

DESEGREGATING DIGITAL SPACES: *A Facebook Field Experiment in Jerusalem**

Alexandra Scacco[†] Alexandra Siegel[‡] Chagai M. Weiss[§]

April 22, 2021
Preliminary Draft

Abstract

Can online exposure to outgroup viewpoints reduce prejudice and promote empathy in deeply-divided societies? What types of exposure are most likely to do so? We address these questions through an online field experiment in Jerusalem, where we assigned Jewish Facebook users to receive posts in their newsfeeds describing Palestinian life in East Jerusalem for a period of two weeks. We find that, in aggregate, exposure did not shift Jewish attitudes toward Palestinians. However, when we consider the effect of different *types* of posts, we find that exposure to content highlighting personal Palestinian experiences improved attitudes towards Palestinians. In contrast, exposure to posts “without a human face” had no effect on outgroup prejudice and reduced the perceived importance of exposure to outgroup perspectives. With respect to behavior, exposure to treatment decreased engagement with additional content highlighting outgroup viewpoints. Taken together, our findings contribute to the prejudice-reduction literature by providing naturalistic evidence of the effects of exposure to outgroup perspectives on attitudes and behavior, and inform the ongoing debate over the political consequences of online “echo-chambers.”

*We thank Thomas Zeitzoff, Pablo Barbará, Salma Mousa, Laura Jakli, Shelby Grossman, Leah Rosenzweig, Justine Davis, Macartan Humphreys, Sascha Riaz, Bernd Beber, Lotem Bassan-Nygate, Hanno Hilbig, Sampada KC, Nina McMurry, Rebecca Littman, Kate Baldwin, and participants in the Institutions and Political Inequality workshop at WZB Berlin Social Science Center, MPSA Identity Across the Subfields,t and the MEI fellows workshop at Harvard University for helpful comments and suggestions. We gratefully acknowledge funding from the WZB and the University of Colorado at Boulder. This research was approved via University of Wisconsin IRB Protocol 2020-1259. Our pre-analysis plan is available via OSF Registries at <https://osf.io/xz2yv/>.

[†]Senior Research Fellow, WZB Berlin. ✉: alex.scacco@wzb.eu

[‡]Assistant Professor, University of Colorado at Boulder. ✉: alexandra.siegel@colorado.edu

[§]Phd Candidate, UW - Madison. Predoctoral fellow, Harvard University. ✉: cmweiss3@wisc.edu

Introduction

Can online exposure to outgroup viewpoints reduce prejudice and promote empathy in deeply divided societies? Does such exposure encourage more interest in outgroup perspectives? Existing research suggests that social media platforms can help democratize the spread of information and facilitate virtual exchanges across political and sectarian divides (Aday et al., 2010; Siegel et al., 2021). However, the homogeneity of many social media networks risks turning online spaces into ideological “echo-chambers,” limiting the potential for information shared on social media to cross group borders and facilitate perspective-taking.¹

We examine the consequences of diversifying content in online environments. To do so, we conducted a field experiment designed in collaboration with *0202 Points of View from Jerusalem* (hereafter “0202”), a Jerusalem-based NGO that translates popular Facebook posts and disseminates them to Facebook users across communities in Jerusalem.² Using Facebook advertisements, we embedded daily posts describing Palestinians’ interactions with Israeli state authorities into the newsfeeds of Jewish Facebook users for a period of two weeks. Treated participants were randomly assigned to receive posts highlighting either personal stories or non-personal stories regarding Palestinian experiences in East Jerusalem. To evaluate the effects of exposure to this content on intergroup attitudes and interest in consuming additional outgroup content, we distributed a short endline survey to participants (1-7 days post-treatment). We also invited participants to engage with additional outgroup content—by watching a short video describing the struggles of a Palestinian from East Jerusalem, and following our partner organization’s organic Facebook page (8-14 days post-treatment).

We find that, on average, exposure to outgroup viewpoints did not improve attitudes toward Palestinians or spark interest in consuming more outgroup content. However, these pooled results obscure important variation. When we consider the effect of different *types* of content, we find that exposure to posts highlighting personal Palestinian experiences improved attitudes towards Palestinians. In contrast, exposure to non-personal posts decreased the perceived importance of exposure to outgroup perspectives.

Turning to behavioral outcomes, we find that daily exposure to outgroup content for a period of two weeks significantly *reduced* online engagement with further outgroup content. This result may be driven by saturation or backlash. According to the saturation explanation, our treatment simply

¹For an overview of the ongoing debate about the consequences of online bubbles or echo-chambers, see (Barberá, 2020).

²This experiment was designed to help 0202 determine whether their activity improves intergroup relations in Jerusalem, and whether refining the content they create can increase Jerusalem residents’ interest in Palestinian current events.

satisfied demand for outgroup perspectives, and therefore treated subjects were less likely to consume additional out-group content. In contrast, a backlash explanation would suggest that Facebook users were unhappy with exposure to Palestinian perspectives, and therefore they avoided future exposure to such perspectives. overwhelmingly hostile anti-Palestinian comments on posts inviting participants to consume more outgroup content provide support for the latter explanation.

Our study contributes to the literature on polarization in online environments. Building on a set of recent studies on polarization in online settings, most of which focus on the United States (Levy, 2020; Guess and Coppock, 2020; Mosleh et al., 2021), we design a novel intervention to consider the effects of diversifying social media environments on intergroup prejudice in Jerusalem—a hyper-segregated city (Rokem, Weiss and Miodownik, 2018) located in the center of an ongoing territorial conflict (Bar-Tal, 2013). Our contextual focus on Jerusalem is important, as offline geographic segregation and linguistic barriers contribute to extreme segregation in Israeli social networks (Mor, Ron and Maoz, 2016; Lissitsa, 2017; Harel, Jameson and Maoz, 2020).

In addition, we provide new insight regarding the diverging effects of different types of content. In contrast to recent studies which diversify online environments by encouraging users to follow organic accounts (Bail et al., 2018; Levy, 2020), we employ advertisements and embed carefully curated posts into the newsfeeds of consenting study-participants. Doing so allows us to unbundle previously identified treatments, and consider how exposure to personal and non-personal Facebook posts has diverging effects on attitudes and behaviors. Our research design demonstrates the potential of using Facebook’s advertising tools and messenger bots to randomly assign consenting participants to receive content in a naturalistic setting, and to measure attitudinal and behavioral effects, all in a replicable and scalable way.

We also engage with the literature on intergroup prejudice, by designing and testing a scalable intervention in a setting of ongoing intergroup conflict (Paluck et al., 2020). Our findings are somewhat discouraging, as they suggest that in general, mere exposure to online content from the outgroup perspective does not reduce prejudice, and reduces interest in and engagement with outgroup content. However, our findings point to an important pattern by which exposure to personal posts, as opposed to non-personal posts, increases affect and empathy towards the outgroup.

Jerusalem: The Empirical Context

As depicted in Figure 1, Jerusalem is an ethnically segregated and politically divided city. The origins of this segregation are rooted in the aftermath of the 1948 war, when control of the city was divided between Israeli and Jordanian rule. Following the 1967 war, Israel gained control over the Eastern-Palestinian component of the city, and later annexed this territory granting Palestinians

residency rather than citizenship status (Bollens, 2000). Inequality, segregation, and linguistic barriers largely limit intergroup interactions between Jews living in West Jerusalem and Palestinians living in East Jerusalem. Existing literature suggests that the low frequency of meaningful interactions across communities likely serves to further impair the already tense intergroup relations in a highly contested city (Kasara, 2013, 2014; Weiss, 2020).

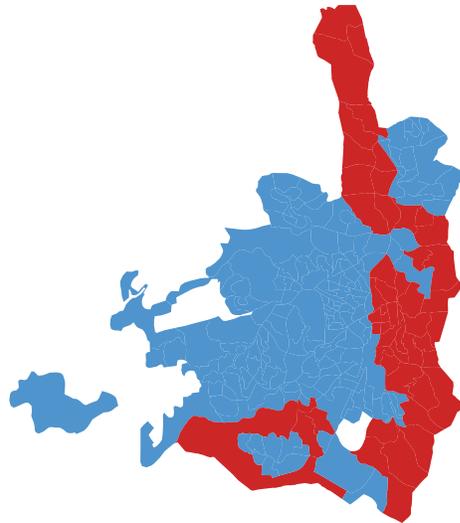


Figure 1: **Residential Segregation in Jerusalem** – This map depicts statistical areas in Jerusalem (an administrative unit equivalent to neighborhood). Blue statistical areas are predominantly Jewish, whereas red statistical areas are predominantly Palestinian.

A core challenge for intergroup relations in segregated cities like Jerusalem is that they provide few meaningful opportunities to learn about outgroup members. This challenge is exacerbated when group cleavages include a linguistic dimension. Not only is Jerusalem a physically segregated city, but recent research suggests that Jewish-Palestinian online spaces are highly segregated as well (Mor, Ron and Maoz, 2016; Lissitsa, 2017; Harel, Jameson and Maoz, 2020). Survey research suggests that the frequency of Israeli Jews’ contact with Arabs is extremely limited. In a 2017 representative survey of Israeli Jews who regularly use social media, about three-quarters of respondents reported no online contact with Arabs whatsoever; 17% reported low frequency of online contact, 6% reported medium frequency and only a small minority (3%) indicated high or very high frequency of online contact with Arabs (Lissitsa, 2017).

