Journal Pre-proofs

Something distressing this way comes: The effects of trigger warnings on avoidance behaviors in an analogue trauma task

Victoria M.E. Bridgland, Melanie K.T. Takarangi

PII: S0005-7894(21)00130-1
DOI: https://doi.org/10.1016/j.beth.2021.10.005
Reference: BETH 1127

To appear in: *Behavior Therapy*

Received Date: 17 May 2021
Accepted Date: 18 October 2021

Please cite this article as: V.M.E. Bridgland, M.K.T. Takarangi, Something distressing this way comes: The effects of trigger warnings on avoidance behaviors in an analogue trauma task, *Behavior Therapy* (2021), doi: https://doi.org/10.1016/j.beth.2021.10.005

Something distressing this way comes: The effects of trigger warnings on avoidance behaviors in an analogue trauma task

Victoria M. E. Bridgland

&

Melanie K.T. Takarangi

Flinders University

Corresponding Author:
Melanie Takarangi
Flinders University
GPO Box 2100
Adelaide
South Australia, 5001
melanie.takarangi@flinders.edu.au

Abstract

Avoidance is one of the purported benefits and harms of trigger warnings—alerts that upcoming content may contain traumatic themes. Yet, previous research has focused primarily on emotional responses. Here, we used a trauma analogue design to assess people's avoidance behavior in response to stimuli directly related to an analogue trauma event. University undergraduates (n = 199) watched a traumatic film and then viewed film image stills preceded by either a trigger warning or a neutral task instruction. Participants had the option to "cover" and avoid each image. Apart from a minor increase in avoidance when a warning appeared in the first few trials, we found that participants did not overall avoid negative stimuli prefaced with a trigger warning any more than stimuli without a warning. In fact, participants were reluctant overall to avoid distressing images; only 12.56% (n = 25)

participants used the option to cover such images when given the opportunity to do so. Furthermore, we did not find any indication that trigger warning messages help people to pause and emotionally prepare themselves to view negative content. Our results contribute to the growing body of literature demonstrating that warnings seem trivially effective in achieving their purported goals.

Keywords: *trauma, trigger warnings, avoidance*

A scene depicting dubious sexual consent in Netflix's regency drama series *Bridgerton* (2020) was the recent cause of online uproar, with one Twitter user noting: "I was considering watching…however I have decided I will not because rape scenes on screen are a trigger for me" (Logan, 2020). Echoing these concerns, news outlets were quick to point out that the "controversial sex scene needs a trigger warning"—an alert that upcoming content may contain themes similar to traumatic experiences that could "trigger" someone to re-experience a traumatic event (Jean-Philippe, 2020), for example via vivid thoughts, feelings or flashbacks about the event (Ehlers, Hackmann, & Michael, 2004). These criticisms are not new, and in 2021 Netflix is under increasing pressure to add trigger warnings to their content (Medhora, 2021). Advocates of this idea—and of trigger warning use in other domains (e.g., academic and online contexts)—claim that such warnings are necessary so that people have the chance to emotionally prepare for, or to completely avoid, the content (George & Hovey, 2020; Strothman, 2021). However, thus far, the extant research has focused primarily on people's emotional reactions to novel content following trigger warnings and less so on how trigger warnings may—or may not—lead to avoidance behaviors. Moreover, no previous research has accounted for the *uniqueness* of trauma triggers—which typically relate to stimuli similar to a traumatic event (Ehlers et al., 2004). Here, we addressed these shortcomings of previous research. Using a trauma analogue design, we assessed people's

avoidance behaviors in response to stimuli directly related to an analogue trauma event. We had two main aims: First, to investigate if participants would be more likely to avoid content associated with an analogue trauma film when that content is preceded by a trigger warning message; and second, to assess if participants' avoidance behavior would be related to a) their emotional responses to the stimuli, and b) individual difference factors such as state and trait anxiety, experiential avoidance, perceptions of harm, reported use of avoidance strategies, and perceptions that trigger warnings are beneficial.

Paradoxically, avoidance is both one of the purported harms *and* one of the potential benefits of implementing trigger warnings. Critics (e.g., Lukianoff & Haidt, 2015) argue that trigger warnings contribute to a "counsel of avoidance" culture (McNally, 2019, as quoted in Flaherty, 2019). In support of these concerns, a substantial literature implicates avoidance as a primary maintaining factor in Posttraumatic Stress Disorder (PTSD; Badour, Blonigen, Boden, Feldner, & Bonn-Miller, 2012) as well as a central characteristic of a broad range of mental disorders (Krypotos, Effting, Kindt, & Beckers, 2015). In some exceptions to this rule, avoidance can assist with fear reduction within the early stages of treatment (Rachman, Radomsky, & Shafran, 2008) and increase adherence to concurrent exposure therapy (Levy & Radomsky, 2014). Indeed, warning advocates claim trauma survivors should be able to decide if they want to avoid content that may trigger re-experiencing symptoms, arguing that avoidance aids recovery (Cripps, 2020). However, regardless of long-term harms or benefits, little to no research has addressed the underlying assumption that trigger warnings influence—i.e., by promoting or not—avoidance-type behaviors.

