Helping or Harming? The Effect of Trigger Warnings on Individuals with Trauma Histories

Payton J. Jones

Benjamin W. Bellet

Richard J. McNally

Harvard University, Cambridge, MA

Correspondence concerning this article should be addressed to Payton J. Jones, Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138. E-mail: payton_jones@g.harvard.edu

Abstract

Objective: Trigger warnings alert trauma survivors about potentially disturbing forthcoming content. However, most empirical studies on trigger warnings indicate that they are either functionally inert or cause small adverse side effects. These evaluations have been limited to either trauma-naïve participants or mixed samples. Accordingly, we tested whether trigger warnings would be psychologically beneficial in the most relevant population: survivors of serious trauma.

Method: Our experiment was a preregistered replication and extension of a previous one (Bellet, Jones, & McNally, 2018); 451 trauma survivors were randomly assigned to either receive or not receive trigger warnings prior to reading potentially distressing passages from world literature. They provided their emotional reactions to each passage; self-reported anxiety was the primary dependent variable.

Results: We found no evidence that trigger warnings were helpful for trauma survivors, for those who self-reported a PTSD diagnosis, or for those who qualified for probable PTSD, even when survivors' trauma matched the passages' content. We found substantial evidence that trigger warnings countertherapeutically reinforce survivors' view of their trauma as central to their identity. Regarding replication hypotheses, the evidence was either ambiguous or substantially favored the hypothesis that trigger warnings have no effect.

Conclusions: Trigger warnings are not helpful for trauma survivors. It is less clear whether trigger warnings are explicitly harmful. However, such knowledge is unnecessary to adjudicate whether to use trigger warnings – because trigger warnings are consistently unhelpful, there is no evidence-based reason to use them.

Keywords: trigger warning, trauma, PTSD, resilience, pre-registered replication

Helping or Harming? The Effect of Trigger Warnings on Individuals with Trauma Histories

Giving a trigger warning means providing prior notification about forthcoming content that may be emotionally disturbing (Boysen, 2017). In this sense, trigger warnings are similar to PG-13 or "viewer discretion advised" warnings that are common across many different forms of media. Trigger warnings are distinct in that they originated as a measure of protection specifically for survivors of trauma. For those with posttraumatic stress disorder (PTSD), viewing reminders of trauma can spark painful reexperiencing symptoms (e.g., flashbacks; American Psychiatric Association [APA], 2013). Trigger warnings originated in online discussion groups for survivors of sexual trauma, where individuals would warn readers before discussing their experiences. Since their inception, trigger warnings have expanded far beyond the boundaries of specialized online communities. Trigger warnings are now used in educational settings, social media, entertainment, and other venues. In addition to their expansion in setting, they have also expanded in scope beyond sexual violence (Wilson, 2015).

Trigger warnings have sparked considerable debate in higher education. Proponents of trigger warnings emphasize their importance in creating an inclusive atmosphere for disadvantaged groups on campus (e.g., Karasek, 2016). They argue that trigger warnings provide agency to engage or not to engage and that they allow trauma survivors to adequately prepare to engage with difficult material. Critics suggest that trigger warnings imperil free speech, academic freedom, and effective teaching, preventing students from engaging with challenging material (e.g., Ellison, 2016). Other critics have suggested that trigger warnings foster unreasonable expectations about the world, hampering natural resilience among young people (e.g., Lukianoff & Haidt, 2015). Further, trigger warnings could also be problematic for trauma survivors in particular (McNally, 2016). Those who view trauma as a core part of their identity

have worse symptoms (Berntsen & Rubin, 2006; Brown, Antonius, Kramer, Root, & Hirst, 2010; Robinaugh & McNally, 2011). Therefore, trigger warnings might iatrogenically reinforce the importance of past traumatic events for the very people they were originally designed to help.

The arguments surrounding trigger warnings are often complex. Before diving into this complexity, a much more basic question should be answered: do trigger warnings actually work? From the vantage point of clinical science, trigger warnings are a type of community-based clinical intervention intended to foster emotional well-being among trauma survivors. Yet due to their grassroots origin in a non-clinical setting, trigger warnings have expanded for years without the rigorous scientific evaluation that normally accompanies such interventions.

Bellet, Jones, and McNally (2018) were among the first to experimentally test the effect of trigger warnings. In a crowd-sourced sample of individuals who had not experienced past trauma, they found that trigger warnings given before literature passages had no significant effect on anxiety. Further, they found that trigger warnings undermined participants' sense of their resilience to potential future traumatic events, and that of others. They also reported a moderation effect – among individuals who believed that words were emotionally harmful, trigger warnings acutely increased anxiety reactions.

Since this original study, the scientific literature has quickly expanded. Bellet et al. (2019) conducted a preregistered replication of the same protocol of Bellet et al. (2018) with undergraduate college students. Their results suggest that trigger warnings created a trivially small, yet genuine increase in anxiety. However, they found strong evidence that the previously observed effects on projected vulnerability and the moderation effect from Bellet et al. (2018) did not replicate among college students. In the most comprehensive set of studies to date, Sanson, Strange, and Garry (2019) concluded that trigger warnings had trivially small effects

overall. Across six studies of varying sample characteristics, they found that negative affect and intrusive memories were similar regardless of whether individuals received trigger warnings.

Bridgland, Greene, Oulton, and Takarangi (2019) similarly found that trigger warnings had trivially small effects on arousal levels when participants viewed photos. Importantly, however, their results differentiated *anticipatory* anxiety from *response* anxiety. Anticipatory anxiety refers to levels of anxiety after viewing the trigger warning but before viewing the stimulus, whereas response anxiety refers to anxiety after viewing the stimulus. Although trigger warnings appeared to have a trivial effect on *response* anxiety, they reliably increased *anticipatory* anxiety. Relatedly, Bruce (2017) found that trigger warnings produced greater physiological markers of anticipatory anxiety compared to PG-13 warnings or no warnings.