Our study partner is a Jewish-Palestinian NGO that seeks to overcome this challenge by exposing Jerusalem residents to a more diverse set of Facebook posts about community and daily

life in the city. The organization’s mission is to provide an unfiltered glimpse into the daily life of different communities in Jerusalem. To that end, they translate Arabic-language Facebook posts about local Jerusalem experiences into Hebrew (and vice versa) and post this content daily in their 35,000+ Facebook followers’ newsfeeds. As we describe below, our study evaluates how Jewish Jerusalem residents who do not already follow the 0202 Facebook page react to exposure to this type of content. Findings from this evaluation contribute to ongoing academic debates regarding the consequences of online “echo chambers” (Barberá, 2020), and bear important policy implications for 0202 and similar NGOs which seek to employ effective strategies for disseminating online knowledge and promoting interest in outgroup content, and empathy toward the outgroup.

Israel leads the world in social media use, with 77% of the population using social media as of 2019 (Taylor and Silver, 2019). Social media marketing data suggests that around 85% of Israeli internet users regularly use Facebook, and Israelis spend an astounding average of eleven hours daily on social networks, which is almost double the world average of six hours (Stats, 2018). As a result, Facebook provides an ideal platform on which to test our intervention on Jerusalem residents.

Theoretical Motivation and Expectations

A rich literature in political science has considered the effects of physical segregation on intergroup relations. Existing evidence suggests that residential segregation, for example, is detrimental for intergroup relations (Enos, 2017), as it limits the flow of information across groups (Weiss, 2020), reduces opportunities for shared experiences, erodes trust (Kasara, 2013), and can promote violence (Kasara, 2014). Segregated *online* environments, often referred to as “filter bubbles” or “echo chambers” (Bakshy, Messing and Adamic, 2015; Del Vicario et al., 2016; Cinelli et al., 2021), are similarly believed to foster intolerance across social group lines and exacerbate political polarization, for at least two reasons.

First, in such environments, social media users are exposed to an onslaught of ingroup-oriented content. Particularly in deeply divided societies, this content may reinforce existing negative beliefs about outgroup members, or signal exclusionary and intolerant social norms (Munger, 2017; Siegel and Badaan, 2020; Siegel et al., 2021). Recent evidence suggests that in-person interactions in homogeneous environments can increase discriminatory behavior against outgroups (Scacco and Warren, 2018). Therefore, homogenous environments may impair intergroup relations as a consequence of increased exposure to and engagement with ingroup members and the content they create.

Second, homogeneous online environments shield social media users from outgroup members and their points of view. The absence of outgroup members in homogenous online environments limits opportunities for the kinds of friendship-formation highlighted as crucial for prejudice-reduction in the literature on intergroup contact (Scacco and Warren, 2018; Mousa, 2020; Lowe, 2020; Weiss, 2021; Rossiter, 2020), and limits occasions to develop empathy for outgroup members and practice perspective-taking (Kalla and Broockman, 2021; Simonovits, Kezdi and Kardos, 2018; Adida, Lo and Platas, 2018; Broockman and Kalla, 2016), each of which has been shown to improve intergroup relations in online and offline contexts.

Systematic empirical evidence for a link between homogenous online environments, prejudice, and polarization remains thin on the ground, however, and what evidence there is about the consequences of online exposure to outgroup or counter-attitudinal content points in opposite directions.³ Levy (2020) demonstrates that exposure to counter-attitudinal news in the United States can improve attitudes toward the opposing political party. Similarly Siegel et al. (2021) finds that Egyptian Twitter users embedded in more diverse networks express more tolerant views over time. In contrast, Bail et al. (2018) provide evidence from the U.S. context that exposure to opposing political views on Twitter may actually increase ideological polarization. Lastly, Guess and Coppock (2020) find little to no effect of exposure, in three large-scale survey experiments.

Given these disparate findings about the effect of exposure to counter-attitudinal perspectives in recent studies, we do not have clear prior expectations about how exposure to outgroup content will affect attitudes and behaviors in the Israeli-Palestinian context. On the one hand, online exposure to the perspective from Palestinians living in the same city might generate empathy and spark interest in consuming additional information about outgroup viewpoints among Jewish Jerusalem residents. On the other hand, we might expect backlash if exposure reinforces negative beliefs, triggers concerns about outgroup threat, or encourages Jewish study subjects to justify or defend the behavior of Israeli state authorities in East Jerusalem. These competing possibilities motivate our first two bi-directional hypotheses.

H1: Exposure to outgroup viewpoints will improve [worsen] attitudes toward the outgroup.

H2: Exposure to outgroup viewpoints will increase [decrease] interest in and engagement with additional content highlighting outgroup perspectives.

But exposure to different types of outgroup content may have varying effects. Even if posts about life in East Jerusalem cover the same topics, the way they are framed may shape the impact

³See Barberá (2020) for a recent review.

of exposure. Here, we focus on the distinction between personal or human-centered stories in contrast to non-personal stories that highlight outgroup experiences more generally. Human-centered frames bring “a human face or an emotional angle to the presentation of an event, issue, or problem” (Semetko and Valkenburg, 2000). They are thought to generate empathy by use of human examples, adjectives or personal vignettes, consideration of how individuals are affected, personal or private information, and emotionally evocative visual content (Figenschou and Thorbjørnsrud, 2015). On the one hand, we expect personal stories to be particularly effective in reducing prejudice and promoting perspective-taking, as they have been shown to generate empathy, higher levels of interest in policy issues, and greater engagement online and offline (Trilling, Tolochko and Burscher, 2017; Grabe et al., 2017). On the other hand, if exposure to outgroup content provokes hostility or backlash, it is possible that exposure to more evocative personal stories will prompt an even stronger negative response than non-personal stories. This motivates our second set of bi-directional hypotheses:

H3: Exposure to personal outgroup viewpoints will be more [less] effective than non-personal content in improving [worsening] attitudes, empathy and perspective-taking toward the outgroup.

H4: Exposure to personal outgroup viewpoints will be more [less] effective than non-personal content in encouraging interest in and engagement with additional outgroup content.

Research Design

To test whether exposure to outgroup viewpoints shapes attitudes and behavior, we designed and implemented a field experiment in collaboration with our partner organization, 0202 Points of View from Jerusalem. The goal of the experiment was to evaluate the routine activities of 0202, and determine whether and how their work may affect Jerusalem Facebook users’ attitudes and online behavior. Our experiment followed four steps, including: (1) a baseline survey, (2) a 14 day Facebook intervention exposing treated participants to content from 0202, (3) an endline survey, and (4) two post-treatment behavioral measures. Figure 2 summarizes this design, which we describe in more detail below.

This design allows us to make several primary comparisons.⁴ First, we can assess the aggregate effect of exposure to outgroup viewpoints by comparing respondents who received any type of such content (personal and non-personal) with the pure control group (which received no content).

⁴Each of these comparisons was pre-registered as part of our primary analysis: <https://osf.io/xz2yv/>



Figure 2: **Primary Experimental Design**

We can consider the result of this comparison as the overall effect of exposure to content from an outgroup perspective. Second, we can evaluate the relative impact of the *type* of outgroup content by comparing the effects of exposure for subjects treated with personal versus non-personal posts. We can consider the result of this comparison as the effect of exposure to personal outgroup content. In order to determine whether any effects of personal and non-personal posts work in the same or in opposite directions, we also compare the effect of each type of outgroup news with the pure control group.

Recruitment and Baseline Survey

In order to recruit participants for our study, we targeted Jewish Facebook users from West Jerusalem with advertisements in Hebrew inviting them to participate in two waves of a short social survey. To encourage participation, we offered study participants the opportunity to enter a raffle to win an iPhone as well as ten 50 NIS vouchers.⁵ Respondents who clicked on our advertisement were redirected to Facebook Messenger, where we administered our survey natively within Messenger using a digital marketing platform called SurveyBot.⁶

The baseline survey opened with a consent form. If respondents consented to participate, they were then asked a short set of demographic questions, feeling thermometer questions about sev-

⁵This recruitment strategy is not expected to yield a representative sample of Facebook users from Jerusalem. However, evidence regarding the similarity of average treatment effects implemented on convenience and representative samples (Mullinix et al., 2015) increases our confidence that even if our sample is not representative of the population of Facebook users in Jerusalem, our average treatment effects will be a meaningful quantity of interest.

⁶For more information about SurveyBot see: <https://surveybot.io/>.

eral types of outgroups, questions relating to empathy and perspective taking, and items relating to social media consumption. The survey was designed to take no more than ten minutes to complete. A description of all items collected in our baseline survey is depicted in Section S2 of the supplementary materials.

Our study recruitment advertisement reached 273,219 unique Facebook users living in Jerusalem. This is about 40% of the 700,000 estimated Facebook users living in Jerusalem.⁷ 8,092 people clicked on the ad to take our baseline survey. This represents a 3% click through rate, about three times the average click through rate for ads across industries.⁸ 6,153 of these individuals consented to participate in our study. Of the 6,153 people who consented to participate in our study, 4,532 answered at least one baseline survey question and 3,041 completed all 23 questions at baseline.

Treatment

To administer treatment, we focused on Jewish Facebook users from Jerusalem who consented to participate in our study, even if those consenting users did not complete our baseline survey, and randomly assigned them to receive 14 posts in their Facebook newsfeeds over a period of two weeks (approximately one post per day). Users were block randomized by self-reported political leaning,⁹ to receive either: (1) Facebook posts about Palestinians’ interactions with Israeli that center the stories of individuals (personal posts), (2) Facebook posts about about Palestinians’ interactions with Israeli authority that do not highlight individual stories (non-personal posts), or (3) no Facebook posts at all.¹⁰ Facebook posts with and without “a human face” were selected to focus on the same substantive issues facing East Jerusalem residents including arrests, destruction of property, targeted violence, harassment by state authorities, and other difficult features of daily life. We present an example of each type of post in Figure 3, and the full set of post in Section S1 of the appendix. A detailed discussion of the criteria we used to select the 28 posts employed in our experiment is reported in Section S1 of the appendix.