Advocates also claim trigger warnings allow time for people to emotionally "prepare" for, or cope with content. For instance, Strothman (2021) of the Digital Citizens Academy—which focuses on educating the public about mental health—writes that trigger warnings allow people "*to take a step back and pause or pass over the content*." That is, seeing a

trigger warning should help someone to bring helpful strategies to mind (e.g., reappraisal, e.g., "*a television show is not real life and therefore poses no real threat*"), thus reducing negative emotional reactions. However, previous research—focused primarily on people's emotional responses towards content following warning messages (e.g., how warning messages make people feel; Bellet, Jones, & McNally, 2018; Boysen, et al., 2021; Bridgland, Green, Oulton, & Takarangi, 2019; Bruce, 2019)—shows that trigger warnings *do not* ameliorate distress. One possibility could be that despite advocates' claims, trigger warnings do not help people "*pause and prepare*" by bringing coping strategies to mind.

In the first empirical investigation to explore how trigger warning messages may or may not change the coping strategies people bring to mind, we asked participants to report what they would do when encountering a trigger warning related to their most stressful/traumatic experience (Bridgland, Barnard, & Takarangi, 2021). These participants reported a similar number of approach-based strategies (e.g., reappraising a situation in order to reduce its emotional impact) and avoidance-based strategies (e.g., leaving the situation) to participants who imagined the same hypothetical situation of encountering trauma-related content, but without a warning. However, due to the intention-behavior gap—i.e., that intentions do not always map onto future behavior (Sheeran & Webb, 2016)—asking people to hypothetically simulate the future may not capture actual behavior. To directly investigate avoidance behaviors, participants would need the opportunity to avoid content following a trigger warning.

Only a handful of studies have explicitly examined whether people avoid material accompanied by warning messages. Gainsburg and Earl (2018) asked participants to select a film to watch from a series of titles and found no difference in how often participants selected titles accompanied by a trigger warning or no warning. But, participants' anticipated anxiety about warned-of content, as well as the belief that trigger warnings are protective from real

harm (vs. coddling/overprotective), was associated with increased avoidance of material accompanied by trigger warnings. These results suggest that trigger warnings may encourage avoidance only among certain people.

However, other research has not found increased avoidance among groups who are traditionally associated with the belief that trigger warnings are helpful in reducing distress (i.e., trauma survivors). Bruce and Roberts (2020) found no preference for articles labelled with trigger warnings compared to the same titles without warnings, including among participants who had experienced trauma matching the article (i.e., interpersonal violence). Kimble et al. (2021) found only a small minority of participants (< 6%)—including those with a history of trauma or with probable PTSD—avoided potentially triggering text when provided with a non-triggering alternative option. Similarly, we found that when presented with a single Instagram sensitivity screen (a version of a trigger warning focused on blurring graphic content), only 10-15% of people opt to avoid potentially distressing content, while the remainder opt to uncover and reveal the image (Bridgland, Bellet & Takarangi, 2021). Indeed, Simister, Bridgland, and Takarangi (2021) found that more than 90% of people repeatedly uncover content covered by a sensitivity screen, even after exposure to graphic photo imagery underneath. Importantly, neither study found evidence that users with mental health concerns (e.g., symptoms of depression or PTSD) were any more likely to use the screens to avoid sensitive content. In fact, in one of Bridgland et al.'s studies, participants' desire to view potentially negative content covered by a warning screen was associated with risk markers for psychopathology (e.g., lowered wellbeing). Trigger warnings in this instance may therefore foster a "Forbidden Fruit effect" (Ringold, 2002)—where a restricted behavior becomes more desirable—and encourage morbid curiosity about distressing content (Oosterwijk, 2017). Taken together, early research provides an incomplete account of how warnings may affect avoidance behaviors.

A persisting problem with *all* previous trigger warning research is the *uniqueness* of triggers. That is, trauma triggers typically relate to stimuli with similarities to the traumatic event, unique to a person's individual situation (Ehlers et al., 2004). For instance, while a fictional depiction of sexual assault (e.g., on *Bridgerton*) may trigger one person to have re-experiencing symptoms (i.e., be 'triggered'), the same scene may not elicit any response in another person who has also experienced sexual assault. Therefore, a crucial piece of information is still missing from trigger warning research: when warned about and given the chance to avoid material *that reminds someone of their trauma*, do they actually avoid it?