Gainsburg and Earl (2018) found that trigger warnings increased negative anticipatory affect, but slightly decreased negative response affect. However, they found that participants who were given trigger warnings were more likely to avoid both film clips and essays, which may exacerbate anxiety in the long run (McNally, 2016). This particular finding is contrasted by Kimble (2019), who found that individuals rarely avoided material due to trigger warnings. Articles evaluating the effect of trigger warnings on anxiety or negative affect are summarized in Table 1.

Table 1

Effect of Trigger Warnings on Anticipatory and Response Anxiety

Authors (*Year*)	*n*	Source	Trauma Exposure	Stimuli	Outcome	Anticipatory Anxiety *(d)*	Response Anxiety *(d)*
Bellet, Jones, & McNally (*2018*)	270	Crowd-sourced	No	Literature passages	Self-reported anxiety		0.06 [-0.18, 0.30]
Gainsburg & Earl (*2018*)[1]	276; 979	Crowd-sourced	Mixed	Essay	Negative affect (SAM)	0.75** [0.58, 0.92]; 0.26** [0.10, 0.42]	-0.17* [-0.33, -0.01]
Bellet, Jones, Meyersburg, Brenneman, Morehead, & McNally (*2019*)	462	Students	No	Literature passages	Self-reported anxiety		0.20* [0.02, 0.38]
Sanson, Strange, & Garry (*2019*)[2]	1880	Students / Crowd-sourced	Mixed	Story / Film clip	Negative affect (PANAS)		0.02 [-0.08, 0.13]
Bridgland, Greene, Oulton, & Takarangi (*2019*)[2]	1600	Crowd-sourced	Mixed	Photos	State anxiety (STAI)	1.36** [0.99, 1.74]	0.07 [-0.03, 0.16]
Authors (*Current Study*)	451	Crowd-sourced	Yes	Literature passages	Self-reported anxiety		0.08 [-0.11, 0.26]

Note: Positive *Cohen's d* type effect sizes indicate an increase in anxiety. SAM = Self-Assessment Manikin, PANAS = Positive and Negative Affect Scale, STAI = State Trait Anxiety Inventory. Cells are left blank for studies that did not measure anticipatory anxiety.
*$p < 0.05$, **$p < 0.01$
[1]Anticipatory anxiety effects are from Studies 2 & 3 respectively ($n = 276, 979$) and based on reported *t*-values; response anxiety is from Study 3. Confidence intervals are estimated based on incomplete information.
[2]Results are internal meta-analyses across all experiments. For Bridgland et al. (2019) response anxiety, we meta-analyzed the effects reported in Table 5

This encouraging growth of studies has begun to converge on the consensus that trigger warnings are not typically helpful in reducing anxiety. This finding has been consistent across various operationalizations of trigger warnings and types of stimuli. For instance, Sanson et al. (2019) find similar effects regardless of whether trigger warnings mention potential emotional reactions (e.g., "You might find this content disturbing") or whether they only mention content (e.g., "The following story contains violence and death"). Similar effects are found with literature passages, stories, photos, and film clips. The literature also suggests several different types of harm potentially caused by trigger warnings (e.g., anticipatory anxiety, avoidance, perception of vulnerability), but with occasionally mixed or contradictory results.

There remain several important limitations to this area of research. First, none of the studies has exclusively focused on the primary intended target of trigger warnings – survivors of trauma. Although some of the studies have included trauma survivor subgroups in their sample (e.g., Sanson et al., 2019), this has not been the main focus of any study. If trigger warnings are considered as a clinical intervention to promote the well-being of trauma survivors, this is an important limitation. Second, trigger warnings present unique concerns for trauma survivors, especially those who are experiencing symptoms of PTSD. Trigger warnings might reinforce survivors' belief that their trauma is central to their identity, and the severity of PTSD symptoms might also moderate trigger warnings' effects. Third, many studies used different operationalizations and stimuli. Although the consistency of results across diverse studies inspires confidence, direct replications are also essential.

In the current study, we tested the effect of trigger warnings in a large sample of trauma survivors recruited from Amazon Mechanical Turk, a crowd-sourcing platform. This preregistered study includes a direct replication of the experiment in Bellet et al. (2018) and

extends the paradigm to address questions specific to trauma survivors. Further exploratory analyses examined vulnerable subpopulations (e.g., those who had received a diagnosis of PTSD).

Method

Participants

The preregistration for the study design and analysis plan are available on the Open Science Framework (https://osf.io/gdxtr/). Any deviations or exploratory analyses that were not specified in the preregistration are marked as such within this manuscript. Participants were recruited online from Amazon Mechanical Turk. We prespecified a sequential data collection procedure with a stopping rule based on Bayes Factors. However, the evidential criteria for our stopping rule were not met at any intermediate step, so we recruited participants until the specified ceiling of 600 participants had completed the study. Participants were excluded from the study if they incorrectly answered an attention check or if they failed an English fluency verifier[1]. This left a final sample of 451 participants.

Procedure

This study was a randomized controlled experiment with one control group (no trigger warnings) and one experimental group (trigger warnings for distressing passages). After providing informed consent, participants were asked to complete a CAPTCHA and answer three questions to verify US residency (e.g., "What is the most common emergency number in the United States? [0-0-0 / 9-1-1 / 0-0-0-9-5 / 9-9-9]"). Participants failing these items were immediately barred from completing the study. Remaining participants then completed a single-

[1] "In the space provided below, please describe your activities last weekend in exactly 3 sentences. (This question is very important and will be used to determine payment. Please use complete English sentences)."

item question that screened for the presence of trauma according to PTSD diagnostic Criterion A in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013). The DSM-5 defines trauma as "exposure to actual or threatened death, serious injury, or sexual violence." Participants who had not experienced such a trauma were excluded from the study.