Because the SurveyBot platform stored unique identifiers for each person who consented to participate in our study, we were able to target each Facebook user in our sample with personal or

⁷Estimate provided by Facebook’s Ads Manager.

⁸<https://bit.ly/39w9n50>

⁹Our right-left ideology score ranges from 1-7, and study subjects who did not respond to this question were randomized as an eight block into treatment and control conditions.

¹⁰We randomly assigned a subset of Facebook users to receive “organic” content (without curating posts to focus only on Palestinian interactions with Israeli authorities, but addressing a wide range of topics, including sports, culture, and other content) created by our partner organization for partner program evaluative purposes, but as noted in our pre-analysis plan, we do not consider these respondents in our main analyses.



(a) Personal Post



(b) Non-Personal Post

Figure 3: **Example of 0202 Facebook Post**– the personal post provides information about a Palestinian women who was investigated for a seventh time by the police for operating an unauthorized news station. The non-personal post provides information about Jerusalem municipality officers confiscating vegetables from a local market in Shuaafat refugee camp.

non-personal ads depending on their treatment assignment. Targeting individuals with Facebook advertisements does not guarantee that the ads will appear on their newsfeeds, but we optimized spending to increase visibility of our ads. According to Facebook Ads manager metrics, 92% of our treated participants saw our ads over course of the treatment period, and these users saw an average of 14.3 posts, suggesting that the vast majority of our treated users indeed received all 14 days of treatment.¹¹

Our treatment ads received high levels of engagement relative to typical Facebook ads. The click through rate for our ads, which includes all forms of engagement, was 12%, more than 12 times the average click through rate of promoted posts on Facebook. Additionally, our ads received more than 7 times as many comments as comparable posts on 0202’s own Facebook page aimed at Israeli Jews, which has over 35,000 followers—about 7.5 times the number of people we assigned to treatment. Taken together, these figures offer reassurance that our treatment was delivered effectively to study participants, which engaged with our experimental content in relatively high rates.

Outcomes

One day after the intervention period, we invited all treated subjects to participate in an endline survey. In order to increase response rates, we sent reminders to subjects for seven additional days. Our endline includes measures of attitudes toward the outgroup, including: prejudice, empathy, perspective-taking, and interest in outgroup-oriented media. Like the baseline survey, the endline was administered natively within Facebook Messenger using SurveyBot.

Out of a total of 6,153 treated Facebook users, 1,270 users responded to at least one question on the endline survey. In Section S4 of the appendix, we address potential concerns that might arise from attrition in three ways. First, we demonstrate that treatment status does not predict participation in our endline survey. Second, we show that treatment status does not predict non-response to particular questions on our endline survey. Third, we use baseline survey data to examine correlates of participation in the endline. We are reassured by the fact that, although age and gender each predict participation in the endline, political ideology, religiosity and affect towards Palestinians (measured with a feeling thermometer) do not. A detailed description of all attitudinal outcomes used in the main analysis is reported in Table 1.

We supplemented our endline survey with two behavioral measures tracking whether users click on promoted advertisements embedded in their Facebook newsfeeds. As reported in Figure 4, these promoted advertisements asked respondents to (1) click in order to follow a Facebook page

¹¹In some cases, study subjects saw the same post more than once on their Facebook timelines.

<i>Outcome</i>	<i>Question Wording</i>	<i>Possible Responses</i>
Affect	On a scale of 1-10, where 1 resembles very cold feelings, and 10 resembles very warm feelings, how do you feel about the following group of people living in Jerusalem: Palestinians	1-10
Peace	Do you agree, or not agree with the following statement: "Most Palestinians in Jerusalem want to live in peace"	Four point scale
Empathy (Index)	<p>1. On a scale of 1-4, to what extent does the following describe your views: "It is important for me to try and understand Palestinians"</p> <p>2. On a scale of 1-4, to what extent does the following describe your views: "Even if I disagree with Palestinians, I try to think why they take different positions than me"</p>	Average of two four point scales
Exposure (Index)	<p>1. Agree/disagree: "Exposure to diverse opinions on social media is valuable"</p> <p>2. Agree/disagree: "I would benefit from more exposure to news about Palestinian life in East Jerusalem"</p>	Average of two four point scales

Table 1: **Attitudinal Outcomes**



(a) Like 0202 Page Ad



(b) Watch Video Ad

Figure 4: **Behavioral Measures** — The left-hand side advertisement states: “Interested in learning what is happening in East Jerusalem? Follow 0202 Points of View from Jerusalem in order to receive daily updates regarding Palestinian life in Jerusalem.” The right-hand side advertisement states: “Palestinians in East Jerusalem experience many challenges in their daily life. Take a moment to view a video about these challenges.”

sharing news from East Jerusalem, and (2) click in order to watch a video sharing the perspective of an East Jerusalem Palestinian resident. Both behavioral outcomes were measured from 8 to 14 days after treatment ended.

Ethics

Before describing our estimation strategy, we discuss ethical considerations as they pertain to our experimental design. First, during the process of recruitment for our study, subjects were informed when asking for their consent that they would be taking part in an experimental study in partnership with a Jerusalem-based NGO whose mission is to connect members of different communities in Jerusalem through shared news about daily life in the city. During the consenting process, subjects were told in a direct way that the study “might expose you to information about local news and events in Jerusalem via Facebook” and that the researcher “will consider how you react to different prompts on Facebook,” such as “how you respond to a prompt inviting you to follow specific pages or RSVP to events on Facebook.” The translation of our consent form is presented in Section S5 of the appendix.

Second, while the ethics of Facebook advertisements are a subject of current debate among social scientists and practitioners (Guess, 2021), receiving advertisements is part of the standard

experience for Facebook users, who see many ads and sponsored posts each day in their newsfeeds. These ads span a wide range of topics, from products advertised by retail stores and pharmaceutical companies, to advertisements from political campaigns, dating apps, charities, and religious organizations soliciting donations and recruiting members. Facebook users are generally aware that advertisements target individuals based on their demographic characteristics, previous purchasing habits, online browsing activity, and other metadata. It is common to receive advertisements from the same sources for days, weeks, or months at a time. The experience of receiving our sponsored treatment posts is therefore well within the normal range of Facebook user experiences. Furthermore, if people wish to opt out of receiving advertisements from a certain source (such as our treatment page), they are able to do so by adjusting the settings in their Facebook profiles. In practice, we did not see evidence of study participants opting out of treatment, but they were free to do so. Because we obtained informed consent at the outset and subjects had the opportunity to opt out of treatment and the study at any time, we were able to respect their autonomy throughout the experiment.

Third, the aim of our partner organization’s work is to reduce intergroup prejudice by translating and disseminating outgroup-focused content online, and this was our own hope as researchers. But it is important to acknowledge that several recent studies suggest that diversifying social media environments might elicit negative online responses (Bail et al., 2018). For that reason, evaluating 0202’s activity in light of the potential for backlash offers important insights that can inform the organization’s future activities. With this possibility in mind, we took several steps to minimize the potential for harm, including the spread of online hate speech. First, given disparities in power and resources across communities within Jerusalem, we chose to limit our study to Jewish residents of Jerusalem, rather than exposing members of the more vulnerable community of Palestinians in East Jerusalem to content they might prefer to avoid. Second, we chose to disable public viewing of user comments on 0202 treatment posts and on posts inviting respondents to participate in post-treatment behavioral measures. As such, we strove to conduct this experiment in as naturalistic a way as possible, while controlling any potential public backlash.¹²

Estimation Strategy

To identify the effects of our treatments, we use an OLS regression with block fixed-effects, where we regress our outcomes of interest over a treatment indicator (See equation 1). Throughout our analyses, we consider four different treatment indicators, which are used to compare respondents assigned to: (1) pooled (personal and non-personal) 0202 treatment vs. pure control, (2) personal

¹²We received IRB approval from the University of Wisconsin IRB office (Approval #: 2020-1259).

treatment vs. control, (3) non-personal treatment vs. control, and (4) personal vs. non-personal 0202 treatments.

$$y_{ib} = \beta_{treatment} + \gamma_b + \varepsilon_{ib} \quad (1)$$

We initially intended to focus our analyses on variables measured in both pre- and post-treatment surveys, and have y_{ib} denote the difference in a given survey measure, between our pre- and post-treatment periods. However, because roughly half of our endline respondents did not complete pre-treatment surveys, in our main analysis we consider post-survey responses (rather than pre-post differences) as outcomes. We expect y_{ib} to be affected by our treatment indicator $\beta_{treatment}$. γ_b represents our block fixed effect, and ε_{ib} is our model's error term.

Results

In Figure 5 we report treatment effects on our main measures of attitudes toward the outgroup — affect toward Palestinians, beliefs about whether Palestinians want to live in peace, and an empathy index (composed of one measure of empathy for the outgroup and one measure of respondents' beliefs about the importance of perspective-taking). In the upper-left panel (red), we consider the overall effect of receiving outgroup media content (personal or non-personal). We find no evidence that our pooled treatments affected general affect towards Palestinians, beliefs that most Palestinians want to live in peace, or empathy towards Palestinians. These pooled results obscure important variation, however.

