.000	2.400	3.400	5.000
Support Annexation	1,540	3.591	1.759	0	2	5	7

	annex (N=770)	arms (N=759)	Overall (N=1540)
sex_1			
Female	410 (53.2%)	372 (49.0%)	790 (51.3%)
Male	360 (46.8%)	387 (51.0%)	750 (48.7%)
age_1			
18-22	95 (12.3%)	90 (11.9%)	186 (12.1%)
23-29	135 (17.5%)	131 (17.3%)	270 (17.5%)
30-39	169 (21.9%)	171 (22.5%)	343 (22.3%)
40-49	139 (18.1%)	134 (17.7%)	275 (17.9%)
50-70	232 (30.1%)	233 (30.7%)	466 (30.3%)
ethnicity_1			
Ashkenazi	270 (35.1%)	286 (37.7%)	561 (36.4%)
Ethiopian	52 (6.8%)	48 (6.3%)	100 (6.5%)
Mizrahi	285 (37.0%)	278 (36.6%)	568 (36.9%)
Other	4 (0.5%)	2 (0.3%)	6 (0.4%)
Russian	140 (18.2%)	123 (16.2%)	264 (17.1%)
Missing	19 (2.5%)	22 (2.9%)	41 (2.7%)
locality_1			
Jerusalem	89 (11.6%)	78 (10.3%)	167 (10.8%)
North	197 (25.6%)	201 (26.5%)	406 (26.4%)
Sharon	64 (8.3%)	73 (9.6%)	137 (8.9%)
South	178 (23.1%)	162 (21.3%)	341 (22.1%)
Tel Aviv	242 (31.4%)	245 (32.3%)	489 (31.8%)
religiosity_1			
Haredi	13 (1.7%)	32 (4.2%)	45 (2.9%)
Religious	114 (14.8%)	96 (12.6%)	213 (13.8%)
Secular	407 (52.9%)	396 (52.2%)	811 (52.7%)
Traditional	236 (30.6%)	235 (31.0%)	471 (30.6%)
left_right_1			
Mean (SD)	3.10 (1.55)	3.16 (1.61)	3.13 (1.58)
Median [Min, Max]	3.00 [1.00, 7.00]	3.00 [1.00, 7.00]	3.00 [1.00, 7.00]

Figure A1: Balance across treatment and control.

C Manipulation check

Figure A2 demonstrates that treated respondents were more likely to believe that Israel was shamed for its policy.

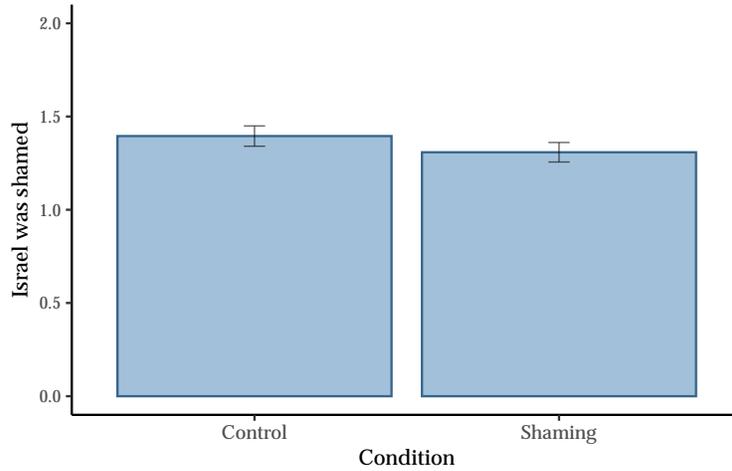


Figure A2: Manipulation check.