Participants then read literature passages typical of a high school or college English class. Passages were standardized by length, and participants were shown the passages for a minimum of 20 seconds before they were allowed to proceed to the next screen. The passages were previously rated on the degree to which they provoked anxiety in a pilot study (Bellet et al., 2018). Depending on their content, passages are hereafter classified as either *neutral* (no disturbing content; e.g., a character description from Melville's *Moby-Dick*), *mildly distressing* (general themes of violence or harm with no graphic details; e.g., a description of a battle from Bradley's *Flags of our Fathers*), or *markedly distressing* (graphic scenes of violence, injury, or death; e.g., the murder scene from Dostoevsky's *Crime and Punishment*). After each passage, participants rated their emotional state by using slider bars ranging from 0 (not at all) to 100 (very much) on seven emotions: *sad*, *happy*, *afraid*, *anxious*, *angry*, *content*, and *disgusted*. The responses to the *anxious* rating were used as the primary outcome measure.

Participants in both conditions first read three mildly distressing passages in random order to establish a baseline. Next, participants read a series of five neutral passages and five markedly distressing passages intermixed in random order. In the experimental condition, markedly distressing passages were preceded by a trigger warning (*TRIGGER WARNING: The passage you are about to read contains disturbing content and may trigger an anxiety response, especially in those who have a history of trauma*). In the control condition, passages were

preceded by a screen that indicated they were about to view the next passage, which was acknowledged by clicking a radio button. After these 10 passages, participants read three more mildly distressing passages appearing in random order that served to test for any sensitization effects.

After reading all passages, participants completed the questionnaires detailed below. Participants also answered questions about demographic information and psychiatric history, completed an English fluency verifier, and answered validity-related questions that did not impact payment (e.g., "What do you think was the purpose of this study?", "Is there any reason you think that your data should not be used (this will not impact payment)?"). At the end of the study, they received a debriefing form explaining the purpose of the study in detail.

Measures

Centrality of Event Scale (CES). The CES is a 7-item questionnaire that measures the extent to which participants view the memory of their worst event as a reference point for personal identity and the attribution of meaning to other experiences in their life (Berntsen & Rubin, 2006). Items (e.g., "I feel that this event has become a central part of my life story") are rated on a 5-point Likert scale (1 = *totally disagree*, 5 = *totally agree*). The CES displayed excellent internal consistency in the current study (α = 0.94).

Perceived Posttraumatic Vulnerability Scale – Self (PPVS-S). This 19-item questionnaire measures participants projections of their own emotional impairment and posttraumatic symptoms if they were to hypothetically experience a trauma in the future (Bellet et al., 2018). Participants are asked to imagine being exposed to an attempt on their life, and then indicate their agreement with the effects of that experience (e.g., "I would not be able to work a

job, or take care of myself") on a 100-point scale (1 = *disagree*, 100 = *agree*). The PPVS-S displayed excellent internal consistency in the current study (α = 0.95).

Perceived Posttraumatic Vulnerability Scale – Other (PPVS-O). This 19-item questionnaire measures participants' belief that if an "average" person were to experience a trauma, they would experience persistent and debilitating emotional harm (Bellet et al., 2018). Participants are asked to imagine an average person being exposed to an attempt on his or her life, and then indicate their agreement with the effects of that experience (e.g., "he/she would have nightmares of the event") on a 100-point scale (1 = *disagree*, 100 = *agree*). The PPVS-O displayed excellent internal consistency in the current study (α = 0.95).

Life Events Checklist for DSM-5 (LEC-5). The LEC-5 is a self-report instrument that identifies specific traumatic events that have occurred in one's lifetime (Weathers, Blake, et al., 2013). The LEC-5 contains 16 events known to potentially result in PTSD or distress (e.g., "life threatening illness or injury") and an additional option for "any other very stressful event or experience". In our study, participants were initially screened by a question assessing the presence of a Criterion A trauma. Later in the study, they were provided with the LEC-5 and asked to choose the event description that best matched their most stressful or traumatic event.

PTSD Checklist for DSM-5 (PCL-5). The PCL-5 is a 20-item questionnaire that assesses the presence and severity of PTSD symptoms in the past month (Weathers, Litz, et al., 2013). When answering the PCL-5, participants in our study were instructed to answer the questions keeping in mind their worst event as selected on the LEC-5. Items on the PCL-5 correspond closely to DSM-5 criteria for PTSD (e.g., "In the past month, how much were you bothered by repeated, disturbing, and unwanted memories of the stressful experience"). The PCL-5 is often used to monitor symptoms over time, to screen for PTSD, or assist in making a provisional

diagnosis of PTSD. For exploratory analyses involving the PCL-5, we used the cutoff score of 33 recommended by the United States Department of Veterans Affairs (Weathers, Litz, et al., 2013), and based on research (Bovin et al., 2017; Wortmann et al., 2017). The PCL-5 displayed excellent internal consistency in the current study (α = 0.96).

Words Can Harm Scale (WCHS). The WCHS is a 10-item scale that measures the degree to which participants believe that words can cause serious and lasting emotional harm (Bellet et al., 2018). Participants rated their agreement with each statement (e.g., "Even a simple phrase can be emotionally traumatizing for someone vulnerable") on a 100-point scale (1 = *disagree*, 100 = *agree*). The WCHS displayed excellent internal consistency in the current study (α = 0.92).

Trigger Warnings Attitudes Assessment (TWAA). We administered three items to assess participants' prior exposure to and attitudes about trigger warnings. First, we provided participants with a definition of trigger warnings (i.e., "A trigger warning is a statement given prior to presented material that allows the viewer to prepare for or avoid distress that it may cause, particularly if the viewer has clinical mental health issues"). Participants were then asked to give a binary rating of whether they believe that trigger warnings should be given prior to potentially distressing material. If the participants selected "yes", they were shown a checklist asking why they think trigger warnings should be used (e.g., "Trigger warnings help to protect vulnerable populations…") including an "Other" option with the ability to write in a response. Participants were then asked to rate their agreement with the statement "I have personally seen many trigger warnings used before" on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). Only the last item was used for analysis in the present study, as specified in the

preregistration. The other items are to be used in future studies addressing attitudes about trigger warnings.

