Empirical Article

Clinical Psychological Science
2019, Vol. 7(4) 778–793

DOI: 10.1177/2167702619827018
www.psychologicalscience.org/CPS

Trigger Warnings Are Trivially Helpful at Reducing Negative Affect, Intrusive Thoughts, and Avoidance

Mevagh Sanson[1], Deryn Strange[2], and Maryanne Garry[1]
[1]School of Psychology, the University of Waikato, and [2]Department of Psychology, John Jay College of Criminal Justice, City University of New York

Abstract
Students are requesting and professors issuing *trigger warnings*—content warnings cautioning that college course material may cause distress. Trigger warnings are meant to alleviate distress students may otherwise experience, but multiple lines of research suggest trigger warnings could either increase or decrease symptoms of distress. We examined how these theories translate to this applied situation. Across six experiments, we gave some college students and Internet users a trigger warning but not others, exposed everyone to one of a variety of negative materials, then measured symptoms of distress. To better estimate trigger warnings' effects, we conducted mini meta-analyses on our data, revealing trigger warnings had trivial effects—people reported similar levels of negative affect, intrusions, and avoidance regardless of whether they had received a trigger warning. Moreover, these patterns were similar among people with a history of trauma. These results suggest a trigger warning is neither meaningfully helpful nor harmful.

Keywords
clinical cognition, teaching, trigger warning

Received 8/7/18; Revision accepted 11/3/18

Universities around the world are grappling with demands for *trigger warnings*—cautions to students about upcoming course content that may cause them distress (Medina, 2014; Palmer, 2017). The ideas that various topics may trigger distress—because the material itself is negative or reminds people of prior negative experiences—and warnings about the material's topic can prevent this distress have long circulated online (Vingiano, 2014). But now these ideas have spread to universities: Two recent surveys of U.S. professors found over half reported using trigger warnings about their course content (Kamenetz, 2016; National Coalition Against Censorship, 2015). Some professors believe trigger warnings help decrease their students' distress following an encounter with negative material, rather than merely allowing students to avoid that material altogether (e.g., Gust, 2016; Manne, 2015). Other professors, however, believe trigger warnings are not only an affront to academic freedom but might actually increase students' distress either by allowing students to avoid material altogether (thereby preventing them from learning to cope effectively with reminders of prior negative experiences) or encouraging a more negative reaction to material they do encounter and contributing to the rising levels of anxiety and depression among college students (American Association of University Professors, 2014; Center for Collegiate Mental Health [CCMH], 2016; Lukianoff & Haidt, 2018; McNally, 2014).

When it comes to the effects of trigger warnings, these conflicting positions are not simply ideological; they are also psychological. Yet when we turned to the psychological literature to find out what effects trigger warnings have, we found research suggesting they would be helpful, research suggesting they would be harmful, but no data directly addressing their effects—until we were writing this manuscript, when Bellet, Jones, and McNally (2018) published a single experiment. In this experiment,

Corresponding Author:
Maryanne Garry, School of Psychology, The University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand
E-mail: maryanne.garry@waikato.ac.nz

trigger warnings produced a small increase in people's self-reported anxiety after reading negative passages—but only among people who strongly believed words can cause emotional damage. In addition, trigger warnings led people to believe they and others were slightly more susceptible to emotional harm from future, hypothetical traumatic experiences. Taken together, these results suggest trigger warnings mostly did very little.

But there are at least three reasons Bellet et al.'s (2018) findings are not conclusive on the effects of trigger warnings. First, it is plausible that at least some of their primary findings become nonsignificant once corrected for the increased false discovery rate that arises from multiple comparisons. The rest of their findings regarding the effects of trigger warnings are nonsignificant without correction. This collection of null findings is difficult to interpret, particularly against a backdrop of a single experiment and a modest sample size. Second, anyone with a history of exposure to an extremely distressing event was not permitted to complete the experiment (comprising roughly 50% of those who started it; Jones, 2018). Considering that the majority of the population has been exposed to a potentially traumatic event (Breslau et al., 1998), this exclusion limits generalizability. Third, inasmuch as Bellet et al. tell us something about peripheral effects of trigger warnings, most crucially, they do not tell us much about trigger warnings' putative ability to alleviate people's symptoms of distress.

The fact that this issue remains unresolved is a problem—one we aim to remedy in this article. We systematically and empirically examined the consequences of trigger warnings for three symptoms of people's distress: negative affect following exposure to negative material, intrusive thoughts related to the negative material, and avoidance of reminders of the negative material. These symptoms at their extreme can constitute part of a posttraumatic stress disorder (PTSD) diagnosis (American Psychiatric Association [APA], 2013). But many people experience these symptoms following exposure to a traumatic experience or negative material without developing PTSD (Breslau et al., 1998; Rothbaum, Foa, Riggs, Murdock, & Walsh, 1992). These symptoms are also therefore the very ones that negative course materials could evoke—and that trigger warnings might alleviate or exacerbate.

Reasons Trigger Warnings Might Alleviate Distress

There are at least two reasons to expect that trigger warnings could alleviate people's symptoms of distress following their exposure to negative material. First, trigger warnings may prompt people to better regulate their emotions. We know from the literature on emotion regulation that people can decrease the negativity of their emotional response to something by anticipating that response and then proactively using one or more strategies to down-regulate it (for a review, see Gross, 2015; for a meta-analysis, see Webb, Miles, & Sheeran, 2012). In one study, before people watched a disgusting film clip, some were told to regulate their emotional response to the clip by reappraising it while they watched; others were told to act as though they were not emotionally affected, or to simply watch (Gross, 1998). Compared with people given these latter instructions, those instructed to regulate by reappraising had lesser emotional responses to the film on subjective, behavioral, and physiological measures. Suppose trigger warnings are akin to those regulation instructions—leading people to anticipate they will have a negative response to the negative material and prompting them to use strategies to down-regulate that negative response. If so, we should see people who received a trigger warning would report less negative affect following exposure to negative material than their unwarned counterparts. What is more, given that less negative material is less accessible in memory, people who received a trigger warning should also experience fewer intrusive thoughts related to the material (Hall & Berntsen, 2008).

Second, trigger warnings may play into people's existing beliefs about the helpful effects of being forewarned. We know from the literature on response expectancies that when people expect that a set of factors will produce a certain response, when encountering those factors, they unwittingly act in such a way to produce that response—the classic example is the placebo effect of a sham medical treatment (for reviews, see Kirsch, 1997; Price, Finniss, & Benedetti, 2008). Suppose people believe that a trigger warning about negative material will decrease their distress following exposure to that material. If so, receiving a warning might encourage people to be less aware of their internal states, such as emotions and intrusive thoughts, and so report fewer symptoms of distress compared with their unwarned counterparts. Of course, we do not know if trigger warnings have these helpful effects.

Reasons Trigger Warnings Might Exacerbate Distress

Conversely, there are at least four reasons to expect that trigger warnings could exacerbate symptoms of distress. First, inasmuch as the emotion regulation literature suggests trigger warnings may prompt people to better regulate their negative emotions, it is also true that the same literature suggests trigger warnings may lead people to be worse. That is, just as people can decrease the negativity of their emotional response,

they can also increase it (Gross, 2015; Webb et al., 2012)—and people might interpret trigger warnings as a tacit instruction to up-regulate their negative response to the material. Indeed, in two studies, when people were told they would see disturbing footage, they were more upset by it (Cantor, Ziemke, & Sparks, 1984; de Wied, Hoffman, & Roskos-Ewoldsen, 1997)—findings in line with classic cognitive work showing that prefacing information can alter how people interpret and remember the material that follows (Bransford & Johnson, 1972). Suppose trigger warnings lead people to use strategies that up-regulate their negative response. Then we should see that people who received a trigger warning would report more negative affect than their unwarned counterparts.

Second, trigger warnings may lead people to remember the material as more negative afterward. In one study, telling people an experience of theirs was especially negative led them to remember it more negatively; that is, they made judgments of their memory in line with the social feedback they received about it (Takarangi & Strange, 2010). This finding suggests that people who receive a trigger warning about negative material may likewise recast that material as more negative. If trigger warnings lead people to up-regulate their negative response, or recast the material as more negative, then either route (or both) should lead that material to be more accessible in memory. As a result, people who receive a trigger warning would experience more related intrusive thoughts (Hall & Berntsen, 2008).

Third, and relatedly, if trigger warnings lead people to find the material more negative and have more intrusive thoughts about it, people also might try to avoid thinking about that material more. Of course, trying to avoid thinking about something can, counterproductively, lead people to have even more intrusive thoughts about it (Harvey & Bryant, 1998; for a review, see Wenzlaff & Wegner, 2000).