When we consider the effect of exposure to different *types* of outgroup media content, we find that exposure to stories highlighting the experiences individual Palestinians improved attitudes towards Palestinians and increased the perceived importance of exposure to outgroup perspectives, when compared with exposure to non-personal outgroup media content. In the upper right panel (blue), we demonstrate that, when compared with non-personal content, exposure to personal content increases respondents' affect and empathy towards Palestinians. The effect on our measure of belief that Palestinians want to live in peace is not statistically significant.

In the bottom two-panels, we consider whether these results are driven by the impact of personal or non-personal posts, by comparing respondents' assigned to the personal (green) and non-personal (purple) conditions, with respondents from the pure control group. The increase in positive affect towards Palestinians is largely driven by the positive effects of exposure to personal content. Indeed, when comparing respondents from the personal treatment condition with respon-

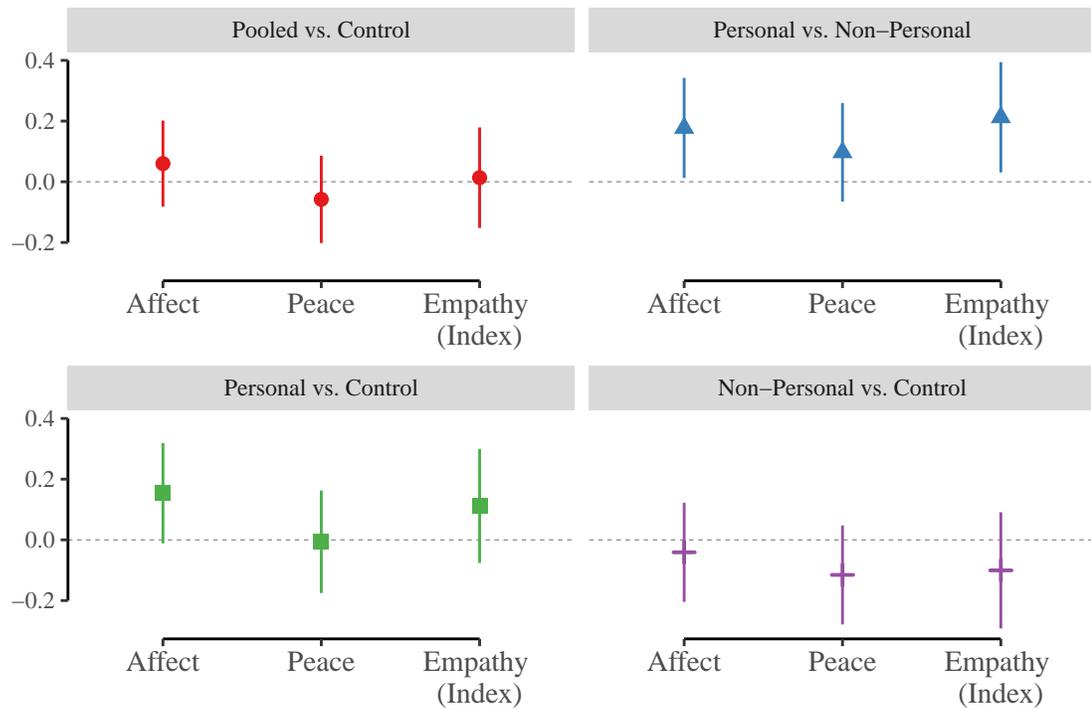


Figure 5: **Treatment Effects on Attitudinal Outcomes** – Each point estimate is extracted from an OLS regression with block fixed effects, where we regressed an outcome of interest (e.g. Affect, Peace, Empathy), over a treatment indicator (e.g. control vs. pooled, control vs. personal, control vs. non-personal, personal vs. non-personal).

dents from the control condition, we identify a positive increase of one sixth of an SD in affect ($p < 0.07$). The effect of our non-personal treatment on intergroup affect in the bottom-left panel, is negative but not statistically significant.

Additionally, the point estimates for our empathy outcome in both bottom panels are not statistically significant. However, the coefficient sign on personal content is positive, whereas coefficient sign on non-personal content is negative. We interpret this finding, to suggest that our increase in empathy towards Arabs (in the personal vs. non-personal comparison), is likely driven by the positive effects of personal content, and the negative effects of non-personal content.

In Figure 6, we consider whether our treatments shaped respondents' interest in consuming additional content highlighting outgroup perspectives. Our results suggest that treated respondents are actually *less* likely to report that exposure to outgroup media offering different opinions is valuable. Indeed, we observe a decrease of over a tenth of a SD in responses to our exposure scale, among respondents in the pooled treatment condition ($p < 0.06$).

We further investigate whether the type of outgroup posts is driving these general effects. We find that decreased interest in outgroup content is likely driven by the significant negative effect of exposure to non-personal media content. As Figure 6 demonstrates, the point estimate of our personal treatment (compared to the control group) on the exposure index is very small and imprecisely estimated. In contrast, the effect of communal content (compared to the control group) on the exposure index is negative and equivalent to more than a quarter of a SD decrease. This suggest—in line with our pooled findings (red point estimate)—that exposure to non-personal content decreases the value that Facebook users see in exposure to different opinions online.

In Figure 7, we turn to examine whether these attitudinal patterns are in line with respondents' revealed preferences for additional exposure to outgroup viewpoints. Figure 7 displays the effect of exposure to 0202 posts on individuals' interest in engaging with outgroup content. Since differences between the personal and non-personal content are statistically insignificant, we focus on the general effects of out-group exposure (i.e. our pooled comparison). Figure 7 displays a *decrease* in study subjects' engagement with additional outgroup viewpoints by following the 0202 Facebook page. While the effect on study subjects' willingness to watch a video on Palestinian experiences in East Jerusalem is not statistically significant, the direction of the effect is the same. Looking at both behavioral measures together (the combined effect) we also see a statistically significant decrease in engagement with additional outgroup content.

At first glance, our behavioral results suggest two competing interpretations. First, they may reflect a saturation effect—people who received content expressing outgroup viewpoints were simply no longer interested in engaging with additional outgroup content because their demand for such

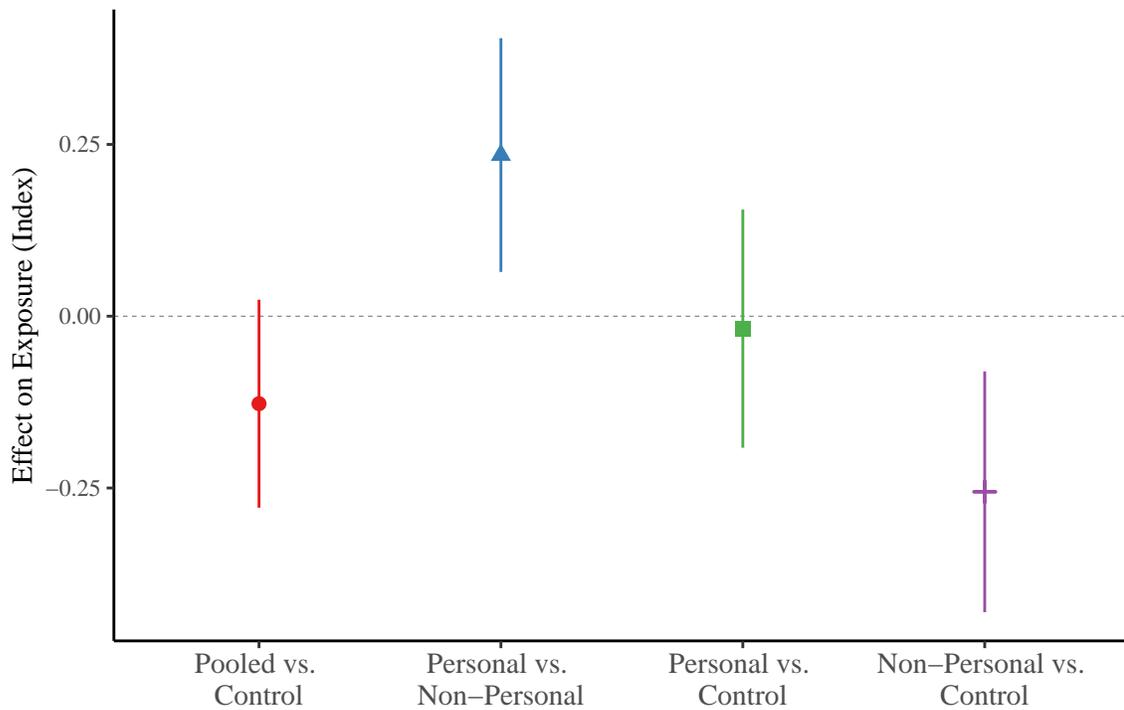


Figure 6: **Treatment Effects on Attitudes regarding Exposure to Outgroup Content** – Each point estimate is extracted from an OLS regression with block fixed effects, where we regressed our exposure index, over a treatment indicator (e.g. control vs. pooled, control vs. personal, control vs. non-personal, personal vs. non-personal).