Trauma-Matching Passages. We asked participants if any of the literary passages they read during the study reminded them of their worst event. If participants answered yes, we provided them with a checklist of passages, and asked them to identify which ones reminded them of their worst event. These passages were marked as "trauma-matching" passages.

Demographics Questionnaire. We asked participants to report their gender, race, ethnicity, religiosity, political orientation, and age. Religiosity and political orientation were assessed with a 5-point Likert scale (1 = *not religious*, 5 = *extremely religious*; 1 = *very liberal*, 5 = *very conservative*). We also asked participants to report whether they are currently a full-time undergraduate student.

Psychiatric History. At the beginning of the study, all participants were given a screener assessing for the presence of a Criterion A traumatic event. Participants were only included in the study if they indicated the presence of a Criterion A event. Near the end of the study, we asked participants whether they had “ever been diagnosed with a psychiatric or psychological problem.” If participants answer yes, we asked them to choose all diagnoses that apply from a list including PTSD and “Other” (to allow for a free response of any disorders not listed).

Analyses

All analyses were conducted in R (R Core Team, 2019). Analyses used linear regressions with trigger warning condition (trigger warnings versus no trigger warning) as the primary independent variable. For all analyses, we used Bayes Factors (BFs) as our inference criteria. BFs give relative evidence between two competing hypotheses. For all tests, we used a preregistered minimum BF value of 3 (or 1/3) as a criterion for "substantial evidence" relative to

the null or alternative hypothesis. As indicated in the preregistration, we first examined whether demographic or psychiatric history differed by condition. If this were the case, we added those variables as covariates in regression analyses. Code for all analyses is available in the supplemental materials.

Preregistered Replication Tests. We preregistered five replication tests, each related to a previous effect observed in Bellet et al. (2018). For replication tests, our two competing hypotheses were that (1) the observed effect was equal to zero ($t_{obs} = 0$) or that (2) the observed effect was equal to the effect in the previous study by Bellet et al. ($t_{obs} = t_{orig}$). Replication BFs were computed following the *t*-value comparison procedure described by Verhagen and Wagenmakers (2014). We tested the effect of trigger warnings on (1) participants' perceptions of their own posttraumatic vulnerability via the PPVS-S, (2) participants' perceptions on others' posttraumatic vulnerability via the PPVS-O, (3) immediate anxiety response following markedly distressing passages, (4) subsequent anxiety response to mildly distressing passages presented without a trigger warning, and (5) an interaction effect between trigger warning condition and the WCHS on immediate anxiety response (including a simple slopes analysis if the interaction was significant).

Trauma-Specific Preregistered Tests. We preregistered several additional tests to answer specific questions about trauma survivors. For these tests, our two competing hypotheses were (1) the observed effect was equal to zero ($t_{obs} = 0$) or that (2) the observed effect was not equal to zero ($t_{obs} \neq 0$). Specifically, this is done by comparing a linear model which includes the parameter of interest (e.g., condition) against a linear model without that parameter (e.g., intercept only model) using the *lmBF* function in the *BayesFactor* package (Morey & Rouder, 2018). First, we tested whether trigger warnings affected participants' ratings of trauma

centrality on the CES. Second, we tested whether PTSD severity scores on the PCL moderated any of the previous tests (e.g., effect on PPVS-S, PPVS-O, etc.). Third, we tested whether participants' self-reported prior exposure to trigger warnings (see TWAA) moderated any of the previous tests.

Exploratory Tests. Based on critiques we received of our past work, we were interested in whether the effect of trigger warnings differed in specific subgroups of our sample. It may be that trigger warnings are not helpful for trauma survivors broadly but are indeed helpful for those who have severe PTSD symptoms or have been diagnosed with PTSD. Accordingly, we tested the effect of trigger warnings among the subgroup of individuals who (1) met a clinical cutoff for a probable PTSD diagnosis or (2) reported a past diagnosis of PTSD.

Another possibility is that trigger warnings are only helpful when the content of the passage matches the traumatic experience of the survivor (i.e., the passage actually triggers remembrance of the trauma). We asked participants to identify "trauma-matching" passages, allowing for a direct test of this hypothesis. We selected only the individuals who specified trauma-matching passages and selected only the responses in reference to those specific passages. We then tested whether trigger warnings prior to these trauma-matching passages affected anxiety. In addition to testing trauma matching, we also tested whether the effect of trigger warnings on anxiety was moderated by the type of trauma reported by participants.

Results

Sample Characteristics

Our sample contained a majority of self-identified females (n = 239, 53%) with a significant minority of males (n = 208, 46%) and a small number of participants who specified another gender (n = 4, 1%). Participants had a mean age of 37 (SD = 11.2 years), and identified

their race as Caucasian (*n* = 336, 75%), Black/African American (*n* = 39, 9%), Asian/Pacific Islander (*n* = 23, 5%), Hispanic (*n* = 23, 5%), Native American/Alaska Native (*n* = 5, 1%), or multi-racial/selected multiple categories (*n* = 25, 6%). A substantial minority of participants identified their ethnicity as Hispanic (*n* = 41, 9%). Participants identified as not religious (*n* = 201, 45%), somewhat religious (*n* = 72, 16%), moderately religious (*n* = 79, 18%), very religious (*n* = 67, 15%), or extremely religious (*n* = 32, 7%). A minority of participants identified themselves as full-time undergraduate students (n = 44, 10%). Participants were skewed slightly toward liberal political orientation (*mean* = 2.64; 1 = *very liberal* to 5 = *very conservative*). Participants reported a wide diversity of traumatic experiences on the LEC-5. All 16 categories were represented, with the largest categories being natural disaster (*n* = 95, 21%), transportation accident (*n* = 79, 18%), sexual assault (*n* = 78, 17%), and physical assault (*n* = 47, 10%).