Of course, ascertaining individual participants' unique trauma triggers and then matching study stimuli to those triggers would be challenging. The "trauma film paradigm," a well-established method to simulate exposure and reactions to psychological trauma, provides a practical alternative (Holmes, Brewin, & Hennessy, 2004; Lazarus, Speisman, Mordkoff, & Davison, 1962; Holmes & Bourne, 2008; James et al., 2016). Typically, non-clinical participants watch a short film depicting a traumatic event and then answer questions about the event and their emotional responses. This experience reliably induces PTSD-like symptoms—such as intrusive memories of the film footage—that are similar to, but occur to a lesser extent and shorter duration, as those for real trauma (Holmes & Bourne, 2008; James et al., 2016). This paradigm therefore offers a suitable method to study people's avoidance of warned of stimuli that *directly reminds* them of a trauma analogue event (i.e., stimuli taken from the film).

Here, participants watched a traumatic film and then viewed 32 still images taken from the film for 5(s) each. On half the trials, a trigger warning message preceded each image; on the remaining trials a neutral task instruction appeared before each image. We instructed participants to view each image for the entire time it was displayed, but also gave them the option of pressing a "stop viewing" button that would take them to a blank screen

for the equivalent time if they did not want to view the image. We therefore measured avoidance behaviors in two distinct ways: passive avoidance—operationalized as more time spent avoiding viewing the image by remaining on the instruction screen, and active avoidance—operationalized as the number of times participants chose to cover photos and the subsequent time spent viewing each image. We also examined if avoidance behaviors were related to emotional responses or individual difference factors as suggested by previous research (e.g., Gainsburg & Earl, 2018). To do so, we measured state anxiety, projected PTSD-like symptoms, perceptions of how others might be harmed by the content, trait anxiety, experiential avoidance (the tendency to avoid feelings associated with anxiety), beliefs that general trigger warnings are helpful/were helpful in the study, use of and perceptions that avoidance/approach strategies are beneficial, prior exposure to traumatic events, and also prior personal experience with the topic of the film.

We recruited an undergraduate university population. This sample is particularly appropriate to evaluate the reported efficacy of trigger warning messages for three reasons. First, the debate about the purported benefits and harms of trigger warnings is centralized around the use of trigger warnings on college campuses (e.g., Lukianoff & Haidt, 2015). Second, many universities now mandate the use of trigger warnings as part of mental health initiatives (e.g., Harris, 2016; Palmer, 2017). Third, at least one third of first-year students around the world screen positive to an anxiety, mood, or substance abuse disorder (DSM–IV; Auerbach et al., 2018).

We had two competing predictions regarding avoidance behaviors. On the one hand, if critics' and advocates' claims that trigger warnings lead to avoidance are true, then we would expect participants to engage in more avoidance behaviors—more time spent waiting on the message screen, more instances of covering an image and less time spent viewing an image before it is covered—on trigger warning message trials compared to neutral task

instruction trials. On the other hand, given other research has found that trigger warnings do not promote avoidance behaviors (Bruce & Roberts, 2020, Gainsburg & Earl, 2018), it is also possible that we may find little difference in avoidance behaviors between warning and control trials.

We made additional predictions about the emotional consequences of avoidance behaviors. If advocates' claims—that trigger warnings help people to *pause* and emotionally *prepare* (Strothman, 2021) to view distressing content—are true, then we should find that the time spent waiting on the message screen before viewing each image, particularly the trigger warning screen, is negatively associated with extent of negative emotional reactions throughout the study (i.e., state anxiety after the image task, projected future PTSD-like symptoms and harm to others). However, based on critics' claims, as well as other previous empirical work showing that trigger warnings have negligible effects on emotional responses (e.g., Bridgland et al., 2019) and reported coping strategies (Bridgland, Barnard & Takarangi, 2021), we may instead find that time spent on the message screens is not associated with reduced distress.

Finally, based on previous work showing that certain populations (e.g., people with increased anxiety; Gainsburg & Earl, 2018) may be more likely to avoid content marked with a trigger warning, we expected avoidance behaviors would be positively correlated with trait anxiety, experiential avoidance, reported use of and perceptions that avoidance strategies are beneficial overall and within the study (i.e., perceptions that being able to press a key to stop viewing the images was beneficial), and perceptions that trigger warnings are helpful overall/within the study. Further, avoidance behaviors should be negatively associated with participants' reported use of approach strategies and perceptions that approach strategies are beneficial overall and within the study (i.e., perceptions that viewing the images in the image

task was beneficial in reducing distress). Finally, we expected these correlations to be stronger when participants saw a trigger warning message versus control message.