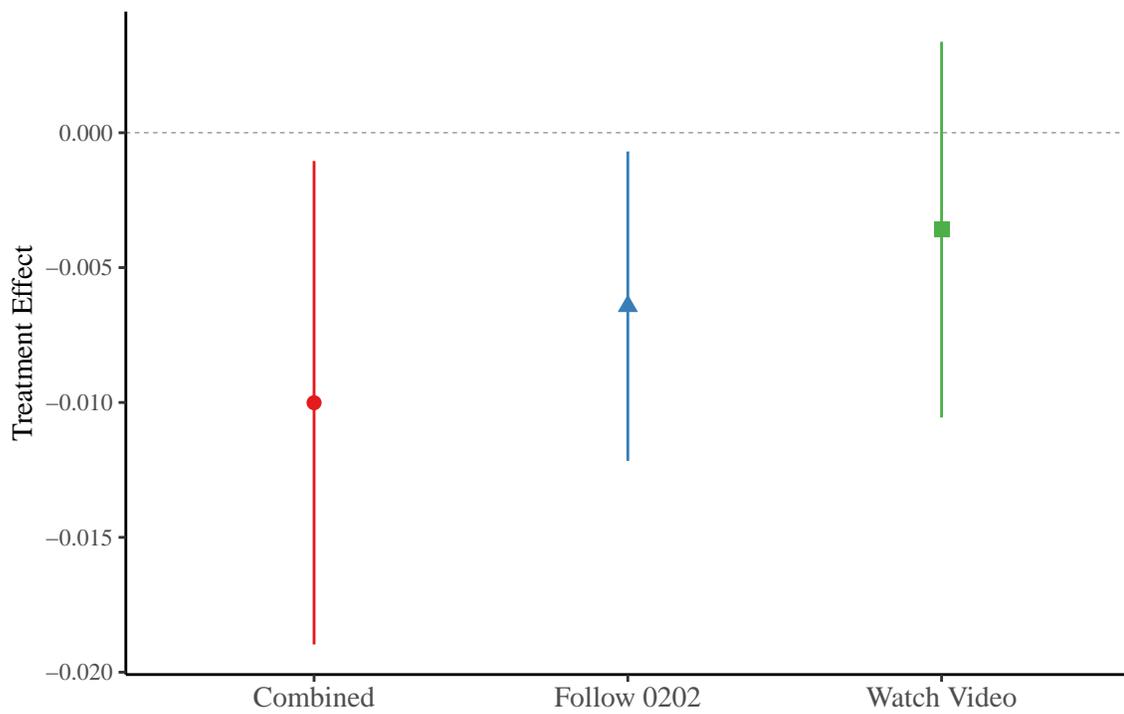


Figure 7: **Treatment Effects on Behavioral Outcomes** – Each point estimate is extracted from an OLS regression with block fixed effects, where we regressed a our behavioral outcome over our pooled treatment indicator.

information had been met after two weeks of posts from 0202. Second, the behavioral results may reflect a backlash effect.

Although either interpretation is possible, comments left by study subjects on the behavioral ads (invitations to like the 0202 page and to watch a video about daily life in East Jerusalem) suggest that diversifying online environments may lead to backlash, at least among some Facebook users. Comments on both the advertisement to like the 0202 page and the video ad, point to hostility rather than fatigue or satisfaction of demand for this type of content.¹³

Translated Comments on 0202 Page Ad:

- *All the stories will not do well for you. You are the invaders. You are not Palestinians, you are Israeli citizen traitors.*
- *Why do we need to get updates online? We hear shooting all day. We smell fires and enjoy the company of their youth that attend our places with manners of a wild horse.*
- *They are welcome to leave ... Whoever is living by them is suffering*

Translated Comments on Video Ad:

- *The videos are biased and untruthful. Enough with the Palestinian terror.*
- *Trying to understand if because of these challenges we hear shooting and find bullets in our porches and windows and if because of this...young people in their 40s drive by with loud music and I won't continue because everyone knows what happens*
- *All of Jerusalem belongs to the Israeli people*

The limited number of comments left on our ads makes it difficult to determine with confidence whether our treatment directly caused these negative responses. The tone and content of the comments above is consistent with comments left on the 0202 treatment posts during the two week intervention period. As before, comments left on our treatment ads were uniformly negative, ranging in severity from mockery of to outright hostility toward Palestinians as a group.¹⁴ Breaking the comments down by treatment assignment, we received 33 comments on personal 0202 treatment posts and 25 comments on non-personal posts.

¹³These uniformly anti-Palestinian comments stand in contrast to comments targeted Facebook users posted on our baseline survey invitation advertisement, for example, which expressed irritation for having been shown the ad too many times.

¹⁴It is important to note that our intervention did not provide information that would allow study participants to identify the Facebook accounts or other contact details of authors of 0202 posts.

An important feature of our design that merits discussion is our decision to make comments invisible to other participants in our study. We made this decision in order to minimize the potential for harmful or abusive exchanges among study participants. Because we hid comments from participants in our study, each commenter on a treatment post believed their comments were being made publicly, but viewed themselves as the first person leaving a comment.¹⁵ Representative examples of comments left by study participants on 0202 treatment posts (personal and non-personal) are displayed in Section S6 of the Appendix.

The content of comments we received on our treatment posts and behavioral measures suggests that, at least for some of subjects, our treatment elicited negative reactions. But why did the content of our study ads spark this type of reaction? Although our study design does not allow us to provide a definitive answer, we suspect this may be for several reasons.

First, part of 0202's mission as an organization is to provide an "unfiltered" look at media content generated by communities in Jerusalem that are often in conflict with one another. As such, popular posts created by Palestinians living in East Jerusalem are simply translated from Arabic into Hebrew and made publicly available on 0202's page aimed at Jewish Facebook users. In designing our intervention, we followed this principle and did not add disclaimers to "soften" the treatment messages (though, as we discuss in the appendix, we did not use any posts in our study that included graphic or potentially disturbing images). We did not alter the language of the posts in any way. As such, a number of personal and non-personal posts used terms like "occupation," "occupying force," or "repression" in describing their interactions with the Israeli state and its authorities. Such language may have alienated Jewish Facebook users unaccustomed to seeing this language in their typical engagement with Facebook content.

This unfiltered approach to diversifying online content differs from a number of studies in the emerging literature in political science on intergroup perspective-taking and perspective-getting, where content is curated and presented to study subjects in ways that explicitly establish a "non-judgmental context" before sharing narratives from outgroup members. In their research on perspective-getting, Kalla and Broockman (2021) discuss the type of unfiltered approach to sharing stories that we employ here as largely untested and worthy of exploration in future research.

Second, the fact that Jerusalem is a deeply divided city in the middle of an ongoing territorial conflict, cannot be overlooked. It is possible that, even if we had decided to soften or curate the language used in our study's treatment post that the reaction to this content would have been largely negative. The intensity of conflict and antagonism across the Jewish-Palestinian divide may simply not be comparable to settings like the United States, even given high levels of political polarization

¹⁵Public pages and individuals frequently hide comments on ads and posts on Facebook in this way. We automated this process by hiding any comments on content from our page that contained Hebrew stopwords, profanity, or emojis.

there today.

We also consider a third possible reason for the negative content of these (seemingly) public comments, recalling the radical segregation of online networks in Israel. Given highly homogeneous Facebook networks, comments might reasonably be considered as public signals to fellow ingroup members. From the viewpoint of study participants, their comments are likely to be seen by friends, family members, and other Hebrew-speaking Facebook users. Indeed, we observe here that many of the comments explicitly defend or praise the ingroup while denigrating the outgroup.

The social psychology literature on intergroup relations suggests that ingroup members may use public displays of outgroup prejudice to signal their belonging to their ingroup. As Brewer (1999) explains in their review of forty years of prejudice research, “Many discriminatory perceptions and behaviors are motivated by the desire to promote and maintain positive relationships within the ingroup rather than by any direct antagonism toward outgroups.” Indeed, ingroup favoritism has been shown to be a stronger driver of prejudicial behavior than hostility toward the outgroup in a wide variety of settings.¹⁶

Exploratory analysis of our survey data on the impact of our treatment on ingroup attitudes provides additional suggestive evidence that exposure to our 0202 treatments may have intensified feelings of ingroup loyalty. As Figure 8 suggests, participants who were exposed to outgroup media content were more likely to report higher levels of ingroup affect. These patterns are in the positive direction for all treated individuals, though only the effect of exposure to personal versus non-personal content is precisely estimated ($p < .05$). In short, much as we observed for outgroup affect, we see that personal posts increased ingroup affect in comparison to non-personal posts. Moreover, examining differences in commenters’ baseline characteristics relative to treated individuals who did not comment on posts, we see that commenters have higher ingroup affect than non-commenters, offering further suggestive evidence.

Conclusion

In this paper, we consider the effects of desegregating social media environments. To do so, we report results from a Facebook field experiment in which we randomly assigned Jewish Facebook users living in Jerusalem to receive content about life in East Jerusalem from a Palestinian perspective for a period of two weeks. We find that, in aggregate, exposure to outgroup Facebook posts did not shift Jewish attitudes toward Palestinians. When we consider the effect of different *types* of posts, we find that exposure to content highlighting personal Palestinian experiences

¹⁶See Greenwald and Pettigrew (2014) for an overview of research in this area.