Preregistered Replication Tests

The results of the replication tests appear in Figure 1. Overall, replication tests either favored the null hypothesis or gave ambiguous evidence. In the original study by Bellet et al. (2018), a significant effect was found by trigger warning condition on perceived vulnerability to self (PPVS-S) and perceived vulnerability of others (PPVS-O). Neither of these significant effects replicated in our sample, with substantial evidence favoring the null hypothesis for an effect on perceived vulnerability of others (PPVS-O). A significant interaction effect was also found in the original experiment, such that participants' belief that words can harm (WCHS) moderated the effect of trigger warnings on immediate increases in anxiety. This interaction effect did not replicate in our sample, with substantial evidence favoring the null hypothesis. For immediate increases in anxiety or sensitization to anxiety (which were nonsignificant in the

original study), we found ambiguous evidence and substantial evidence favoring the null hypothesis, respectively.

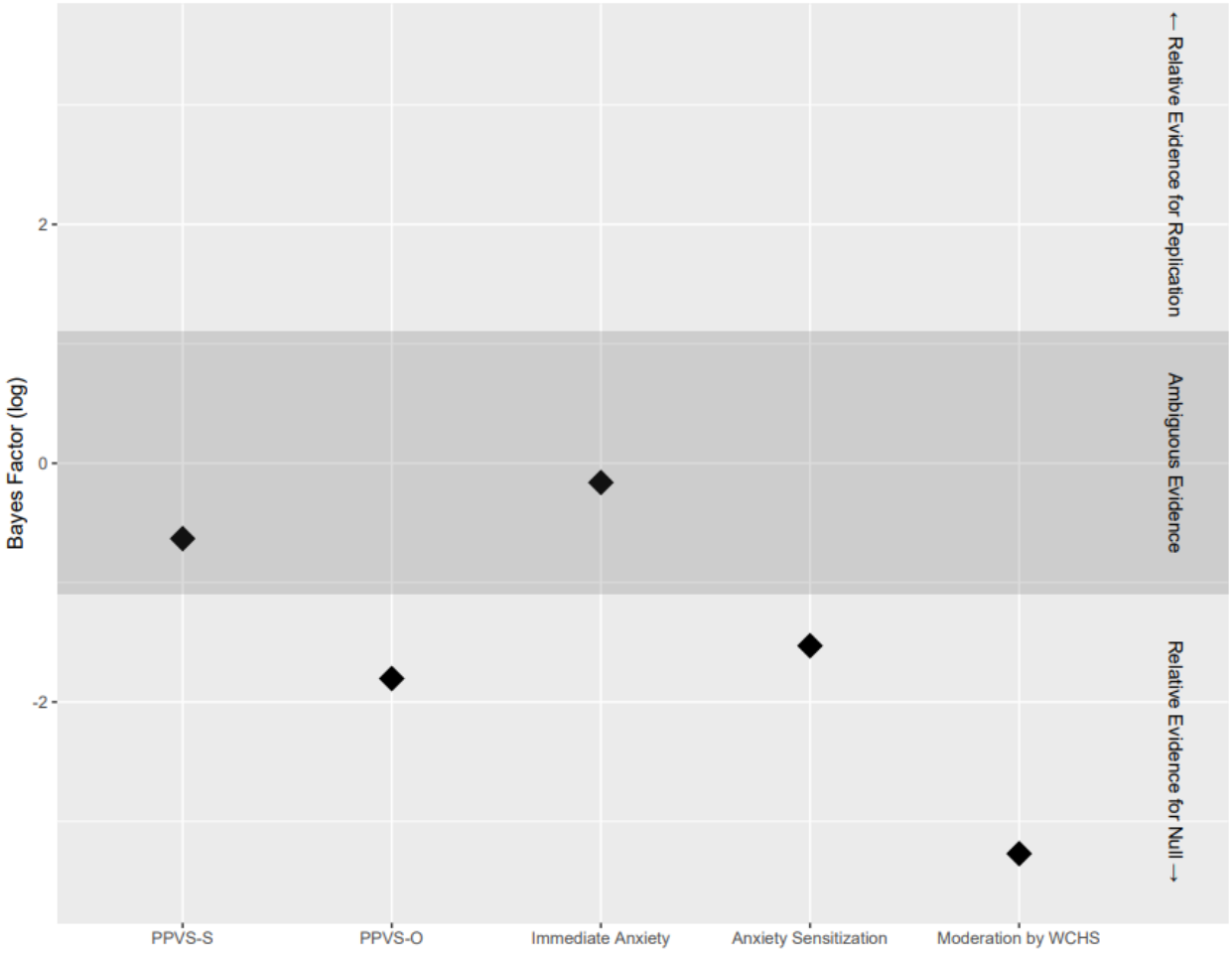

Figure 1. Replication Tests

Bayes Factors (log) are presented representing relative evidence for either the null hypothesis (t_{obs} = 0) or relative evidence for a hypothesis of equivalence with the effect from the original study (t_{obs} = t_{orig}). Overall, evidence was either ambiguous or favored the null hypothesis.

Trauma-Specific Preregistered Tests

First, we tested whether trigger warnings affected participants' ratings of trauma centrality on the CES. We found substantial evidence that trigger warnings increased the degree to which participants viewed their worst event as central to their life narrative (*BF* = 3.26).