Fourth, although the literature on response expectancies suggests trigger warnings might cause people to act in such a way as to decrease their distress, the same literature suggests trigger warnings could instead unwittingly cause them to increase their distress (Kirsch, 1997). After all, merely telling people about possible negative side effects of a treatment can lead them to report experiencing more of those side effects (Wells & Kaptchuk, 2012). Suppose trigger warnings create the belief that people will find the negative material that follows very distressing. If so, receiving a warning might encourage people to become hyperaware of (and inclined to negatively interpret) their emotions and intrusive thoughts, leading them to report more symptoms compared with their unwarned counterparts.

Finally, whether trigger warnings are helpful or harmful as applied to a classroom setting will depend on more than their direct effects on students' symptoms of distress. The more trigger warnings increase (or decrease) students' distracting, intrusive thoughts following their exposure to negative material, the worse (or better) those students should be able to comprehend other, unrelated material they subsequently encounter in class (Takarangi, Strange, & Lindsay, 2014; for a review, see Mooneyham & Schooler, 2013).

Determining what trigger warnings do—both the direction and magnitude of their effects—is practically and theoretically important. On the practical side, obviously it would be good to know how effective trigger warnings are at reducing distress given they are already being deployed at universities. But the effects of trigger warnings have wider practical implications, too. For example, if, on the one hand, we find trigger warnings meaningfully decrease distress, that would suggest they might also decrease the distress jurors sometimes experience when they grapple with graphic evidence (Lonergan, Leclerc, Descamps, Pigeon, & Brunet, 2016). If, on the other hand, we find trigger warnings meaningfully increase distress, that would call into question the use of the warnings that institutional review boards often require scientists include in their consent forms and those that media outlets include before disturbing content.

On the theoretical side, in trying to understand what trigger warnings do, we have to consider how they might do it. In other words, we need to examine how theory on emotion regulation and response expectancies translate into real-world situations—situations that do not tend to come with instructions directing people to implement some regulation strategy or form a particular expectancy. Yet the effectiveness of trigger warnings at reducing distress remains largely unexamined in the empirical literature (cf. Bellet et al., 2018). To fill this gap, we asked: To what extent do trigger warnings affect people's (a) ratings of negative material and their symptoms of distress, namely, (b) negative affect, (c) intrusive thoughts, and (d) avoidance? We gathered data from 1,394 people across six experiments in which we presented some people (but not others) with a trigger warning, exposed everyone to negative material, and measured their acute symptoms of distress. We then conducted mini meta-analyses on these data to more precisely estimate the sizes of trigger warnings' effects.

Method

Across six experiments, we examined the extent to which trigger warnings changed the rates of symptoms of distress—negative affect, intrusive thoughts, and avoidance—that subjects experienced after reading a negative story (Experiments 1a and 1b) or watching a more negative or less negative film clip (Experiments

2a, 2b, 3, and 4). We also measured how negative subjects judged that material to be and their comprehension of other material presented subsequently. In Experiment 3, to address the possibility that trigger warnings do not lead people to anticipate negative material, we additionally examined the extent to which trigger warnings influenced subjects' expectations about the material to follow. In Experiment 4, to address the possibility that trigger warnings' effects are different for people who have experienced traumatic events, we additionally asked about subjects' history of trauma.

Subjects

In Experiments 1a and 2a, we recruited introductory psychology students at Victoria University of Wellington, who participated toward fulfillment of a course requirement (1a: $n = 254$; 2a: $n = 130$). We collected as many subjects as we could according to departmental allocations that semester. In Experiments 1b, 2b, 3, and 4, we recruited workers on Amazon's Mechanical Turk platform (MTurk; https://www.mturk.com/), who participated in exchange for Amazon credit (1b: $n = 203$; 2b: $n = 395$; 3: $n = 460$; 4: $n = 438$). In line with projected rates of noncompliance for MTurk subjects (Oppenheimer, Meyvis, & Davidenko, 2009), we aimed to collect enough subjects so that we could exclude up to 30% of our sample and still retain a mean of 70 subjects per cell.

We used ESCI software's (Cumming, 2012) "precision for planning feature," which showed that our target n should be 69 subjects per warned/unwarned cell given the following three constraints. First, the true size of the effect of a trigger warning (relative to no warning) is $d = 0.4$, Ferguson's (2009) "minimum effect size of practical significance." Second, the target margin of error (i.e., half the width of a 95% confidence interval [CI]) around those effects is 0.35, allowing us to exclude zero as a plausible effect size. Third, we have what Cumming (2012) calls "99% assurance" of achieving this target margin of error.

We met this target cell size in almost every case. Our final cell sizes after exclusions (which we report in Table S1 in the Supplemental Material available online) were as follows: Experiment 1a: warning condition, $n = 121$, and no-warning condition, $n = 119$. Experiment 1b: warning condition, $n = 77$, and no-warning condition, $n = 67$. Experiment 2a: more negative condition, warning condition, $n = 25$, and no-warning condition, $n = 27$; less negative condition, warning condition, $n = 28$, and no-warning condition, $n = 28$. Experiment 2b: more negative condition, warning condition, $n = 70$, and no-warning condition, $n = 72$; less negative condition, warning condition, $n = 64$, and no-warning condition, $n = 73$. Experiment 3: more negative condition, warning condition, $n = 94$, and no-warning condition, $n = 74$; less negative condition, warning condition, $n = 73$, and no-warning condition, $n = 76$. Experiment 4: more negative condition, warning condition, $n = 80$, and no-warning group, $n = 77$; less negative condition, warning condition, $n = 71$, and no-warning condition, $n = 78$.

Subjects retained in the data sets ranged in age from 16 to 78 years, with median ages across experiments ranging from 18 to 33 years. The percentage of females ranged from 56% to 75%, and across our MTurk samples, the percentage of U.S. citizens ranged from 93% to 99%, and the percentage for whom English was their first language ranged from 97% to 99%. We present more detail about the characteristics of each experiment and the subjects who participated in them in Table S1.

Design

In Experiments 1a and 1b, we manipulated presence of trigger warning (warning, no warning); in Experiments 2a, 2b, 3, and 4, we also manipulated the negativity of film (more negative, less negative); both between-subjects. This additional factor allowed us to investigate the effects of trigger warnings across different kinds of materials.

Procedure

Experiments were approved by the School of Psychology Human Ethics Committee at Victoria University of Wellington (with reciprocal and additional approval granted at the University of Waikato when two authors relocated there) and conducted in accordance with the World Medical Association Declaration of Helsinki (other than Provision 35, regarding preregistration). We told subjects we were investigating factors that affect comprehension of different writing styles (Experiments 1a, 1b) or visual and verbal learning (Experiments 2a, 2b, 3, 4). That is, we did not alert subjects that they may be exposed to negative material. Subjects in Experiments 1a and 2a participated in a controlled environment at individual computers. Those in Experiments 1b, 2b, 3, and 4 were instructed to complete the experiment under similar conditions (these instructions appear in the Supplemental Material). The experiments proceeded in three phases (except for Experiment 4, in which there was a fourth phase).

Phase 1. To measure baseline negative affect, subjects completed the Positive and Negative Affect Schedule–Expanded form (PANAS-X; Watson & Clark, 1999; Experiments 1a and 1b) or the shorter, regular PANAS (Watson, Clark, & Tellegen, 1988; Experiments 2a, 2b, 3, and 4). On either version of the PANAS, subjects rate several affect-related words (e.g., *distressed*) according to how

much they feel that way "right now," from 1 (*very slightly or not at all*) to 5 (*extremely*). Ratings for items on the negative affect subscale—the same in both versions—are summed, yielding a total between 10 and 50. When subjects rate "the present moment," the negative subscale has good internal consistency, with Cronbach's alpha of .85 (Watson et al., 1988). Similarly, in the current experiments, Cronbach's alphas for the negative subscale ranged from .88 to .93. The subscale also has reasonable test-retest reliability across 2 months, with a correlation of .45 (Watson et al., 1988). It also has good external validity: Within-subject changes in ratings across a day regarding the present moment correlate with ratings of current stress (Watson et al., 1988). Further, ratings made about "today" correlate with the Hopkins Symptom Checklist, a measure of general distress over the past week, at .65 (Watson et al., 1988). We expected levels of baseline negative affect to be similar across subjects who were later randomly assigned to see a warning and those who were not. Indeed, they were, as Table S2 in the Supplemental Material shows.

Phase 2. Next, subjects either randomly saw or did not see a trigger warning. We developed the warnings by drawing on examples online and guidelines issued by student associations. In Experiments 1a and 1b, *warning* subjects read: "TRIGGER WARNING: The following story contains violence and death," whereas those in Experiments 2a, 2b, 3, and 4 read: "TRIGGER WARNING: The following video may contain graphic footage of [a fatal car crash/violent domestic abuse]. You might find this content disturbing" (brackets indicate differences between counterbalances). *No warning* subjects received no information about the material to follow.

Subjects were then randomly assigned to encounter negative material. Across experiments, all the materials were about topics that would typically be the target of a trigger warning (i.e., child abuse, murder, a car accident, and physical domestic abuse).