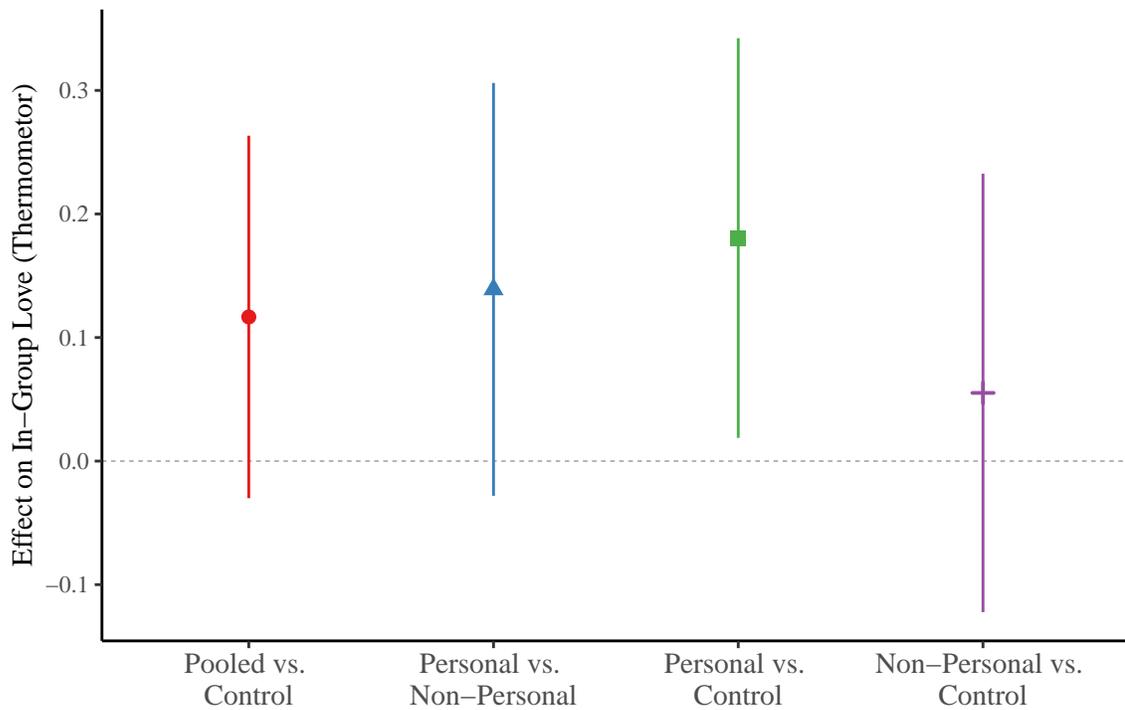


Figure 8: **Treatment Effects on Outgroup Hate and Ingroup Favoritism** – Each point estimate is extracted from an OLS regression with block fixed effects, where we regressed respondents’ ingroup affect (measured with a feeling thermometer), over a treatment indicator (e.g. control vs. pooled, control vs. personal, control vs. non-personal, personal vs. non-personal).

improved attitudes towards Palestinians. In contrast, exposure to non-personal posts decreased the perceived importance of exposure to outgroup perspectives. In addition, general exposure to treatment decreased engagement with additional content highlighting outgroup viewpoints.

Our study highlights the need for caution and care in any process of attempting to diversify the content shared in online spaces and exposing people to content they might normally avoid in homogeneous online networks. Indeed, we find that the type of content used to diversify online environments may condition users' responses in terms of prejudice and willingness to consume outgroup news in the future. Additionally, our findings suggest that, at least among some subjects, attempts to diversify online networks might yield negative responses.

In providing evidence on the effects of online exposure to outgroup news on prejudice and willingness to engage with outgroup news in the future, this work offers several contributions. We build on the growing body of existing research on the consequences of segregation in online settings (Barberá, 2020) and speak to the broader literature on prejudice-reduction in at least three ways: (1) we provide evidence from outside of the U.S. context, (2) we design and test a scalable intervention in a setting of ongoing intergroup conflict (Paluck et al., 2020), and (3) we offer evidence for the differential effects of exposure to different types of content—in this case personal vs. non-personal stories.

Alongside other recent work using Facebook advertisements for social science research¹⁷, our approach highlights the potential benefits of using online platforms' advertising tools to randomly assign consenting participants to receive content in a naturalistic setting. In particular, by using Facebook ads in conjunction with Surveybot—an online tool that enables us to build chat bot surveys and distribute them via Facebook Messenger—we were able to get consent from our subjects before administering treatment, assign subjects to be exposed to particular types of content in a naturalistic setting on their newsfeeds, collect endline survey responses natively within the Facebook Messenger platform, and collect post-treatment behavioral measures. While Facebook ads and messenger bots are each increasingly used in social science research,¹⁸ our approach illustrates the utility of combining these tools to administer treatments.

Our findings also raise several questions that we view as exciting avenues for future research. For example, in our study, all treatments focused on Palestinians' negative experiences with Israeli authorities. We focused on such experiences because they are central to daily life in East Jerusalem, and are prominently featured on Palestinian social media. However, future research might explore how exposure to alternative types of content—focusing on arts, culture or sports, for example—

¹⁷See (Guess, 2021) for a recent overview.

¹⁸See innovative ongoing research by Leah Rozenzweig for examples <https://lehrrosenzweig.com/#/online-misinformation/>.

might affect attitudes and behaviors.

Second, as noted in our discussion, our treatments simply translated and shared Facebook posts written by Palestinians living in East Jerusalem with Jews living in West Jerusalem. We intentionally did not alter the content of these posts, in accordance with the goal of our partner organization to provide an unfiltered glimpse at lived experiences from communities across Jerusalem. Future research might explore how filtering or contextualizing this type of outgroup content would be more effective in reducing prejudice and generating empathy toward outgroup members.

Finally, we identify the effects of randomly-assigned exposure to outgroup viewpoints. In reality, social media users largely (though not exclusively) endogenously select into exposure to content they would like to see. The effects of exposure may be very different when it is chosen rather than exogenously assigned. We hope to build on our design in future research and compare the effects of endogenous and exogenous exposure to outgroup content on social media. Selective (Chassang et al., 2012), and patient preference trial designs (Knox et al., 2019), are promising approaches for future research in this area. We hope that future studies examining the dynamics of online network diversity and intergroup prejudice will tackle these questions to continue to improve our understanding of online political behavior and its offline consequences in diverse contexts.

References

- Aday, Sean, Henry Farrell, Marc Lynch, John Sides, John Kelly and Ethan Zuckerman. 2010. "Blogs and bullets: New media in contentious politics." *United States Institute of Peace* 65:1–31.
- Adida, Claire L, Adeline Lo and Melina R Platas. 2018. "Perspective taking can promote short-term inclusionary behavior toward Syrian refugees." *Proceedings of the National Academy of Sciences* 115(38):9521–9526.
- Bail, Christopher A, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout and Alexander Volfovsky. 2018. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences* 115(37):9216–9221.
- Bakshy, Eytan, Solomon Messing and Lada A Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348(6239):1130–1132.
- Bar-Tal, Daniel. 2013. *Intractable conflicts: Socio-psychological foundations and dynamics*. Cambridge University Press.
- Barberá, Pablo. 2020. *Social Media, Echo Chambers, and Political Polarization*. SSRC Anxieties of Democracy Cambridge University Press pp. 34–55.
- Bollens, Scott A. 2000. *On narrow ground: Urban policy and ethnic conflict in Jerusalem and Belfast*. Suny Press.
- Brewer, Marilyn B. 1999. "The psychology of prejudice: Ingroup love and outgroup hate?" *Journal of social issues* 55(3):429–444.
- Broockman, David and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352(6282):220–224.
- Chassang, Sylvain, Padró I Miquel, Erik Snowberg et al. 2012. "Selective trials: A principal-agent approach to randomized controlled experiments." *American Economic Review* 102(4):1279–1309.
- Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi and Michele Starnini. 2021. "The echo chamber effect on social media." *Proceedings of the National Academy of Sciences* 118(9).
- Del Vicario, Michela, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli and Walter Quattrociocchi. 2016. "Echo chambers: Emotional contagion and group polarization on facebook." *Scientific reports* 6(1):1–12.
- Enos, Ryan D. 2017. *The space between us: Social geography and politics*. Cambridge University Press.

- Figenschou, Tine Ustad and Kjersti Thorbjørnsrud. 2015. "Faces of an invisible population: Human interest framing of irregular immigration news in the United States, France, and Norway." *American Behavioral Scientist* 59(7):783–801.
- Grabe, Maria Elizabeth, Mariska Kleemans, Ozen Bas, Jessica Gall Myrick and Minchul Kim. 2017. "Putting a human face on cold, hard facts: Effects of personalizing social issues on perceptions of issue importance." *International Journal of Communication* 11:23.
- Greenwald, Anthony G and Thomas F Pettigrew. 2014. "With malice toward none and charity for some: Ingroup favoritism enables discrimination." *American Psychologist* 69(7):669.
- Guess, Andrew and Alexander Coppock. 2020. "Does counter-attitudinal information cause backlash? Results from three large survey experiments." *British Journal of Political Science* 50(4):1497–1515.
- Guess, Andrew M. 2021. "Experiments Using Social Media Data." *Advances in Experimental Political Science* p. 184.
- Harel, Tal Orian, Jessica Katz Jameson and Ifat Maoz. 2020. "The normalization of hatred: Identity, affective polarization, and dehumanization on Facebook in the context of intractable political conflict." *Social Media+ Society* 6(2):2056305120913983.
- Kalla, Joshua and David Broockman. 2021. "Which narrative strategies durably reduce prejudice? Evidence from field and survey experiments supporting the efficacy of perspective-getting."
- Kasara, Kimuli. 2013. "Separate and suspicious: Local social and political context and ethnic tolerance in Kenya." *the Journal of Politics* 75(4):921–936.
- Kasara, Kimuli. 2014. "Does local ethnic segregation lead to violence?: Evidence from Kenya." *Evidence from Kenya (February 25, 2014)* .
- Knox, Dean, Teppei Yamamoto, Matthew A Baum and Adam J Berinsky. 2019. "Design, identification, and sensitivity analysis for patient preference trials." *Journal of the American Statistical Association* 114(528):1532–1546.
- Levy, Roe. 2020. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review* .
- Lissitsa, Sabina. 2017. "Online political participation, online contacts with out-groups members and social distances." *Asian Journal of Communication* 27(1):18–32.
- Lowe, Matt. 2020. "Types of contact: A field experiment on collaborative and adversarial caste integration."
- Mor, Yifat, Yiftach Ron and Ifat Maoz. 2016. "'Likes' for peace: Can Facebook promote dialogue in the Israeli–Palestinian conflict?" *Media and Communication* 4(1):15–26.
- Mosleh, Mohsen, Cameron Martel, Dean Eckles and David G Rand. 2021. "Shared partisanship dramatically increases social tie formation in a Twitter field experiment." *Proceedings of the National Academy of Sciences* 118(7).