Method

The Flinders University Social and Behavioural Research Ethics Committee approved this experiment. We preregistered this experiment (https://osf.io/dmeuq). The data and supplementary material for this experiment can be found here: https://osf.io/gcsf8/. We have reported all measures, conditions, and data exclusions.

Participants

According to an a priori power analysis for a two-tailed, matched pairs t-test (using G*Power; Faul, Erdfelder, Lang, & Buchner, 2007) with an alpha of .05, power of .80, and effect size of $d = 0.2$ (the smallest effect we would be interested in) a sample size of $N = 199$ is required.

We recruited 207 participants using the Flinders University participation system via SONA. We excluded four participants because of electronic data collection malfunction and four participants withdrew during the film viewing phase of the study. Our final sample of 199 participants ranged from 18-60 years ($M = 22.78$, $SD = 6.58$), and were predominately female (71.9%), and Caucasian/White/European (41.7%), followed by Asian (19.6%), Middle Eastern (3.0%), African (2.01%), 1.0% Indigenous Australian, and 1.0% Polynesian. Some participants specified nationality[1] (Australian 23.6%), nationality and ethnicity (3.5% European-Australian; 1.5% Asian-Australian; 0.5% Asian-European; 1.0% Middle Eastern-Australian; 0.5% South American-Australian) or provided no answer (0.5%).

Materials

[1] Participants were asked to report their ethnicity in an open text box resulting in the range of answers provided—i.e., some people interpreted "ethnicity" as "nationality", and some participants provided both nationality and ethnicity.

Film stimulus. Participants watched an 8-minute negative film widely employed within analogue trauma research (e.g., Green, Strange, Lindsay, & Takarangi, 2016; Takarangi, Strange, & Lindsay, 2014; Woodward, & Beck, 2017) from the 1988 fictional movie *The Accused*. This clip depicts the gang rape of a female in a bar.

Film ratings. After watching the film, participants indicated (a) how distressing they found the film, and (b) how closely they paid attention to it, and (c) how involved they felt on an 10-point Likert scale (1 = Not at all, 10 = Extremely).

Image stimuli. We generated an image still from every 7(s) of the film, resulting in 64 images. To determine the 32 most distressing images, pilot participants (N = 33) first watched the film, and then answered, "How distressed do you feel at the present moment?" (on a scale of 1 = *Very slightly or not at all*, 5 = *Extremely*) in response to each of the 64 images, presented in a random order (resulting 32 most distressing images: M = 4.55, SD = 0.24).

Image task. All participants saw all 32 images, but we manipulated message within-subjects: half of the image trials were preceded by a trigger warning message ("*Warning: The image you are about to view contains disturbing content that may be distressing*."), and half with a control message ("*Instructions: When you are ready, the next image will be displayed on the next screen*."; counterbalanced). We created four predetermined random sets of image and message screen pairings (control and warning). Participants saw one image set, presented on E-Prime (Psychology Software Tools, Pittsburgh, PA). Image and message pairings appeared in a random order with one exception: the first two trials were always one warned image and one control image (randomized) from four predetermined subsets of images (i.e., one from each larger set). This setup allowed us to examine participants' first two responses in isolation, in case participants habituated to the effect of the warning over extended trials, or the effect of the warning message on behavioral responses was temporally brief.

Participants could spend an unlimited time on each message screen and up to 5(s) on each image. We instructed participants that they should view each image for the entire time it was displayed, but advised them that they could use a "stop viewing" button at any time before the end of the 5(s) which would take them to a blank screen. Thus, participants had the option to stop viewing an image and to instead view a blank screen for the remainder of the viewing time. Participants could not move through the overall task faster by using the "stop viewing" option.

Short-form Spielberger State-Trait Anxiety Inventory (STAI-6; Marteau & Bekker, 1992). Participants rated how they felt at that current moment for three anxiety-present items (e.g., "*I am worried*") and three anxiety-absent items (e.g., "*I feel calm*";1 = *Not at all*, 4 = *Very much*; (present study α = .75-.83). Scores reflect state anxiety responses and are summed (range 4-20).