Subjects in Experiments 1a and 1b were randomly assigned to read one of two negative fictional short stories (more information about these materials is presented in the Supplemental Material). In these early experiments, we used stories because trigger warnings are often used (or requested) before such written material. We selected these materials on the basis of norming data showing that subjects found them negative (norming details are presented in the Supplemental Material).

But in subsequent experiments, we used films, for at least three reasons. First, we suspected they would be more engaging—and therefore be rated as more negative and elicit higher rates of symptoms. Second, each film had a version of lesser negativity and a version of greater negativity (e.g., a pair of road-safety public-service-announcement clips, with one version in which a speeding driver narrowly escapes a serious accident and another version in which the accident occurs). We expected *more negative film* subjects would experience more symptoms than *less negative film* subjects but that trigger warnings would have similar effects on symptom frequency for both kinds of material (Rubin, Berntsen, & Bohni, 2008). Third, the use of similar films in trauma research is well established (Holmes & Bourne, 2008; James et al., 2016).

Thus, subjects in 2a, 2b, 3, and 4 were randomly assigned to watch one of four short clips from public service announcement campaigns. Two counterbalanced clips showed the events of greater negativity; the other two clips were their less negative (or control) counterparts, showing the same events unfolding in a less negative way (more information about these materials is presented in the Supplemental Material). Our norming data verified that subjects experienced the more negative and less negative versions as classified (norming details are presented in the Supplemental Material).

Subjects in Experiment 3 completed an additional step prior to watching the film but after warning subjects had been warned. To determine the extent to which trigger warnings influence subjects' expectations about the material to follow, all subjects in that experiment first rated how negative, positive, surprising, interesting, unpleasant, distressing, and disgusting they thought the upcoming film was going to be from 1 (*not at all*) to 7 (*extremely*).

Phase 3. After subjects had read or watched the material, we measured symptoms of distress using four tasks completed in the following order.

PANAS. First, to measure negative affect, subjects again completed the PANAS-X or PANAS regarding how they felt right now.

Tally of intrusions. Second, to measure intrusions, we asked subjects to read one of two randomly assigned nonfiction articles (about atoms or cells; Smallwood, Nind, & O'Connor, 2009). We told them "your primary purpose is to understand what you are reading," but we also asked them to press the *x* key each time "you notice that you are experiencing an intrusive memory or thought about the [story/video] that you just [read/watched]" (Takarangi et al., 2014). These presses yielded a tally of intrusions. Subjects spent approximately 3 min reading the article and recording intrusions (we report exact reading times in the Supplemental Material). Afterward, they rated their adherence to noting intrusions from 0 (*not at all well*) to 10 (*extremely well*).

Impact of Event Scale. Third, to measure intrusions and avoidance, subjects completed the Impact of Event Scale (IES; Horowitz, Wilner, & Alvarez, 1979), rating how frequently items (e.g., "Pictures about it popped into my mind" and "I tried to remove it from memory") were true for them following a stressful event—here, reading the story or watching the film—on a 4-point scale in which 0 = *not at all*, 1 = *rarely*, 3 = *sometimes*, and 5 = *often*; ratings were then summed. Both subscales have good internal consistency, with Cronbach's alphas of .86 for intrusions and .82 for avoidance; and good external validity, showing moderate to strong correlations with other measures of distress and PTSD (Sundin & Horowitz, 2002). We omitted two nonrelevant items from the intrusion subscale (concerning sleep), so the range for the intrusion subscale was 0 to 25, and for the avoidance subscale, it was 0 to 40. We nonetheless achieved good internal consistency, with Cronbach's alphas in the current experiments ranging from .78 to .90 for the intrusion subscale and .72 to .83 for the avoidance subscale.

Comprehension. Fourth, to measure how well subjects comprehended the nonfiction article, they answered five four-alternative forced choice questions about it (Smallwood et al., 2009). We expected that to the extent subjects experienced more intrusions while reading, their comprehension would decrease (Mooneyham & Schooler, 2013; Takarangi et al., 2014). In addition, subjects in Experiments 1a and 1b answered five questions about their comprehension of the negative story they read (the story comprehension results and more information about those questions are presented in the Supplemental Material).

After the comprehension task, all subjects (except those in Experiment 1a) then rated the story or film clip on how negative, positive, surprising, and interesting it was from 1 (*not at all*) to 7 (*extremely*). Subjects in Experiments 2a, 2b, 3, and 4 also rated how unpleasant, distressing, and disgusting the material was.

Phase 4 (specific to Experiment 4). Subjects in Experiment 4 took part in an additional phase to measure their prior exposure to traumatic events—they completed the Trauma History Screen (THS; Carlson et al., 2011). On this measure, subjects first review a list of potentially traumatic events (e.g., "a really bad car, boat, train, or airplane accident") and report how many times each type of event has happened to them. We divided the accidents item in two, yielding one item that asked specifically about car accidents (relevant to one of our film clip pairs) and one that asked about other accidents. We also added a new item, "domestic abuse—physical or psychological" (relevant to our other film clip pair).

Next, subjects indicate on the THS whether any of the listed experiences had "really bothered [them] emotionally"; those who say "yes" then answer a series of questions about each of up to five of those events. These questions include which type of event it was (with reference to the initial list), how long they were bothered by it (*not at all*, *1 week*, *2–3 weeks*, or *a month or more*), and how much it bothered them emotionally (*not at all, a little, somewhat, much*, or *very much*).

Subjects who report exposure to anything on the initial list of events are said to have experienced *high magnitude stressors* (HMS). Subjects who report exposure to one of these events, which bothered them emotionally, much or very much, and for a month or more, are said to have experienced *persisting posttraumatic distress* (PPD).

Across a variety of samples, both the HMS and PPD scores have high test-retest reliability, with correlations ranging from .74 to .93 for HMS and .73 to .95 for PPD across periods ranging from 1 week to 2 months. Both scores also have good convergent validity: HMS scores correlated highly with other measures of traumatic experiences, ranging from .73 to .81, and both scores correlated at least moderately with measures of PTSD symptoms, ranging from .22 to .41 for HMS and .18 to .38 for PPD. Further, those subjects who indicated at least one PPD event had higher rates of PTSD symptoms than those who indicated no PPD events (Carlson et al., 2011).

Finally, after the experiment proper, subjects in all experiments answered questions relevant to our exclusion criteria (see the following) and their demographics. Some subjects in 2a and 2b made an additional, exploratory set of ratings at this point; we did not analyze those data and do not consider them further. Subjects in 1b, 2b, 3, and 4 also saw attention checks and compliance checks throughout the experiments (Oppenheimer et al., 2009). We report details of these checks and all end-of-experiment questions in the Supplemental Material.

Results

Exclusions

We present the details of our exclusion criteria in Table S1 in the Supplemental Material. In brief, we excluded data from subjects whose responses indicated they had not correctly completed one or more critical tasks, had previously read or watched the negative material, or for other idiosyncratic reasons (see Table S1 note). Across experiments, the rate of exclusions ranged from 6% to 31%.

Manipulation checks

Before addressing our research question, we carried out two manipulation checks: First, we established that most warning subjects remembered seeing the warning,

which they did (Experiment 1a: 95%; Experiment 1b: 97%; Experiment 2a: 81%; Experiment 2b: 93%; Experiment 3: 90%; Experiment 4: 94%). Second, using data from Experiment 3, we checked that preexposure, warning subjects thought the films would be more negative than their no warning counterparts. They did: $M_{Warning}$ = 5.63, 95% CI = [5.45, 5.81]; $M_{NoWarning}$ = 3.15, 95% CI = [2.94, 3.36]; 95% CI for the difference = [2.20, 2.75] (see Table S3 in the Supplemental Material for other preexposure ratings). These findings are important because they indicate that subjects took note of our warning manipulation. Moreover, they demonstrate that trigger warnings notably worsen subjects' expectations about the material to follow, which has been—until now—an open question.

We now turn to our primary research question, which we address in four parts in the following: determining the extent to which trigger warnings affect subjects' (a) ratings of the material and three symptoms of their distress, namely, (b) negative affect, (c) intrusive thoughts, and (d) avoidance. We address our primary question using mini meta-analyses (Cumming, 2012). Because each subject responded to multiple dependent measures and we were interested in the effect of trigger warnings on each of these measures, we followed Borenstein, Hedges, Higgins, and Rothstein's (2009) suggestion to conduct a separate meta-analysis for each measure.[1]

Although responses on some of our measures were skewed, the skew was similar across warning and no warning conditions. Further, transformations did not normalize them, and so we analyzed untransformed data.

Part 1. Rating of material

To what extent did trigger warnings affect how negatively subjects judged the material? Very little, as the rows labeled "Negative rating" in Tables 1 and 2 show: Across experiments, warning subjects who saw a negative story or more negative film rated those materials similarly negatively as their no warning counterparts. Likewise, although subjects who saw a less negative film gave lower ratings than their more negative film counterparts, those who saw a warning gave similar ratings as those who got no such warning (for other ratings, see Table S3). Thus, the materials performed as expected even though trigger warnings had little effect.