D Sensitivity Analysis

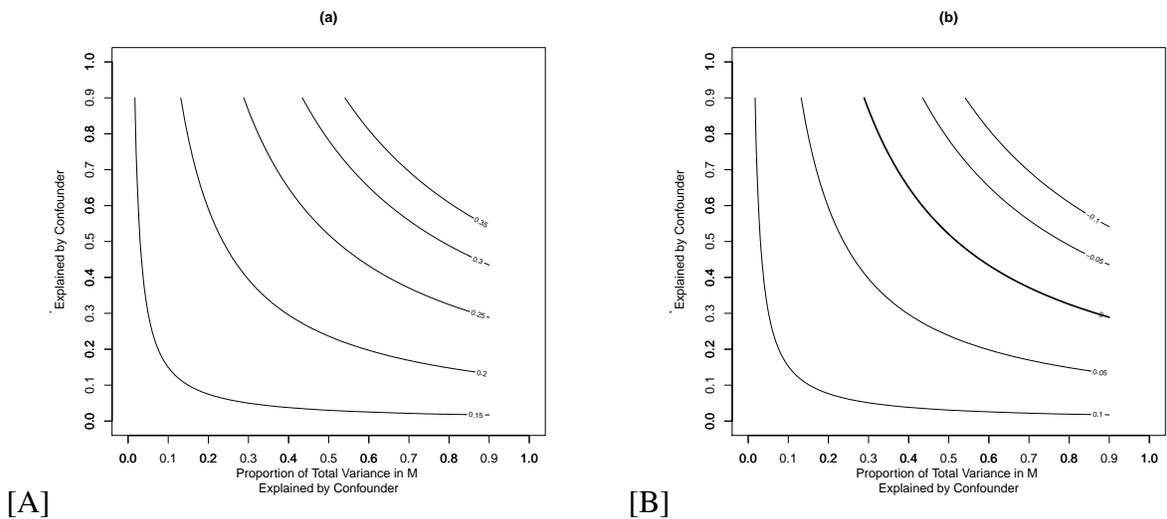


Figure A3: Sensitivity analysis for mediation. In figure A, Treatment is shaming manipulation, Mediator is moral cost (post-treatment), Outcome is support for policy change. In figure B, Treatment is shaming manipulation, Mediator is moral cost (post-treatment), Outcome is support for policy change.

E Heterogeneous Treatment Effects

Table A2: Heterogeneous Treatment Effects on Annexation

	Reputation	
	(1)	(2)
Shaming	0.135 (0.257)	0.169 (0.252)
Hawkishness	-0.335 (0.036)	-0.218 (0.038)
Shaming*Hawkishness	-0.044 (0.050)	-0.050 (0.049)
Demographic Controls	No	Yes
<i>N</i>	1,521	1,521

Table A3: Heterogeneous Treatment Effects of Ideology on Support for Annexation

	Support for Annexation	
	(1)	(2)
Shaming	0.498 (0.186)	0.451 (0.184)
Ideology	-0.328 (0.038)	-0.188 (0.047)
Shaming*Ideology	-0.080 (0.053)	-0.067 (0.053)
Demographic Controls	No	Yes
<i>N</i>	1,529	1,529

In Tables A2 and A3, I explore the moderating effect of hawkishness on reputational concerns, and ideology (left-right scale) on support for annexation, respectively. As depicted in the tables, results are found to be statistically insignificant.

F Placebo Tests

In order to address potential confounders, respondents were asked whether they think Israel has received economic sanctions for its intended annexation. This question is meant to assess whether

other (prevalent) information regarding shaming has “leaked” and bundled the treatment. The result, depicted in the figure, suggest that this association does not exist.

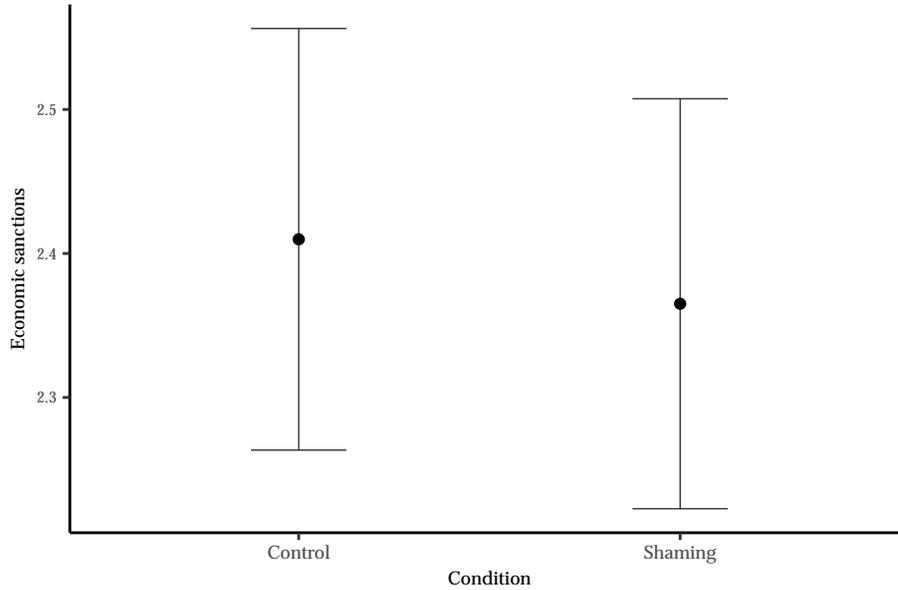


Figure A4: Placebo test. ATE on the likelihood of economic sanctions

Additionally, respondents were asked whether they have thought of a particular country when they have learned of international shaming. This question is meant to explore whether future research should provide more concrete, rather than abstract, information about the identity of the shamer. It appears that only half of respondents who received the shaming treatment have thought of a particular country. A word cloud depicts that most of these subjects thought of European and Arab countries. Future research should attempt to randomize the identity to gain further leverage in the design.