- Mousa, Salma. 2020. "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq." *Science* 369(6505):866–870.
- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2(2):109–138.
- Munger, Kevin. 2017. "Tweetment effects on the tweeted: Experimentally reducing racist harassment." *Political Behavior* 39(3):629–649.
- Paluck, Elizabeth Levy, Roni Porat, Chelsey S Clark and Donald P Green. 2020. "Prejudice reduction: Progress and challenges." *Annual Review of Psychology* 72.
- Rokem, Jonathan, Chagai M Weiss and Dan Miodownik. 2018. "Geographies of violence in Jerusalem: the spatial logic of urban intergroup conflict." *Political Geography* 66:88–97.
- Rossiter, Erin. 2020. "The Consequences of Interparty Conversation on Outparty Affect and Stereotypes."
- Scacco, Alexandra and Shana S Warren. 2018. "Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria." *American Political Science Review* 112(3):654–677.
- Semetko, Holli A and Patti M Valkenburg. 2000. "Framing European politics: A content analysis of press and television news." *Journal of communication* 50(2):93–109.
- Siegel, Alexandra A and Vivienne Badaan. 2020. "# No2Sectarianism: Experimental approaches to reducing sectarian hate speech online." *American Political Science Review* 114(3):837–855.
- Siegel, Alexandra, Jonathan Nagler, Richard Bonneau and Joshua Tucker. 2021. "Tweeting Beyond Tahrir: Ideological Diversity and Political Intolerance in Egyptian Twitter Networks." *World Politics* .
- Simonovits, Gábor, Gabor Kezdi and Peter Kardos. 2018. "Seeing the world through the other's eye: An online intervention reducing ethnic prejudice." *The American Political Science Review* 112(1):186.
- Stats, StatCounter Global. 2018. "Mobile Operating System Market Share Worldwide— StatCounter Global Stats."
- Taylor, K and L Silver. 2019. "Digital connectivity growing rapidly in emerging economies." *Pew Research* .
- Trilling, Damian, Petro Tolochko and Björn Burscher. 2017. "From newsworthiness to share-worthiness: How to predict news sharing based on article characteristics." *Journalism & Mass Communication Quarterly* 94(1):38–60.
- Weiss, Chagai M. 2020. "Sharing Spaces: Segregation, Integration, and Intergroup Relations." *APSA— COMPARATIVE POLITICS* p. 94.
- Weiss, Chagai M. 2021. "Diversity in health care institutions reduces Israeli patients' prejudice toward Arabs." *Proceedings of the National Academy of Sciences* 118(14).

Supplementary Information

Desegregating Digital Spaces

Contents

S1 Selection of Facebook posts for Treatment Conditions	S2
S1.1 Selected Posts	S3
S2 Baseline Survey	S6
S3 Endline Survey	S7
S4 Attrition	S10
S5 Informed Consent Form	S12
S6 Comments on Treatment Posts	S13

S1 Selection of Facebook posts for Treatment Conditions

We followed four steps in selecting a total of 28 Facebook posts from our partner organization, 0202 Points of View from Jerusalem, for our two-week-long intervention (14 personal and 14 non-personal posts).

1. We viewed the full set of 0202 posts about East Jerusalem, spanning the period November 1, 2020 through January 26, 2021. (Note that one treatment post is dated from February 16, 2021. One post was rejected by Facebook due to relatively poor photo image quality. We replaced this post with a post from February 16, 2021.)
2. We excluded any posts that did not focus directly on East Jerusalem residents' interactions with the Israeli state (what East Jerusalem residents would describe as experiences of occupation. In practice, this material primarily focused on arrests and other law-enforcement). We therefore excluded all posts focused on sports, entertainment, art and culture, and news unrelated to East Jerusalem residents' day-to-day experiences interacting with the Israeli state. These interactions typically involved Israeli law enforcement.
3. We then sorted all topically relevant posts into two categories: (a) those focused on individual or personal news, and (b) those focused on non-personal or communal news. Personal posts focus on the experiences of one specific individual (or, in a few cases, several individuals). They refer to this person by name and focus on that individual's experiences of interacting with the Israeli state. In all personal posts, a picture of the individual's face is shown in the post.
4. We deliberately excluded all posts that:
 - (a) could not clearly be categorized as either personal or non-personal in their focus.
 - (b) included graphic or disturbing images of violence or the consequences of violence on human bodies, out of a concern that such images might trigger or traumatize study

subjects.

(c) included blurry photographs.

(d) contained video clips.

S1.1 Selected Posts

We present all posts used in our intervention in this section. As is evident below, the content employed in our personal condition included 5 posts about arrests, 4 posts about house demolitions, 2 posts about targeted violence, and 3 posts about harassment by the state. Similarly, the content employed in our non-personal condition included 5 posts about arrests, 2 posts about house demolitions, and 9 posts about harassment by the state. Each post included a Hebrew translation followed by an original Arabic Facebook post and an image.

Figure S3: Non-Personal Treatment Ads

The figure displays 20 individual advertisements arranged in a grid. Each advertisement is a screenshot of a social media post or flyer. The text in the ads is in Hebrew and describes various non-personal treatment services. Key elements across the ads include:

- Text:** Descriptions of services such as "מטרת הטיפול היא להפחית את רמת האנרגיה של המטופל" (The purpose of the treatment is to reduce the patient's energy level) and "טיפול בשימוש במגנט" (Treatment using magnets). Some ads mention "מטופלים" (patients) and "מטופלים" (patients).
- Images:** Photos showing people in clinical or treatment settings, some with equipment, and some showing outdoor scenes.
- Contact Information:** Repeatedly, the ads provide contact details for "מרכז מידע וידע מרכז מידע וידע מרכז מידע וידע" (Wadi Hishah Information Center) located at "מרכז מידע וידע מרכז מידע וידע" (Silwan, Jerusalem).

S2 Baseline Survey

Demographics

B1 Gender

B2 Age

B3 Religion

B4 Religiosity

B5 Ethnicity

B6 Jerusalem resident?

B7 Hebrew/Arabic proficiency

B8 Ideology (Right-Left scale)

Prejudice

B9 Feeling Thermometer (Foreign workers / Left-wing supporters / Right-wing supporters / Jews in general / Ultra-Orthodox Jews / Jewish immigrants / Arabs)

Perspective Taking and empathy

B10 How well would you say the following statements describe you, where “1” does not describe you at all and “4” describes you very well:

- (a) It is important try to understand ultra-Orthodox Jews by imagining their feelings, suffering, or thoughts
- (b) It is important try to understand Arabs Jews by imagining their feelings, suffering, or thoughts

- (c) Even if I disagree with ultra-Orthodox Jews, it is important to try to think of reasons why that group takes a different point of view
- (d) Even if I disagree with Arabs Jews, it is important to try to think of reasons why that group takes a different point of view.

Social media and news consumption

- B11 How much time do you spend on Facebook on an average day?
- B12 Do you agree or disagree that exposure to diverse points of view on social media is valuable?
- B13 How often do you see news or news headlines on Facebook?
- B14 Thinking about the news you see on Facebook, do you prefer sources that share/challenge your point of view/ do not have a particular point of view?

S3 Endline Survey

Facebook use

- E1 Do you use Facebook for any of the following reasons? (Shop / Search for a job / Play games / Meet new people / Other)
- E2 Do you feel you are (Spending too much time on Facebook / Spending just enough time on Facebook / Spending too little time on Facebook)?

Prejudice

- E3 Feeling thermometer (0 to 10) with respect to left-wing party supporters
- E4 Feeling thermometer (0 to 10) with respect to right-wing party supporters
- E5 Feeling thermometer (0 to 10) with respect to Jews

E6 Feeling thermometer (0 to 10)with respect to Arabs

E7 Do you agree or disagree: “Most Arabs in Jerusalem want to live in peace.” (4 point scale)

Interest in outgroup media

E8 Do you agree with the statement: “Exposure to diverse points of view on social media is valuable?” (4 point scale)

E9 Do you agree with the statement: “I would benefit from more exposure to news about Arab life in East Jerusalem?” (4 point scale)

Empathy and perspective-taking

E10 “It is important to try to understand Arabs by imagining their feelings, suffering, or thoughts.” (4 point scale of alignment/misalignment with personal views)

E11 “Even if I disagree with Arabs, it is important to try to think of reasons why that group takes a different point of view.” (4 point scale of alignment/misalignment with personal views)

E12 Would you agree or disagree: “Residents of East Jerusalem face more difficult problems in their daily lives than residents of West Jerusalem.” (4 point scale)

Policy preferences

E13 Do you support dividing Jerusalem into two cities - one Jewish and one Arab? (Yes/No)

E14 Do you think that Jerusalem municipality should: (Give priority to the needs of its Jewish residents over others living in the city / Give priority to the needs of its Arab residents over others living in the city / Give equal priority to the needs of its Jewish and Arab residents)?