To more precisely estimate the size of warnings' effect on ratings of negativity, we gathered the data from all our experiments to conduct a random effects model mini meta-analysis in ESCI software (Cumming, 2012). This mini meta-analysis revealed that warning subjects rated the materials just 0.15 less negatively, 95% CI = [–0.29, –0.01], than no warning subjects (the maximum possible difference was 6). Notice that the CI around this difference establishes a narrow range of plausible values for the true size of the effect, all of which are trivial. The total variability between effect size estimates was low, Q = 7.98; moreover, meaningful variance across effect size estimates was low, I^2 = 0%, and the estimated standard deviation of the distribution of true effect sizes was small, T = 0, 95% CI = [0, 0.30], indicating little heterogeneity. In standardized terms, this reduction is $d_{Unbiased}$ = –0.14, 95% CI = [–0.26, –0.03] (also known as Hedge's g, calculated by dividing the mean difference by the pooled standard deviation and applying a bias correction; Cumming, 2012). We display this effect along with the standardized effect of trigger warnings on each of our other measures in Figure 1.

One might reasonably wonder if by including the less negative conditions in the meta-analysis, we washed out the effect of trigger warnings in the more negative conditions. The answer is no: When we repeated this meta-analysis using only data from subjects who encountered a negative story or more negative film clip, the outcome was very similar (we report those results in the Supplemental Material). In summary, the best estimate indicates that providing a trigger warning about material—written or visual, of greater or lesser negativity—only very slightly decreased how negative subjects judged that material to be.

Part 2. Negative affect

To what extent did trigger warnings influence how negative subjects felt? Again, the answer is very little, as the rows labeled "PANAS negative" in Tables 1 and 2 show. That is, after reading a negative story or watching a film of greater or lesser negativity, warning subjects felt similarly negative as their no warning counterparts.

To more precisely estimate the size of trigger warnings' effect on negative affect, we conducted another random effects model mini meta-analysis. This meta-analysis revealed that after exposure to the material, warning subjects felt 0.25 points more negative affect, 95% CI = [–0.51, 1.00], than no warning subjects (maximum possible difference was 40). Again, the CI around this difference establishes a narrow plausible range for the true effect size, all of which are trivial. Further, there was little heterogeneity, Q = 7.11, I^2 = 0%, T = 0, 95% CI = [0, 1.30]. As Figure 1 shows, this increase is $d_{Unbiased}$ = 0.02, 95% CI = [–0.08, 0.13] in standardized terms.

When we repeated this meta-analysis using only data from subjects who encountered the more negative materials, the outcome was very similar (see the Supplemental Material). Thus, seeing a trigger warning had

Table 1. Descriptive Statistics for Main Measures of Symptoms of Distress Classified by Presence of Warning and Negativity of Material

		Warning and more negative material		No warning and more negative material		Warning and less negative material		No warning and less negative material	
Measure	Experiment	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
Negative rating	1a	—	—	—	—				
	1b	6.16	[5.94, 6.37]	6.10	[5.87, 6.34]				
	2a	6.52	[6.23, 6.81]	6.30	[5.81, 6.78]	2.96	[2.26, 3.67]	3.14	[2.54, 3.75]
	2b	6.46	[6.21, 6.71]	6.61	[6.43, 6.79]	2.78	[2.34, 3.22]	3.03	[2.58, 3.48]
	3	6.17	[5.88, 6.46]	6.49	[6.28, 6.69]	1.96	[1.65, 2.27]	2.61	[2.19, 3.02]
	4	6.35	[6.09, 6.61]	6.47	[6.20, 6.74]	2.65	[2.24, 3.05]	2.87	[2.48, 3.27]
PANAS negative	1a	19.78	[18.31, 21.25]	18.70	[17.46, 19.93]				
	1b	17.45	[15.73, 19.18]	16.15	[14.47, 17.83]				
	2a	19.56	[16.98, 22.14]	20.48	[17.53, 23.43]	15.68	[13.04, 18.31]	15.82	[12.76, 18.88]
	2b	21.04	[19.06, 23.03]	22.18	[20.06, 24.30]	13.89	[12.51, 15.27]	13.79	[12.61, 14.98]
	3	20.83	[18.90, 22.76]	21.69	[19.82, 23.56]	14.64	[13.29, 15.99]	15.17	[13.51, 16.83]
	4	21.61	[19.42, 23.81]	22.61	[20.65, 24.57]	15.59	[13.87, 17.31]	13.71	[12.45, 14.96]
Intrusions tally	1a	6.86	[5.93, 7.80]	7.06	[6.08, 8.03]				
	1b	4.48	[3.32, 5.64]	5.21	[3.84, 6.58]				
	2a	10.83	[5.76, 15.90]	8.16	[5.76, 10.57]	7.08	[4.95, 9.21]	8.71	[5.44, 11.99]
	2b	7.02	[5.42, 8.62]	7.50	[6.04, 8.96]	4.97	[3.72, 6.21]	4.79	[3.53, 6.06]
	3	7.27	[5.69, 8.85]	5.55	[4.49, 6.61]	4.51	[3.37, 5.65]	5.41	[4.00, 6.83]
	4	6.64	[5.22, 8.05]	7.27	[5.84, 8.70]	4.78	[3.66, 5.90]	5.74	[4.33, 7.16]
IES intrusions	1a	10.69	[9.46, 11.91]	11.15	[9.97, 12.33]				
	1b	9.61	[7.94, 11.28]	10.33	[8.50, 12.15]				
	2a	11.48	[9.16, 13.80]	14.22	[11.75, 16.69]	11.11	[8.66, 13.55]	9.64	[7.23, 12.06]
	2b	11.54	[9.95, 13.14]	13.81	[12.10, 15.52]	7.56	[6.07, 9.05]	8.41	[6.87, 9.95]
	3	11.55	[9.95, 13.16]	13.05	[11.55, 14.56]	7.95	[6.34, 9.55]	9.29	[7.67, 10.91]
	4	12.45	[10.81, 14.09]	12.70	[10.96, 14.44]	8.28	[6.69, 9.88]	8.09	[6.59, 9.59]
Comprehension	1a	0.44	[0.39, 0.48]	0.46	[0.42, 0.51]				
	1b	0.65	[0.59, 0.71]	0.59	[0.53, 0.66]				
	2a	0.62	[0.51, 0.74]	0.40	[0.30, 0.50]	0.59	[0.49, 0.70]	0.44	[0.32, 0.55]
	2b	0.60	[0.53, 0.67]	0.60	[0.53, 0.67]	0.57	[0.51, 0.64]	0.53	[0.46, 0.61]
	3	0.55	[0.50, 0.61]	0.52	[0.47, 0.58]	0.56	[0.50, 0.63]	0.57	[0.50, 0.63]
	4	0.50	[0.43, 0.57]	0.53	[0.46, 0.60]	0.54	[0.48, 0.61]	0.50	[0.43, 0.57]
IES avoidance	1a	14.60	[13.06, 16.13]	14.96	[13.42, 16.49]				
	1b	16.78	[14.43, 19.13]	15.10	[12.76, 17.45]				
	2a	13.84	[10.37, 17.31]	16.30	[13.51, 19.08]	13.64	[9.75, 17.54]	12.89	[10.07, 15.72]
	2b	17.66	[15.80, 19.51]	19.19	[17.14, 21.25]	13.80	[11.31, 16.29]	14.95	[12.69, 17.20]
	3	16.15	[14.04, 18.26]	18.12	[16.00, 20.24]	13.16	[10.79, 15.54]	13.74	[11.54, 15.94]
	4	17.35	[15.24, 19.46]	17.73	[15.89, 19.56]	13.34	[11.10, 15.58]	12.36	[10.42, 14.29]

Note: Dashes indicate that subjects in Experiment 1a did not rate how negative the story was. PANAS = Positive and Negative Affect Schedule (Watson & Clark, 1999; Watson, Clark, & Tellegen, 1988); IES = Impact of Event Scale (Horowitz, Wilner, & Alvarez, 1979).

almost no effect on how negative people felt following a variety of materials.

Part 3. Intrusions

To what extent did trigger warnings influence the frequency of intrusions subjects experienced? The answer is very little—on any of the three measures, as the rows in Tables 1 and 2 labeled, respectively, "Intrusions tally," "IES intrusions," and "Comprehension" show. That is, warning and no warning subjects reported similar numbers of intrusions about the stories or films while they read the nonfiction article; they rated their intrusions as similarly frequent on the IES, and they performed similarly well on the article comprehension test.