Second, we tested whether PTSD severity scores on the PCL moderated any of the previous tests (e.g., effect on PPVS-S, PPVS-O, etc.). We found substantial evidence favoring the null hypothesis for a moderation effect on trauma centrality (*BF* = 0.11) and on perceived vulnerability (self or other; *BFs* = 0.10, 0.13). We found ambiguous evidence for a moderation effect of PTSD severity on anxiety sensitization (BF = 0.86). We found substantial evidence that PTSD severity moderates immediate anxiety reactions (BF = 3.14). That is, individuals who scored higher on the PCL had *increased* anxiety when they were given trigger warnings.

Third, we tested whether participants' self-reported prior exposure to trigger warnings (see TWAA) moderated any of the previous tests. We found substantial evidence favoring the null hypothesis for a moderation effect on trauma centrality (*BF* = 0.27), perceived vulnerability (self or other; *BFs* = 0.19, 0.22), and anxiety sensitization (*BF* = 0.21). We found ambiguous evidence for a moderation effect on immediate anxiety reaction (*BF* = 2.17).

Exploratory Tests

Critics of recent trigger warning research have suggested the plausible hypothesis that whereas trigger warnings may not be helpful for college students generally (e.g., Bellet et al., 2019) or even for trauma survivors generally, they may be helpful for more specific subpopulations. For instance, it is possible that trigger warnings are only helpful for (1) individuals with clinical-level PTSD symptoms or (2) individuals who have received a diagnosis of PTSD. Furthermore, it may be that trigger warnings are only helpful when (3) the content of

the literature passage directly matches the content of their trauma (i.e., it triggers a remembrance of the trauma). We tested each of these hypotheses in exploratory tests. The results of these tests are presented in Figure 2.

Full Sample. When comparing the null hypothesis ($t_{obs} = 0$) to an open alternative hypothesis ($t_{obs} \neq 0$) , the full sample showed substantial evidence favoring the null hypothesis ($BF = 0.14$, $n = 451$). In other words, trigger warnings did not appear to affect immediate anxiety reactions in our full sample.

Clinical Cutoff. When examining only individuals who met the cutoff of 33 on the PCL for a probable diagnosis of PTSD recommended by the United States Department of Veterans Affairs (Weathers et al., 2013), we found substantial evidence favoring the alternative hypothesis ($BF = 3.86$, $n = 150$). Among these individuals, trigger warnings increased immediate anxiety reactions. This is consistent with our preregistered test suggesting that PTSD severity scores moderated the effect of trigger warnings on anxiety reactions.

Self-Reported Diagnosis of PTSD. For individuals who self-reported receiving a past diagnosis of PTSD, we found substantial evidence favoring the null hypothesis ($BF = 0.32$, $n = 53$). That is, trigger warnings did not affect anxiety reactions for individuals who reported a diagnosis of PTSD.

Matching Trauma Passages. We asked individuals whether the passages triggered memories of their worst event and asked them to identify which passages were triggering. Examining only the individuals who reported triggering passages, and examining only the relevant passages, we found ambiguous evidence ($BF = 0.88$, $n = 133$) for an effect of trigger warnings on anxiety. The effect was in the direction of increasing anxiety. That is, individuals

who saw trigger warnings had trivially increased anxiety, suggesting that trigger warnings did not reduce anxiety reactions when passages matched past traumatic experiences.

Trauma Type. We used the LEC-5 to assess the type of trauma that best characterized each individual's worst event. Using the 16 categories from the LEC-5, we tested whether the type of trauma impacted the effect of trigger warnings. We found substantial evidence favoring the null hypothesis ($BF < 0.001$). However, some of the 16 categories had very few observations, limiting the statistical validity of the test. Therefore, we tested for the influence of trauma type by condensing the LEC-5 categories into 5 broad groups: sexual violence, other interpersonal violence, accidental injury or illness, natural or other disaster, and other. Using these categories, we again found substantial evidence favoring the null hypothesis ($BF = 0.003$). That is, the type of trauma did not moderate the effect of trigger warnings.