Humanization

E15 To what extent are east Jerusalem residents rational and logical (1-7)

E16 To what extent are east Jerusalem residents refined and cultured (1-7)

E17 To what extent are east Jerusalem residents Lacking self-restraint, like animals (1-7)

E18 To what extent are east Jerusalem residents Superficial, lacking in depth (1-7)

E19 On a scale of 1-6, where 1 means “very different from one another” and 6 means “very similar to one another”, please indicate how different or similar Arabs living in Jerusalem are to each other, with regards to intellect

E20 On a scale of 1-6, where 1 means “very different from one another” and 6 means “very similar to one another”, please indicate how different or similar Arabs living in Jerusalem are to each other, with regards to morality

Openness to outgroup experiences

E22 How interested would you be in the following online experiences, where “1” means not interested at all and “4” means very interested:

- (a) A virtual tour of Palestinian neighborhoods in West Jerusalem
- (b) An online lecture about life in East Jerusalem
- (c) A video showcasing the life of a Palestinian resident of East Jerusalem

S4 Attrition

In this section, we empirically assess several concerns relating to attrition. First, in table S2, we demonstrate that participation in our endline survey does not correlate with treatment condition. Indeed, we show that receiving any content, personal content, or non-personal content, does not predict participation in endline, participation in endline conditional on participating in baseline, and participation in endline conditional not participating in baseline.

Table S2: Correlation of Treatment with Survey Participation

	Endline		Endline No Baseline		Endline and Baseline	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Personal	-0.002 (0.014)		0.009 (0.009)		-0.011 (0.012)	
Non-Personal	-0.009 (0.014)		0.004 (0.008)		-0.012 (0.012)	
Pooled		-0.006 (0.012)		0.006 (0.007)		-0.012 (0.010)
Num. obs.	4615	4615	4615	4615	4615	4615

* $p < 0.05$. All models include block fixed effects.

To further consider the relationship between our treatments and responses to the endline survey, we created binary variables for each survey question, taking the value of 1 if a respondent answered a given question, and a value of 0 otherwise. We regress these variables over our main treatment indicators, and report results in Figure S4. The models reported in Figure S4, suggest that receiving a particular treatment did not increase the probability of responding to particular survey items.

Lastly, in Table S3, we focus on respondents who reported demographics in our baseline survey, and consider the extent to which reported demographics predict selection into endline survey. We find that men were more likely to select into our endline survey, and older respondents were less likely to select into our endline survey. However, religiosity, ideology, and pre-treatment affect towards Palestinians do not predict selection into our endline survey.

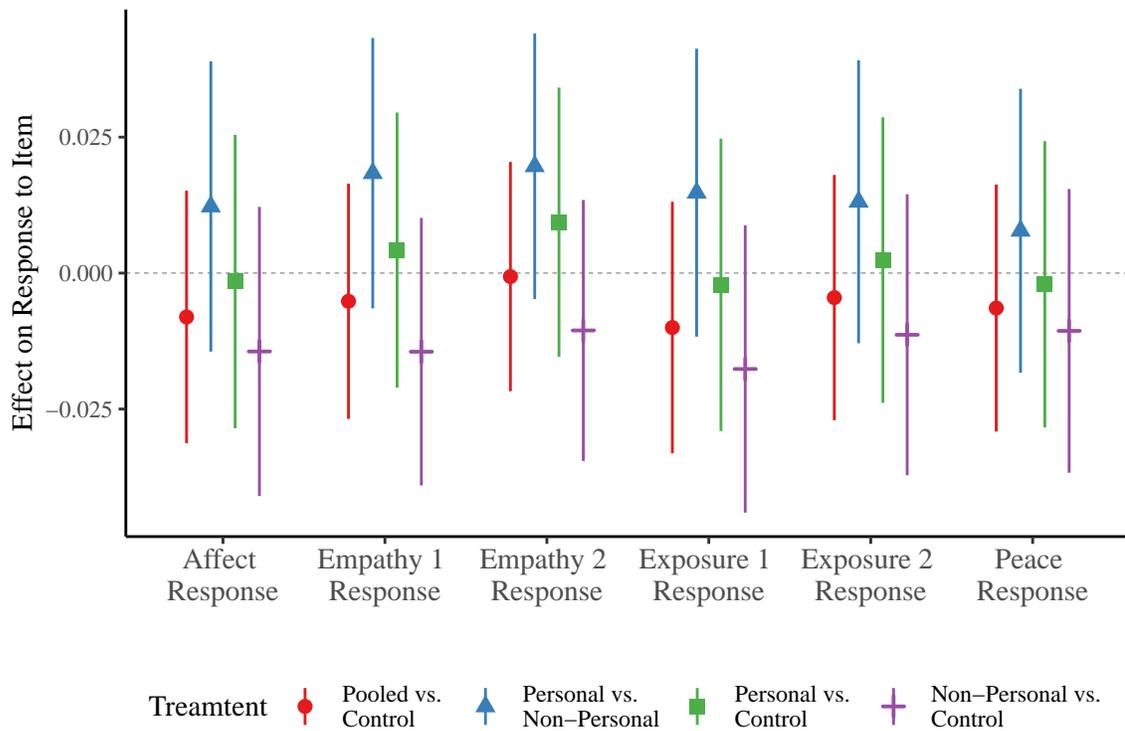


Figure S4: **Treatment effects on response to survey items** – Each point estimate is extracted from an OLS regression with block fixed effects, where we regressed an item response binary variable, over a treatment indicator (e.g. control vs. pooled, control vs. personal, control vs. non-personal, personal vs. non-personal).

Table S3: Participation in Endline Amongst Baseline Respondents

	Endline Participation				
	Model 1	Model 2	Model 3	Model 4	Model 5
Male	0.047*				
	(0.019)				
Age Group		-0.018*			
		(0.006)			
Religiosity (Secular:Ultra-Orthodox)			-0.001		
			(0.007)		
Ideology (Right:Left)				0.009	
				(0.006)	
Pre-Treatment Thermometer					-0.003
					(0.004)
Intercept	0.225*	0.282*	0.266*	0.242*	0.297*
	(0.013)	(0.012)	(0.015)	(0.022)	(0.022)
Num. obs.	2195	2078	2027	1979	1828

* $p < 0.05$.

S5 Informed Consent Form

Description of the research

You are invited to participate in a research study about online information and intergroup relations implemented by researchers at UW-Madison, University of Colorado Boulder, and WZB Berlin. You have been asked to participate because we are interested in learning how you respond to information about local news and events in Jerusalem online. The purpose of the research is to better understand whether sharing information about other groups in Jerusalem affects intergroup relations. This study's sample includes residents of Jerusalem.

What will my participation involve?

If you decide to participate in this research, you will be asked to fill out this brief survey, and you will be invited to participate in a follow up survey in several weeks. Between surveys we may expose you to information about local news and events in Jerusalem via Facebook. In addition, after the first survey, we will consider how you react to different prompts on Facebook. Specifically, we may check to see how you respond to a prompt inviting you to "follow" specific pages or RSVP to events on Facebook. These prompts will be administered by 0202 Points of view from Jerusalem. You can choose to fully ignore or engage with these different Facebook prompts which are administered by 0202 Points of View from Jerusalem as you wish. Participation in this first survey will take about five minutes to complete. Several weeks from now, we will send you an invitation to participate in another survey which will take between five and ten minutes per survey. In total, your participation will require no more than 15 minutes for all surveys.

Are there any risks to me?

Risks associated with this study include confidentiality breach. In addition, you may find parts of the surveys we distribute as upsetting, and therefore you can feel free to skip questions you feel uncomfortable answering.

Are there any benefits to me?

We don't expect any direct benefits to you from participation in this study.

Will I be compensated for my participation?

If you participate in this study, and complete the follow-up survey, you will receive a chance to win one of several prizes (an iPhone, or one of ten 100NIS gift cards from Tav Ha-Zahav which can be used in many stores across Israel) in a raffle we will conduct. The research team will conduct the raffle on December 25, 2020. The research team will distribute 11 prizes (1 iPhone, and ten 100NIS gift cards) between 10,000 possible study participants selected at random, using a computerized random selection procedure. We will notify study participants if they won or didn't win a prize via email and Facebook message, and will deliver all prizes to winners via Israeli mail by January 20, 2021.

How will my confidentiality be protected?

This study is confidential. Neither your name or any other identifiable information will be published. Only our research team will have access to all data collected in this study. In order to enter the raffle described above, you will need to provide your full name and email address. This information will never be published, and it will only be used to contact you if you win one of the raffle prizes. Coded data will be shared between researchers on the study team, but all research data will only be used for this study and not retained for future use.

Whom should I contact if I have questions?

You may ask any questions about the research at any time. If you have questions about the research after you leave today you should contact Chagai Weiss at the following email address: cmweiss3@wisc.edu. If you are not satisfied with response of research team, have more questions, or want to talk with someone about your rights as a research participant, you should contact the Education and Social/Behavioral Science IRB Office at 608-265-4312. If you decide not to participate or to withdraw from the study, you may do so without penalty. If you agree to participate in the study – please select “I agree to participate”.

S6 Comments on Treatment Posts

Examples of Translated Comments on Treatment Ads: Warning: Some of the comments below use strong or abusive language.

- *What is a Palestinian? It's like someone from the made up place Atlantis, Cripton, Never-*

land, Oz, Hogwarts. Stop making up nations and places that don't exist in reality. Palestinians, Elfs, Orks, Batman, and Hobbits exist only in fairy tales just like Atlantis, Cripton, Neverland, Palestine and Scubia.

- *The only occupiers are the Muslims who invaded Israel*
- *Damn you collaborators of the enemy.*
- *They are lucky to deal with the Israeli security forces, and not security forces in Arab countries. There an attack or disruption of public order ends up with bodies. I hope this post is meant to thank the Israeli security forces for their considerate approach in dealing with those who tried to kill and hurt.*