Subjects reported high adherence to noting intrusions while reading (Experiment 1a: $M = 7.30$, 95% CI = [7.06, 7.55]; Experiment 1b: $M = 8.75$, 95% CI =

Table 2. Raw Effect Sizes of Presence of Warning and Negativity of Film Conditions on Main Measures of Symptoms of Distress

		Effect of warning presence on more negative material		Effect of warning presence on less negative material		Effect of film negativity after warning		Effect of film negativity after no warning	
Measure	Experiment	*M* difference	95% CI for the difference	*M* difference	95% CI for the difference	*M* difference	95% CI for the difference	*M* difference	95% CI for the difference
Negative rating	1a	—	—						
	1b	0.05	[–0.26, 0.37]						
	2a	0.22	[–0.34, 0.79]	–0.18	[–1.08, 0.73]	3.56	[2.78, 4.33]	3.15	[2.39, 3.92]
	2b	–0.15	[–0.46, 0.15]	–0.25	[–0.87, 0.38]	3.68	[3.19, 4.16]	3.58	[3.10, 4.07]
	3	–0.32	[–0.69, 0.06]	–0.65	[–1.16, –0.13]	4.21	[3.79, 4.63]	3.88	[3.42, 4.34]
	4	–0.12	[–0.49, 0.25]	–0.22	[–0.79, 0.34]	3.70	[3.23, 4.17]	3.60	[3.12, 4.07]
PANAS negative	1a	1.08	[–0.83, 2.99]						
	1b	1.31	[–1.10, 3.71]						
	2a	–0.92	[–4.77, 2.93]	–0.14	[–4.09, 3.80]	3.88	[0.27, 7.50]	4.66	[0.50, 8.82]
	2b	–1.14	[–4.02, 1.75]	0.10	[–1.69, 1.88]	7.15	[4.71, 9.59]	8.39	[5.99, 10.78]
	3	–0.86	[–3.58, 1.86]	–0.53	[–2.66, 1.61]	6.19	[3.71, 8.67]	6.52	[4.04, 9.00]
	4	–1.00	[–3.92, 1.93]	1.89	[–0.20, 3.97]	6.02	[3.20, 8.84]	8.91	[6.60, 11.21]
Intrusions tally	1a	–0.20	[–1.54, 1.15]						
	1b	–0.73	[–2.50, 1.04]						
	2a	2.67	[–2.67, 8.01]	–1.63	[–5.45, 2.18]	3.75	[–1.40, 8.91]	–0.55	[–4.54, 3.44]
	2b	–0.48	[–2.63, 1.66]	0.18	[–1.59, 1.94]	2.05	[0.01, 4.08]	2.71	[0.80, 4.62]
	3	1.72	[–0.28, 3.72]	–0.90	[–2.71, 0.91]	2.76	[0.72, 4.80]	0.13	[–1.63, 1.89]
	4	–0.64	[–2.64, 1.36]	–0.96	[–2.77, 0.85]	1.85	[0.03, 3.68]	1.53	[–0.46, 3.53]
IES intrusions	1a	–0.47	[–2.16, 1.23]						
	1b	–0.72	[–3.16, 1.73]						
	2a	–2.74	[–6.06, 0.58]	1.46	[–1.89, 4.82]	0.37	[–2.94, 3.68]	4.58	[1.21, 7.95]
	2b	–2.26	[–4.58, 0.06]	–0.85	[–2.99, 1.29]	3.98	[1.81, 6.15]	5.39	[3.11, 7.67]
	3	–1.50	[–3.73, 0.73]	–1.34	[–3.61, 0.92]	3.61	[1.32, 5.90]	3.76	[1.57, 5.96]
	4	–0.25	[–2.62, 2.12]	0.19	[–1.98, 2.36]	4.17	[1.89, 6.45]	4.61	[2.33, 6.89]
Compre-hension	1a	–0.03	[–0.09, 0.04]						
	1b	0.06	[–0.03, 0.15]						
	2a	0.22	[0.08, 0.37]	0.16	[0.01, 0.31]	0.03	[–0.12, 0.18]	–0.04	[–0.18, 0.11]
	2b	0.00	[–0.09, 0.10]	0.04	[–0.06, 0.14]	0.03	[–0.07, 0.12]	0.07	[–0.03, 0.16]
	3	0.03	[–0.05, 0.11]	–0.01	[–0.10, 0.09]	–0.01	[–0.09, 0.08]	–0.04	[–0.13, 0.04]
	4	–0.03	[–0.13, 0.07]	0.04	[–0.05, 0.13]	–0.04	[–0.14, 0.05]	0.03	[–0.07, 0.13]
IES avoidance	1a	–0.36	[–2.52, 1.80]						
	1b	1.67	[–1.63, 4.98]						
	2a	–2.46	[–6.76, 1.85]	0.75	[–3.95, 5.45]	0.20	[–4.94, 5.34]	3.40	[–0.47, 7.28]
	2b	–1.54	[–4.29, 1.21]	–1.15	[–4.47, 2.17]	3.86	[0.82, 6.90]	4.25	[1.23, 7.27]
	3	–1.97	[–4.99, 1.04]	–0.57	[–3.78, 2.63]	2.98	[–0.17, 6.14]	4.38	[1.35, 7.42]
	4	–0.38	[–3.16, 2.41]	0.98	[–1.94, 3.90]	4.01	[0.96, 7.07]	5.37	[2.72, 8.02]

Note: The effect of warning presence (for each type of material) was calculated by subtracting no warning subjects' mean from warning subjects' mean; positive differences indicate higher scores for subjects who received a trigger warning. The effect of film negativity (in each warning condition) was calculated by subtracting less negative subjects' mean from more negative subjects' mean; positive differences indicate higher scores for subjects who watched a more negative film. Dashes indicate that subjects in Experiment 1a did not rate how negative the story was. PANAS = Positive and Negative Affect Schedule (Watson & Clark, 1999; Watson, Clark, & Tellegen, 1988); IES = Impact of Event Scale (Horowitz, Wilner, & Alvarez, 1979).

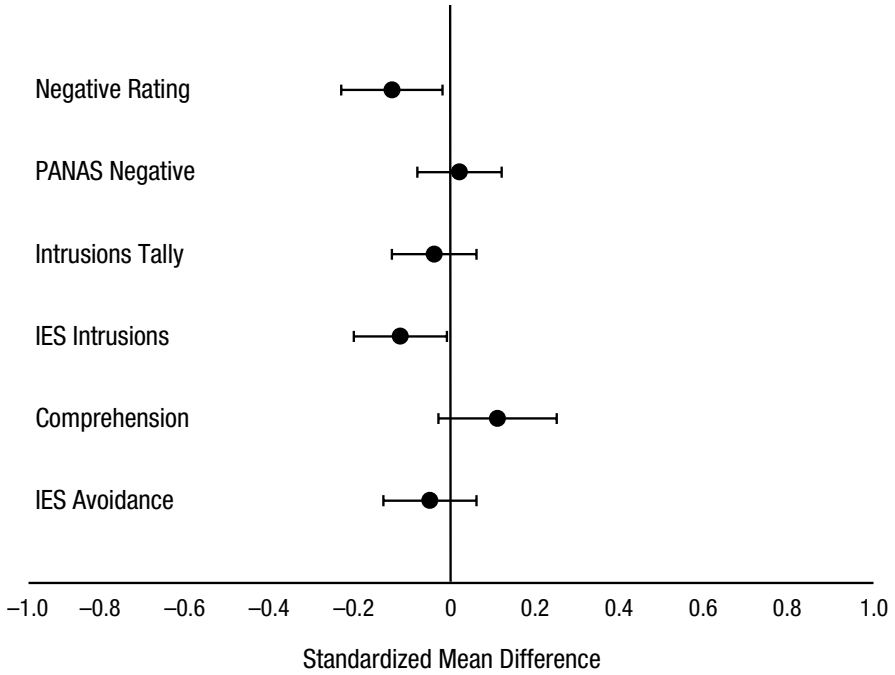

Fig. 1. Standardized meta-analytic effect of trigger warning for each dependent measure, expressed as d_{Unbiased}. Negative values indicate a lower mean score for warning subjects; positive values indicate a higher mean score for warning subjects. The horizontal whiskers show the 95% confidence interval around each effect. PANAS = Positive and Negative Affect Schedule (Watson & Clark, 1999; Watson, Clark, & Tellegen, 1988); IES = Impact of Event Scale (Horowitz, Wilner, & Alvarez, 1979).

[8.47, 9.03]; Experiment 2a: M = 7.81, 95% CI = [7.43, 8.20]; Experiment 2b: M = 8.61, 95% CI = [8.40, 8.83]; Experiment 3: M = 8.53, 95% CI = [8.32, 8.75]; Experiment 4: M = 8.51, 95% CI = [8.27, 8.74]), data that suggest they took the task seriously. But to correct for the small number of subjects who reported very high tallies, we Winsorized the data for each experiment such that the value of any tally exceeding the 95th percentile for that cell was replaced with the value of the 95th percentile (Sheskin, 2003).