Figure A5: Word cloud of the shamers respondents thought of.

G Power analysis

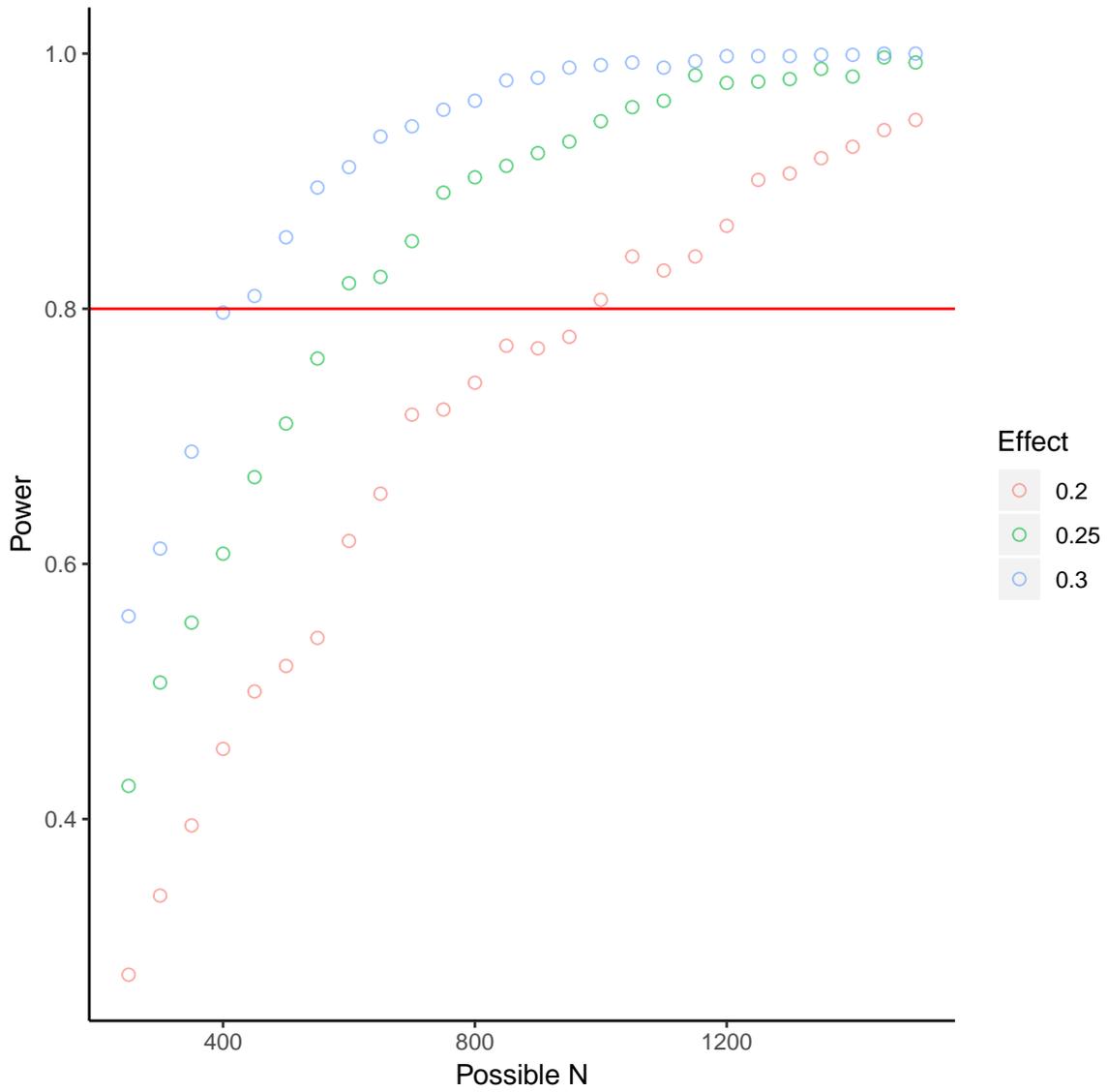


Figure A6: Power analysis, alpha=0.05