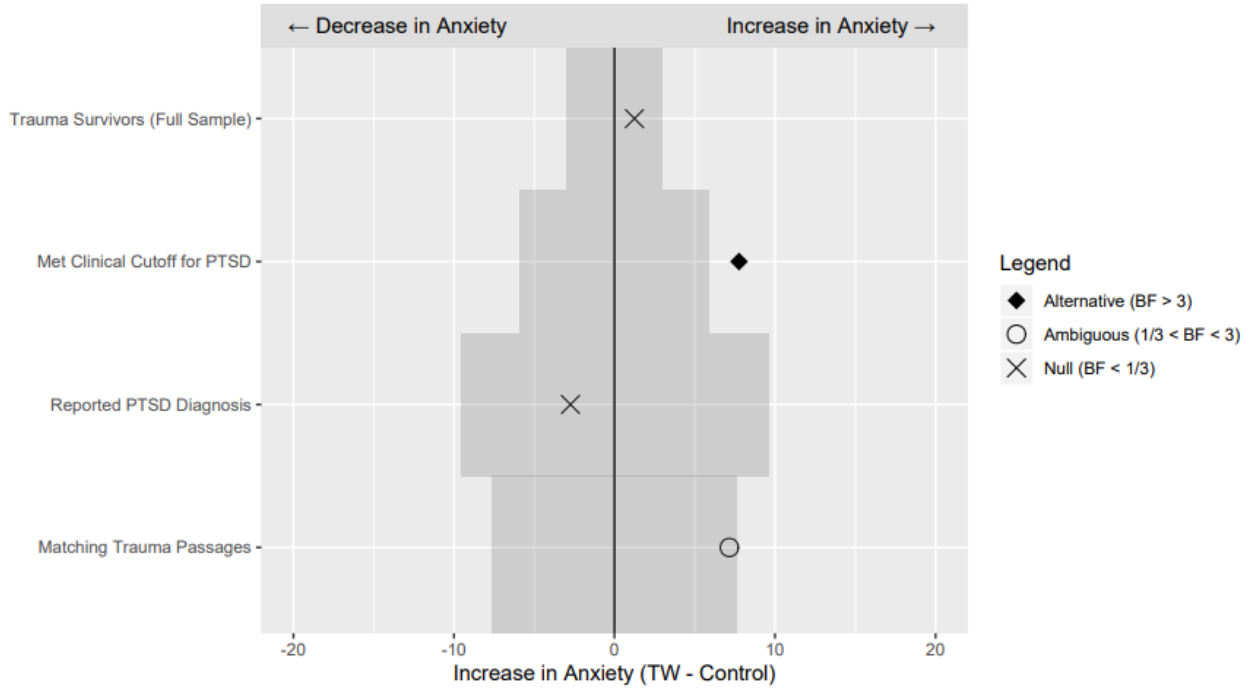

Figure 2. Trigger Warnings' Effect on Anxiety in Vulnerable Groups

Mean difference in anxiety change between the trigger warnings condition and the control condition across subgroups. Shapes correspond to a Bayesian comparison of the null hypothesis ($t_{obs} = 0$) and an alternative hypothesis ($t_{obs} \neq 0$). The shaded region corresponds to the boundaries of frequentist critical regions ($p < 0.05$, two-sided).

Discussion

Past research has indicated that trigger warnings are unhelpful in reducing anxiety. The results of this study are consistent with that conclusion. This study was the first to focus on a sample of people who had survived Criterion A trauma as defined by the DSM-5 (APA, 2013). Trigger warnings did not reduce anxiety for this sample broadly. Trigger warnings also did not reduce anxiety among those who met a clinical cutoff for PTSD symptoms, reported a diagnosis of PTSD, or those who reported that the stimuli matched the content of their past trauma. Trigger warnings appeared to have trivially small effects on response anxiety overall. When effects did emerge, they tended towards small *increases* in anxiety rather than decreases.

Bellet et al. (2018) previously found that trigger warnings increased individuals' projections of their own vulnerability to future trauma, as well as the vulnerability of others. Our results suggested substantial evidence that these effects did not replicate. Bellet et al. (2018) also reported that individuals who endorsed the belief that words are emotionally harmful showed greater anxiety in response to trigger warnings compared to individuals who did not endorse that belief. Again, we found substantial evidence that this effect did not replicate. One possibility is that these effects were unique to the trigger-warning naïve (trauma-naïve), crowd-sourced, older sample used by Bellet et al. (2018). However, given that these effects originally had a small effect size and did not replicate in larger samples of college students (Bellet et al., 2019) or trauma survivors (present study), the original results may have been a false positive.

We found substantial evidence that giving trigger warnings to trauma survivors caused them to view trauma as more central to their life narrative. This effect is a reason for worry. Some trigger warnings explicitly suggest that trauma survivors are uniquely vulnerable (e.g., "…especially in those with a history of trauma"). Even when trigger warnings only mention

content, the implicit message that trauma survivors are vulnerable remains (why else provide a warning?). These messages may reinforce the notion that trauma is invariably a watershed event that causes permanent psychological change. In reality, a majority of trauma survivors are resilient, experiencing little if any lasting psychological changes due to their experience (Bonanno, 2004; Bonanno & Mancini, 2008). Aggregated across various types of trauma, only 4% of potentially traumatic events result in PTSD (Liu et al., 2017). However, trauma survivors who view their traumatic experience as central to their life have elevated PTSD symptoms (Berntsen & Rubin, 2006; Brown et., 2010; Robinaugh & McNally, 2011). Trauma centrality prospectively predicts elevated PTSD symptoms, whereas the reverse is not true (Boals & Ruggero, 2016). Decreases in trauma centrality mediated therapy outcomes (Boals & Murrell, 2016). This suggests that increasing trauma centrality is directly countertherapeutic. In other words, trigger warnings may harm survivors by increasing trauma centrality.

We tested whether the severity of PTSD symptoms in our sample moderated any of our tested hypotheses. In most cases, we found either evidence for no moderation or ambiguous evidence. However, we did find substantial evidence that PTSD symptoms moderated the effect of trigger warnings on response anxiety. For individuals who had more severe PTSD, trigger warnings increased anxiety. This effect is ironic in the sense that trigger warnings may be most harmful for the individuals they were designed to protect. We found no evidence that individuals' prior exposure to trigger warnings moderated any of the previous effects.

A limitation of past research was that trigger warnings were primarily tested among individuals who were trauma-naïve or in mixed samples. That is, the possibility remained that despite being unhelpful for most who view them, trigger warnings may have been helpful for trauma survivors or individuals with PTSD. This study largely puts these questions to rest.

Trigger warnings were not helpful for trauma survivors. For individuals who met a clinical cutoff for severity of PTSD symptoms, trigger warnings slightly *increased* anxiety. Trigger warnings were not helpful for individuals who self-reported a diagnosis of PTSD. Perhaps most convincingly, trigger warnings were not helpful even when they warned about content that closely matched survivors' traumas. That is, when considering only the passages which participants reported as reminding them of past trauma, trigger warnings were still unhelpful.

Public arguments regarding trigger warnings have been politically charged, complex, and data-poor. Recent research on trigger warnings can importantly inform or perhaps even settle some of these debates. Trigger warnings are unhelpful for trauma survivors, college students, trauma-naïve individuals, and mixed groups of participants (Bellet et al., 2018; Bellet et al., 2019, Bridgland et al., 2019; Sanson et al., 2019). Given this consistent conclusion, we find no evidence-based reason for educators, administrators, or clinicians to use trigger warnings.