Then, to more precisely estimate the sizes of trigger warnings' effects on intrusions, we conducted three more random effects model mini meta-analyses. The first of these meta-analyses revealed that warning subjects reported 0.27 fewer intrusions, 95% CI = [–0.88, 0.34], than no warning subjects—a difference of about quarter of a thought. The CI around this difference is narrow and spans a range of trivial values. There was little heterogeneity, Q = 7.33, I^2 = 0%, T = 0, 95% CI = [0, 1.09]. As Figure 1 shows, this reduction is d_{Unbiased} = –0.04, 95% CI = [–0.14, 0.07], in standardized terms. The second of these meta-analyses revealed that warning subjects rated their intrusions on the IES as 0.84 less frequent, 95% CI = [–1.56, –0.11], than no warning subjects (maximum possible difference was 25). The CI around this difference is narrow and spans a range of trivial values. There was little heterogeneity, Q = 6.54, I^2 = 0%, T = 0, 95% CI = [0, 1.14]. In standardized terms, this reduction is d_{Unbiased} = –0.12, 95% CI = [–0.23, –0.02] (see Fig. 1). The third of these meta-analyses revealed that the proportion of comprehension questions warning subjects got correct was 0.03 greater, 95% CI = [–0.01, 0.07], than the proportion no warning subjects got correct (maximum possible difference was 1). The CI around this difference is narrow and spans a range of trivial values. Although there was some variability between effect size estimates, Q = 15.39, and a moderate proportion of this variance was meaningful, I^2 = 41.52%, its absolute magnitude was low, T = 0.04, 95% CI = [0, 0.08]. In standardized terms, this increase is d_{Unbiased} = 0.11, 95% CI = [–0.03, 0.25] (see Fig. 1).

When we repeated these meta-analyses using only data from subjects who encountered the more negative materials, the outcomes were very similar (see the Supplemental Material). Put differently, although most people experienced intrusive thoughts related to the material they saw, a trigger warning beforehand only slightly decreased the frequency of these intrusions.

Part 4. Avoidance

To what extent did trigger warnings influence subjects' avoidance? As the rows labeled "IES avoidance" in Tables 1 and 2 show, the answer, once again, is very

little. Warning and no warning subjects rated their avoidance symptoms as similarly frequent.

To more precisely estimate the size of this effect, we conducted a random effects model mini meta-analysis on these data. This meta-analysis revealed that warning subjects rated their avoidance as 0.50 less frequent, 95% CI = [–1.45, 0.45], than no warning subjects (maximum possible difference was 40). The CI around this difference is narrow and spans a range of trivial values. There was little heterogeneity, $Q = 5.47$, $I^2 = 0\%$, $T = 0$, 95% CI = [0, 1.25]. As Figure 1 shows, this reduction is $d_{\text{Unbiased}} = -0.05$, 95% CI = [–0.16, 0.05] in standardized terms.

When we repeated this meta-analysis using only data from subjects who encountered the more negative materials, the outcome was very similar (see the Supplemental Material). That is, seeing a trigger warning only slightly decreased people's attempts to avoid thinking about the various negative materials to which we exposed them.

History of trauma

Of course, a critic could argue that trigger warnings may be most helpful to particular groups of people. More specifically, one possibility is that trigger warnings have larger effects on people who because they have a history of trauma are predisposed to being upset by the warned-about material. A second possibility is that trigger warnings specifically benefit people whose history of trauma most overlaps with the content of the warned-about material. For instance, someone who has been in a serious car accident may benefit from being warned about upcoming footage depicting such an accident more so than someone who has not. We used data from Experiment 4 to address these two possibilities.

These analyses should be interpreted somewhat cautiously because of two concerns that arise when doing subset analyses—namely, a reduction in the precision with which we are able to estimate effects and the breaking of the random assignment of subjects to condition and counterbalance. That said, the counterpoint to the first concern is that although we are focusing on a smaller number of subjects, if these are the very people for whom trigger warnings are most beneficial, then the effects of trigger warnings on them should be more clearly distinguishable from trivial-sized effects. As for the second concern, it is true we were unable to randomly assign subjects' history of trauma, but as we report in the following, even among these subsets, there were still relatively similar numbers of subjects in the warning versus no warning conditions.

To address the first possibility, we began by determining the proportion of people in our sample who had any history of trauma. We found that almost all of our sample did: 271 subjects, or 89%, 95% CI [85, 92], reported experiencing at least one HMS—a finding in line with other research (Breslau et al., 1998). Put another way, the overall effects of trigger warnings found in Experiment 4 (reported in Table 2) suggest that even for people with a history of trauma, the effects of trigger warnings are trivial.

Still, there is arguably a vast difference between people who have had potentially traumatic experiences and the subset of people for whom those historical traumas caused persisting distress. Would trigger warnings be effective for this subset? The answer is probably no. Of the 89% of people who had experienced at least one HMS, 53% of them, 95% CI = [47, 59] (or, 47% of the sample overall, 95% CI = [42, 53]), reported PPD as a result of at least one of their traumatic experiences (and these subjects were distributed across all our conditions: more negative material: warning condition, $n = 36$, and no-warning condition, $n = 34$; less negative material: warning condition, $n = 37$, and no-warning condition, $n = 37$. We restricted our analyses to just this subset and report those results in Table 3 in the rows labeled "PPD." As Table 3 shows, the effects of trigger warnings were—contrary to what some may have predicted—often in the harmful direction and mostly still small. Moreover, the CIs around these differences were wide and spanned a range of values that in all but one case included zero and effects in the opposite direction.

In standardized terms, the effect sizes for more negative film subjects in this subset were: rating of material: $d_{\text{Unbiased}} = -0.34$, 95% CI = [–0.81, 0.13]; negative affect: $d_{\text{Unbiased}} = 0.19$, 95% CI = [–0.28, 0.66]; intrusion tally: $d_{\text{Unbiased}} = -0.04$, 95% CI = [–0.50, 0.43]; IES intrusions: $d_{\text{Unbiased}} = -0.03$, 95% CI = [–0.50, 0.44]; comprehension: $d_{\text{Unbiased}} = -0.49$, 95% CI = [–0.97, –0.02]; avoidance: $d_{\text{Unbiased}} = 0.11$, 95% CI = [–0.36, 0.58]). The effect sizes for less negative film subjects in the subset were: rating of material: $d_{\text{Unbiased}} = 0.08$, 95% CI = [–0.38, 0.53]; negative affect: $d_{\text{Unbiased}} = 0.40$, 95% CI = [–0.06, 0.86]; intrusion tally: $d_{\text{Unbiased}} = 0.17$, 95% CI = [–0.28, 0.63]; IES intrusions: $d_{\text{Unbiased}} = 0.28$, 95% CI = [–0.18, 0.74]; comprehension: $d_{\text{Unbiased}} = 0.24$, 95% CI = [–0.22, 0.69]; avoidance: $d_{\text{Unbiased}} = 0.03$, 95% CI = [–0.42, 0.49]).

We now turn to the second possibility, that people whose history of trauma overlaps with the content of the warned-about material are the very ones who would benefit the most from trigger warnings. To address this possibility, we first determined which subjects had experienced the events portrayed in our materials. We found that 132 (43%) subjects had experienced a "really bad car accident" and 112 (37%) subjects had experienced physical or psychological domestic abuse. These

Table 3. Descriptive Statistics for Main Measures of Symptoms of Distress and Raw Effect Sizes of Presence of Warning for Subjects With a History of Trauma

		Warning and more negative material		No warning and more negative material		Warning and less negative material		No warning and less negative material		Effect of warning presence on more negative material		Effect of warning presence on less negative material	
Measure	Experiment 4 subset	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M* difference	95% CI difference	*M* difference	95% CI for the difference
Negative rating	PPD	6.53	[6.19, 6.87]	6.79	[6.63, 6.96]	2.62	[2.04, 3.20]	2.49	[1.90, 3.08]	–0.27	[–0.64, 0.11]	0.14	[–0.68, 0.95]
	Content	6.50	[6.19, 6.81]	6.50	[5.99, 7.01]	—	—	—	—	0.00	[–0.54, 0.54]	—	—
PANAS negative	PPD	23.97	[20.49, 27.45]	22.21	[19.36, 25.05]	15.78	[13.06, 18.51]	13.00	[11.20, 14.80]	1.77	[–2.68, 6.21]	2.78	[–0.42, 5.99]
	Content	22.14	[19.15, 25.13]	24.55	[20.34, 28.75]	—	—	—	—	–2.41	[–7.33, 2.52]	—	—
Intrusions tally	PPD	8.22	[5.71, 10.73]	8.47	[6.11, 10.83]	5.50	[3.76, 7.25]	4.58	[2.77, 6.38]	–0.25	[–3.64, 3.13]	0.93	[–1.54, 3.39]
	Content	6.91	[4.63, 9.20]	7.05	[3.88, 10.21]	—	—	—	–	—0.13	[–3.87, 3.61]	—	—
IES intrusions	PPD	13.83	[10.98, 16.69]	14.09	[11.33, 16.85]	8.81	[6.47, 11.16]	6.97	[4.94, 9.00]	–0.25	[–4.16, 3.65]	1.84	[–1.21, 4.89]
	Content	12.61	[10.20, 15.03]	13.32	[10.06, 16.58]	—	—	—	—	–0.71	[–4.62, 3.21]	—	—
Comprehension	PPD	0.41	[0.30, 0.53]	0.57	[0.46, 0.68]	0.57	[0.48, 0.66]	0.50	[0.41, 0.60]	–0.16	[–0.31, 0.00]	0.06	[–0.06, 0.19]
	Content	0.46	[0.35, 0.57]	0.52	[0.38, 0.66]	—	—	—	—	–0.06	[–0.24, 0.12]	—	—
IES avoidance	PPD	18.61	[15.40, 21.82]	17.68	[14.92, 20.43]	13.05	[9.94, 16.17]	12.73	[9.60, 15.86]	0.93	[–3.24, 5.11]	0.32	[–4.01, 4.66]
	Content	19.47	[16.38, 22.56]	19.09	[15.21, 22.97]	—	—	—	—	0.38	[–4.49, 5.26]	—	—

Note: Dashes indicate that we did not calculate values for "Content subset" subjects who saw less negative material. The effect of warning presence (for each type of material) was calculated by subtracting no warning subjects' mean from warning subjects' mean; positive differences indicate higher scores for subjects who got a trigger warning. PPD = subset of subjects who reported persisting posttraumatic distress as a result of a potentially traumatic experience; Content = subset of subjects who were randomly assigned to watch a more negative clip with content that matched a potentially traumatic experience they had had; PANAS = Positive and Negative Affect Schedule (Watson & Clark, 1999; Watson, Clark, & Tellegen, 1988); IES = Impact of Event Scale (Horowitz, Wilner, & Alvarez, 1979).

data suggest our materials depicted common traumas. But more to the point, when we restricted our analyses to the subset of subjects who had been randomly assigned to watch a more negative clip with content that matched a trauma they had experienced (warning condition, n = 36, no-warning condition, n = 22), did we find evidence trigger warnings were helpful? We did not. As Table 3 shows, in the rows labeled "Content," we instead found the effects of trigger warnings were once again trivially small. In standardized terms, the effects for this subset were: rating of material: d_{Unbiased} = 0.00, 95% CI = [–0.40, 0.40]; negative affect: d_{Unbiased} = –0.26, 95% CI = [–0.57, 0.05]; intrusion tally: d_{Unbiased} = –0.02, 95% CI = [–0.61, 0.57]; IES intrusions: d_{Unbiased} = –0.10, 95% CI = [–0.54, 0.34]; comprehension: d_{Unbiased} = –0.17, 95% CI = [–0.74, 0.39]; avoidance: d_{Unbiased} = 0.04, 95% CI = [–0.68, 0.76]).

Taken together, the results of these exploratory analyses do not fit with the idea that trigger warnings are uniquely helpful for people with a history of trauma when "history of trauma" is defined any one of several different ways. Rather, these analyses suggest trigger warnings have trivial effects even among people for whom such warnings may be specifically intended.

Discussion

We conducted six experiments investigating the effects of trigger warnings. Meta-analyses revealed that people who saw trigger warnings, compared with people who did not, judged material to be similarly negative, felt similarly negative, experienced similarly frequent intrusive thoughts and avoidance, and comprehended subsequent material similarly well. Although some measures yielded effects in a "trigger warnings are helpful" direction, these effects were so small as to lack practical significance (Ferguson, 2009). As a reference point, a Cochrane review found the standardized mean difference in self-reported symptoms between those who underwent therapy for PTSD and controls was –1.60, 95% CI = [–2.02, –1.18] (Bisson, Roberts, Andrew, Cooper, & Lewis, 2013). Of course, trigger warnings are not intended to substitute for therapy; nevertheless, the symptom reductions we observed are minuscule in comparison. Moreover, our meta-analytic confidence intervals were narrow, suggesting high precision—yet still showed trigger warnings plausibly have no effect or might even work slightly in the direction of causing harm (Cumming, 2012).

A critic might wonder if some subjects found the warnings helpful because they withdrew from the experiments after being warned, thereby avoiding the material and any ensuing symptoms (Gross, 2015). But when we examined responses from subjects who quit our experiments before completion, we found similar proportions quit in the warned and unwarned conditions, and the number of subjects who quit specifically after seeing the warning was very small (none in Experiment 1a, 1b, or 2a; nine in Experiment 2b; six in Experiment 3; one in Experiment 4), suggesting few if any subjects used the warning to avoid negative material (for more detail about when subjects quit, see the Supplemental Material). Moreover, because avoidance is a PTSD symptom (APA, 2013), the use of warnings to avoid material could be construed as harmful.

There are other possible reasons to explain why trigger warnings exerted only trivial effects, and those accounts are less interesting. For example, perhaps subjects did not notice the trigger warning or the wording did not change expectations about the material to follow. But a large majority of warned subjects said they remembered seeing the warning, and in Experiment 3, we found that warned subjects expected the material to follow would be more negative than unwarned subjects. Of course, it is possible that for people's expectations to matter, warnings need to target people's beliefs about their symptoms rather than about the material (Kirsch, 1997). It is also possible the warnings did not constitute an obvious enough prompt that, or indeed how, people should prepare to regulate their emotions—possibilities that fit with the literature showing people often cannot optimally use warnings to adjust their behavior (Gross, 2015; Wilson & Brekke, 1994). We based our warnings on "real-world" uses, but future research could examine the effects of variations that target a range of theoretical issues.

Our results run contrary to findings that warnings encountered before films made people feel worse afterward (Cantor et al., 1984; de Wied et al., 1997). But given our warned and unwarned subjects found the material similarly negative, it makes sense they then reported symptoms to similar degrees (Hall & Berntsen, 2008; Rubin et al., 2008; Wenzlaff & Wegner, 2000). Indeed, it could be we found little effect of warnings because of a floor effect—our materials did not make people feel sufficiently negative, such that warnings could not help much. Yet for unwarned subjects who saw material of greater negativity, the 95% CIs around their reported negative affect did not overlap with the bottom of the scale. Put another way, although these scores could have been decreased by trigger warnings, they were not. In addition, our finding that more negative materials produced more symptoms of distress fits with the idea that over-accessible memories of traumatic experiences contribute to symptoms of PTSD, and our finding that trigger warnings produced similar effects before material of greater and lesser negativity fits with the idea that the same mechanisms are at play

during more and less traumatic experiences (Rubin et al., 2008). Our results also extend recent work suggesting trigger warnings had little effect on most people's self-reported anxiety; we showed that trigger warnings have little effect on people's distress (Bellet et al., 2018).

Taken together, our findings show that trigger warnings are at best trivially helpful. But this conclusion comes with at least three caveats and limitations. First, we did not recruit people with a history of psychopathology (e.g., those with a diagnosis of PTSD, anxiety, or depression), and so we do not know how well our results generalize to clinical populations (although the results of Experiment 4 fit with the idea that most of our subjects—like most of the population—have had a traumatic experience; Breslau et al., 1998). Second, we did not ask our subjects their socioeconomic status or education level, which limits our ability to characterize the samples on whom we tested the effects of trigger warnings. Our samples were, however, drawn from populations for whom trigger warnings are often provided. Third, trigger warnings may have nontrivial effects we did not measure. For example, we did not ask about the phenomenology of the intrusions, yet warnings may have altered the vividness of the intrusions, for instance (Takarangi & Strange, 2010). Further, we used only self-report measures rather than taking physiological measures of hyperarousal symptoms, for instance (APA, 2013). Indeed, when people can precisely predict the timing of an unpleasant experience (e.g., an electric shock), they have lesser physiological responses to it even though their ratings of its magnitude are unaffected (Lykken, Macindoe, & Tellegen, 1972; for a review, see Lykken & Tellegen, 1974). It is possible, therefore, that if trigger warnings allow people to predict an encounter with negative material, those warnings may reduce people's physiological responses to the negative material. These issues constitute interesting directions for future research.

Where do the current findings leave us? Some might wonder if professors should continue to issue trigger warnings. After all, if the warnings do not worsen distress and students believe the warnings are helpful, then why not? Put simply, people are not always good judges of the effects interventions have on themselves or others (Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2014; Wilson & Brekke, 1994), and the chronic effects of trigger warnings may be different from their acute effects. College students are increasingly anxious (CCMH, 2016), and widespread adoption of trigger warnings in syllabi may promote this trend, tacitly encouraging students to turn to avoidance, thereby depriving them of opportunities to learn healthier ways to manage potential distress.

Action Editor

Scott O. Lilienfeld served as action editor for this article.

Author Contributions

All authors developed the study concepts and contributed to study design. M. Sanson performed data collection, analysis, and interpretation under the supervision of M. Garry and D. Strange. M. Sanson drafted the manuscript, and D. Strange and M. Garry provided critical revisions. All the authors approved the final manuscript for submission.