Whether trigger warnings are explicitly harmful is less clear. We found evidence that trigger warnings increase the narrative centrality of trauma among survivors, which is countertherapeutic (Boals & Murrell, 2016). We also found that trigger warnings increase anxiety for those with more severe symptoms of PTSD. Although these effects were preregistered and found in a large sample, the size of the effects were small and have not yet been rigorously tested across multiple studies. However, such knowledge is unnecessary to adjudicate whether to use trigger warnings – if there is no good reason to deploy them in the first place, we need not require strong evidence of harm before abandoning them. Trigger warnings should serve as an important caution to both clinical and nonclinical professionals who use interventions aimed to improve well-being among trauma survivors. Such practices should be

thoroughly vetted via appropriate scientific techniques before they are adopted. Using unvetted interventions is irresponsible to victims of trauma.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Bellet, B. W., Jones, P. J., & McNally, R. J. (2018). Trigger warning: Empirical evidence ahead. *Journal of Behavior Therapy and Experimental Psychiatry*, *61*, 134-141. https://doi.org/10.1016/j/jbtep.2018.07.002

Bellet, B. W., Jones, P. J., & McNally, R. J. (2019). Trigger warnings and resilience in college students: A preregistered replication. *Retrieved from osf.io/693gj*

Berntsen, D., & Rubin, D. C. (2006). The centrality of event scale: A measure of integrating a trauma into one's identity and its relation to post-traumatic stress disorder symptoms. *Behaviour Research and Therapy, 44*, 219-231.

Boals, A., & Murrell, A. R. (2016). I am > trauma: Experimentally reducing event centrality and PTSD symptoms in a clinical trial. *Journal of Loss and Trauma*, *21*, 471-483.

Boals, A., & Ruggero, C. (2016). Event centrality prospectively predicts PTSD symptoms. *Anxiety, Stress, & Coping, 29*, 533-541.

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events?. *American Psychologist*, *59*, 20.

Bonanno, G. A., & Mancini, A. D. (2008). The human capacity to thrive in the face of potential trauma. *Pediatrics, 121*, 369.

Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., & Keane, T. M. (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders–fifth edition (PCL-5) in veterans. *Psychological Assessment, 28*, 1379.

Boysen, G. A. (2017). Evidence-based answers to questions about trigger warnings for clinically-based distress: A review for teachers. *Scholarship of Teaching and Learning in Psychology, 3,* 163-177. http://dx.doi.org/10.1037/stl0000084

Bridgland, V. M., Green, D. M., Oulton, J. M., & Takarangi, M. K. (2019). Expecting the worst: Investigating the effects of trigger warnings on reactions to ambiguously themed photos. *Journal of Experimental Psychology: Applied*. Advance online publication. http://dx.doi.org/10.1037/xap0000215

Brown, A. D., Antonius, D., Kramer, M., Root, J. C., & Hirst, W. (2010). Trauma centrality and PTSD in veterans returning from Iraq and Afghanistan. *Journal of Traumatic Stress, 23*(4), 496-499.

Bruce, M. J. (2017). *Does trauma centrality predict trigger warning use? Physiological responses to using a trigger warning.* Poster presented at the annual meeting of the Midwest Psychological Association, Chicago, IL.

Ellison, J. (2016). Dear class of 2020 student. Retrieved from https://news.uchicago.edu/sites/default/files/attachments/Dear_Class_of_2020_Students.pdf

Gainsburg, I., & Earl, A. (2018). Trigger warnings as an interpersonal emotion-regulation tool: Avoidance, attention, and affect depend on beliefs. *Journal of Experimental Social Psychology, 79*, 252-263.

Karasek, S. (2016). Trust me, trigger warnings are helpful. *The New York Times*. Retrieved from https://www.nytimes.com/roomfordebate/2016/09/13/do-trigger-warnings-work/trust-me-trigger-warnings-are-helpful

Kimble, M. (2019). *Why the warning? Student responses to triggering material based on trauma history and symptom profiles.* Symposium presented at the International Conference for Psychological Science, Paris, France.

Liu, H., Petukhova, M. V., Sampson, N. A., Aguilar-Gaxiola, S., Alonso, J., Andrade, L. H., ... & Kawakami, N. (2017). Association of DSM-IV posttraumatic stress disorder with traumatic experience type and history in the World Health Organization World Mental Health Surveys. *JAMA Psychiatry, 74*, 270-281.

Lukianoff, G., & Haidt, J. (2015). The coddling of the American mind. *The Atlantic, 316*(2), 42-52.

McNally, R. J. (2016). If you need a trigger warning, you need PTSD treatment. *The New York Times.* Retrieved from https://www.nytimes.com/roomfordebate/2016/09/13/do-trigger-warnings-work/if-you-need-a-trigger-warning-you-need-ptsd-treatment

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. *R package version 0.9.12-4.2.* https://CRAN.R-project.org/package=BayesFactor

Robinaugh, D. J., & McNally, R. J. (2011). Trauma centrality and PTSD symptom severity in adult survivors of childhood sexual abuse. *Journal of Traumatic Stress, 24*(4), 483-486.

Sanson, M., Strange, D., & Garry, M. (2019). Trigger warnings are trivially helpful at reducing negative affect, intrusive thoughts, and avoidance. *Clinical Psychological Science.* Advance online publication. https://doi.org/10.1177%2F2167702619827018

Weathers, F.W., Blake, D.D., Schnurr, P.P., Kaloupek, D.G., Marx, B.P., & Keane, T.M. (2013). The Life Events Checklist for DSM-5 (LEC-5). *Instrument available from the National Center for PTSD at www.ptsd.va.gov*

Weathers, F.W., Litz, B.T., Keane, T.M., Palmieri, P.A., Marx, B.P., & Schnurr, P.P. (2013). The PTSD Checklist for DSM-5 (PCL-5). *Scale available from the National Center for PTSD at www.ptsd.va.gov*.

Wilson, R. (2015). Students' requests for trigger warnings grow more varied. *The Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Students-Requests-for/233043

Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., ... & Mintz, J. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment*, *28*, 1392.