Acknowledgments

We thank Melanie Takarangi, Benjamin Baird, and Women's Aid UK for sharing materials; Cassandra Burton-Wood for help editing films and collecting data; Jeff Foster for Javascript help; Geoff Cumming for his usual sage advice; and Gina Grimshaw for feedback on manuscript drafts.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

M. Sanson was supported by Victoria University of Wellington, the University of Waikato, and Fulbright New Zealand.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/2167702619827018

Note

1. We also thank Geoff Cumming for his advice on this point.

References

American Association of University Professors. (2014). *On trigger warnings*. Retrieved from https://www.aaup.org/report/trigger-warnings

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. doi:10.1016/j.jbtep.2018.07.002.

Bisson, J. I., Roberts, N. P., Andrew, M., Cooper, R., & Lewis, C. (2013). Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults. *Cochrane Database of Systematic Reviews*, *12*, CD003388. doi:10.1002/14651858.CD003388.pub4

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of

comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726. doi:10.1016/S0022-5371(72)80006-9

Breslau, N., Kessler, R. C., Chilcoat, H. D., Schultz, L. R., Davis, G. C., & Andreski, P. (1998). Trauma and posttraumatic stress disorder in the community: The 1996 Detroit area survey of trauma. *Archives of General Psychiatry*, *55*, 626–632. doi:10.1001/archpsyc.55.7.626

Cantor, J., Ziemke, D., & Sparks, G. C. (1984). Effect of forewarning on emotional responses to a horror film. *Journal of Broadcasting*, *28*, 21–31 doi:10.1080/08838158409386512

Carlson, E. B., Smith, S. R., Palmieri, P. A., Dalenberg, C., Ruzek, J. I., Kimerling, R., . . . Spain, D. A. (2011). Development and validation of a brief self-report measure of trauma exposure: The Trauma History Screen. *Psychological Assessment*, *23*, 463–477. doi:10.1037/a0022294

Center for Collegiate Mental Health. (2016). *2015 annual report* (Publication No. STA 15-108). Retrieved from https://sites.psu.edu/ccmh/files/2017/10/2015_CCMH_Report_1-18-2015-yq3vik.pdf

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

de Wied, M., Hoffman, K., & Roskos-Ewoldsen, D. R. (1997). Forewarning of graphic portrayal of violence and the experience of suspenseful drama. *Cognition and Emotion*, *11*, 481–494. doi:10.1080/026999397379890

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532–538. doi:10.1037/a0015808

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*, 224–237. doi:10.1037/0022-3514.74.1.224

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*, 1–26. doi:10.1080/1047840X.2014.940781

Gust, O. (2016, June 14). I use trigger warnings—but I'm not mollycoddling my students. *The Guardian*. Retrieved from https://www.theguardian.com/higher-education-network/2016/jun/14/i-use-trigger-warnings-but-im-not-mollycoddling-my-students

Hall, N. M., & Berntsen, D. (2008). The effect of emotional stress on involuntary and voluntary conscious memories. *Memory*, *16*, 48–57. doi:10.1080/09658210701333271

Harvey, A. G., & Bryant, R. A. (1998). The effect of attempted thought suppression in acute stress disorder. *Behaviour Research and Therapy*, *36*, 583–590. doi:10.1016/S0005-7967(98)00052-7

Holmes, E. A., & Bourne, C. (2008). Inducing and modulating intrusive emotional memories: A review of the trauma film paradigm. *Acta Psychologica*, *127*, 553–566. doi:10.1016/j.actpsy.2007.11.002

Horowitz, M. J., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine*, *41*, 209–218. doi:10.1097/00006842-197905000-00004

James, E. L., Lau-Zhu, A., Clark, I. A., Visser, R. M., Hagenaars, M. A., & Holmes, E. A. (2016). The trauma film paradigm as an experimental psychopathology model of psychological trauma: intrusive memories and beyond. *Clinical Psychology Review*, *47*, 106–142. doi:10.1016/j.cpr.2016.04.010

Jones, P. J. [paytonjjones]. (2018, July 29). We used a pretty strict definition of Criterion A, which ended up excluding about ~50% of people who signed up, if I remember correctly [Tweet]. Retrieved from https://twitter.com/paytonjjones/status/1023701821095915520

Kamenetz, A. (2016, September 7). Half of professors in NPR Ed survey have used "trigger warnings." *National Public Radio*. Retrieved from http://www.npr.org/sections/ed/2016/09/07/492979242/half-of-professors-in-npr-ed-survey-have-used-trigger-warnings

Kirsch, I. (1997). Response expectancy theory and application: A decennial review. *Applied and Preventive Psychology*, *6*, 69–79. doi:10.1016/S0962-1849(05)80012-5

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2014). Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science*, *9*, 355–387. doi:10.1177/1745691614535216

Lonergan, M., Leclerc, M.-E., Descamps, M., Pigeon, S., & Brunet, A. (2016). Prevalence and severity of trauma- and stressor-related symptoms among jurors: A review. *Journal of Criminal Justice*, *47*, 51–61. doi:10.1016/j.jcrimjus.2016.07.003

Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. New York, NY: Penguin Press.

Lykken, D. T., Macindoe, I., & Tellegen, A. (1972). Preception: Autonomic response to shock as a function of predictability in time and locus. *Psychophysiology*, *9*, 318–333. doi:10.1111/j.1469-8986.1972.tb03215.x

Lykken, D. T., & Tellegen, A. (1974). On the validity of the preception hypothesis. *Psychophysiology*, *11*, 125–132. doi:10.1111/j.1469-8986.1974.tb00833.x

Manne, K. (2015, September 19). Why I use trigger warnings. *The New York Times*. Retrieved from https://www.nytimes.com/2015/09/20/opinion/sunday/why-i-use-trigger-warnings.html

McNally, R. J. (2014, May 20). Hazards ahead: The problem with trigger warnings, according to the research. *Pacific Standard*. Retrieved from https://psmag.com/education/hazards-ahead-problem-trigger-warnings-according-research-81946

Medina, J. (2014, May 17). Warning: The literary canon could make students squirm. *The New York Times*. Retrieved from https://www.nytimes.com/2014/05/18/us/warning-the-literary-canon-could-make-students-squirm.html

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology*, *67*, 11–18. doi:10.1037/a0031569

National Coalition Against Censorship. (2015). What's all this about trigger warnings? Retrieved from http://ncac.org/wp-content/uploads/2015/11/NCAC-TriggerWarningReport.pdf

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872. doi:10.1016/j.jesp.2009.03.009

Palmer, T. (2017 March 28). Monash University trigger warning policy fires up free speech debate. Retrieved from http://www.abc.net.au/news/2017-03-28/monash-university-adopts-trigger-warning-policy/8390264

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought. *Annual Review of Psychology*, *59*, 565–590. doi:10.1146/annurev.psych.59.113006.095941

Rothbaum, B. O., Foa, E. B., Riggs, D. S., Murdock, T., & Walsh, W. (1992). A prospective examination of post-traumatic stress disorder in rape victims. *Journal of Traumatic Stress*, *5*, 455–475. doi:10.1007/BF00977239

Rubin, D. C., Berntsen, D., & Bohni, M. K. (2008). A memory-based model of posttraumatic stress disorder: Evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review*, *115*, 985–1011. doi:10.1037/a0013397

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures* (3rd ed.). Boca Raton, FL: CRC Press.

Smallwood, J., Nind, L., & O'Connor, R. C. (2009). When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Consciousness and Cognition*, *18*, 118–125. doi:10.1016/j.concog.2008.11.004

Sundin, E. C., & Horowitz, M. J. (2002). Impact of Event Scale: Psychometric properties. *British Journal of Psychiatry*, *180*, 205–209. doi:10.1192/bjp.180.3.205

Takarangi, M. K. T., & Strange, D. (2010). Emotional impact feedback changes how we remember negative autobiographical experiences. *Experimental Psychology*, *57*, 354–359. doi:10.1027/1618-3169/a000042

Takarangi, M. K. T., Strange, D., & Lindsay, D. S. (2014). Self-report may underestimate trauma intrusions. *Consciousness and Cognition*, *27*, 297–305. doi:10.1016/j.concog.2014.06.002

Vingiano, A. (2014, May 5). *How the "trigger warning" took over the Internet*. Retrieved from https://www.buzzfeed.com/alisonvingiano/how-the-trigger-warning-took-over-the-internet

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule–Expanded Form*. Retrieved from http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1011&context=psychology_pubs

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*, 775–808. doi:10.1037/a0027600

Wells, R. E., & Kaptchuk, T. J. (2012). To tell the truth, the whole truth, may do patients harm: The problem of the nocebo effect for informed consent. *The American Journal of Bioethics*, *12*, 22–29. doi:10.1080/15265161.2011.652798

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*, 59–91. doi:10.1146/annurev.psych.51.1.59

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117–142. doi:10.1037/0033-2909.116.1.117