Anticipated traumatic stress symptoms. Participants completed a modified version of the Pretraumatic Stress Reactions Checklist (PreCL; Berntsen & Rubin, 2014). The original PreCL asks participants to rate traumatic stress symptoms that reflect the DSM-5 symptom criteria for PTSD in relation to possible *future events*. Here, we modified the measure to instead ask about possible *future symptoms* that participants anticipated over the next 24 hours (e.g., *Repeated, disturbing and unwanted images related to the film or images, Avoiding imaginings, thoughts or feelings related to the film or images*) as a result of doing the film/image task (0 = *Not at all,* to 4 = *Extremely*. One item from the original scale was excluded because it did not fit the present context (Blaming yourself or someone else for a possible future stressful experience or what has led up to it). Scores are summed (range 0-76). Despite modifications to the instructions, internal reliability in the present study was excellent (α = .92).

Spielberger Trait Anxiety Inventory (1983; STAI-form Y). Participants rated how often they generally experience a series of anxiety symptoms (e.g., anxiety present item: "*I worry too much over something that really doesn't matter*" and anxiety absent item: "*I am content; I am a steady person*.";1 = *Almost never*, to 4 = *Almost always*). The scale has excellent internal consistency (present study α = .93). Total scores range from 20 to 80, with higher scores indicating greater trait anxiety.

The Acceptance and Actions Questionnaire (AAQ; Hayes et al., 2004). Participants rated how well a series of statements reflecting experiential avoidance—a tendency to avoid particular private experiences (e.g., bodily sensations, emotions, thoughts etc.) typically associated with anxiety (e.g., If I could magically remove all of the painful experiences I've had in my life, I would do so)—generally applies to them (on a scale of 1 = *Never true*, to 7 = *Always true*; present study; α = .71). Scores are summed with higher scores indicating greater experiential avoidance (range 9-63).

Coping strategies following a traumatic event. Participants rated how often they used and the beneficial nature of: a) behavioral approach, b) emotional approach, c) behavioral avoidance, and d) emotional avoidance coping strategies, following a stressful/traumatic event (on a scale of 1 = *Never*, to 5 = *Often*).

Questions related to the image task and warning screens. Participants rated the perceived effectiveness of the image rating task: a) Did you find it beneficial to view images related to the film again? (e.g., do you think it made you feel less distressed over time?), b) Did you find it beneficial to be able to stop viewing the images with the "stop viewing" option? (e.g., did it make you feel less distressed?; 1 = *Not at all*, 5 = *Extremely*), c) I needed the warning messages about the images to prevent them from causing distress, d) People should always receive a warning message before viewing images like this (1 = *Strongly disagree*, 5 = *Strongly agree*).

Topic of the film relevance. Participants indicated a) if they had any personal experience with the topic of the film/images (Yes/No), b) how much anxiety the average person would feel viewing the film and images, and c) how much anxiety someone who had a personal experience with the content would feel viewing the film and images (0 = *None*, to 5 = *An extreme amount*).

Single item Criterion A question. We asked participants to think of their most traumatic or stressful event, and to indicate (Y/N), if they were exposed to: death, threatened death, actual or threatened serious injury, or actual or threatened sexual violence, in any of the following way(s): a) Direct exposure, b) Witnessing the trauma c) Learning that a relative or close friend was exposed to a trauma d) Indirect exposure to aversive details of the trauma, usually in the course of professional duties (e.g., first responders, medics; i.e., Criteria A for PTSD in the DSM-5).

Beliefs about trigger warnings as protective (versus coddling; Gainsburg & Earl, 2018). Participants rated their agreement with two statements, "Trigger warnings that precede distressing content 'coddle' people, hurting them in the long run," and "Trigger warnings that precede distressing content 'protect' people, helping them in the long run" on 7-point scales (1 = *Strongly disagree*, to 7 = *Strongly agree*). A two-item composite score was created by reverse scoring the "coddling" item and averaging the two items. Additionally, participants were also coded as believing that trigger warnings are coddling (those one standard deviation below the mean), of average protectiveness (falling between one standard deviation below and above the mean) or more protective (those one standard deviation above the mean; Gainsburg & Earl, 2018).

Procedure

After completing informed consent procedures participants completed demographic questions and a measure of baseline state anxiety (STAI-6). Next, participants watched the

film, and completed a second measure of state anxiety, in addition to the film rating questions. Participants then completed the image task. To enhance our cover story and obscure our hypothesis, we falsely told participants these images would relate to a memory test later in the experiment. Participants then completed a third measure of state anxiety, possible future post-traumatic stress symptoms related to the film (Pre-CL), trait anxiety (STAI-Form Y), experiential avoidance (AAQ), questions about coping strategies, questions about the warning messages/image task, personal experiences with the topic of the film, the single item Criterion A trauma question, and the question about their belief in trigger warnings as protective or coddling. We also asked participants if they had seen the film before.[2] Participants were then fully debriefed and granted course credit (n = 170) or paid $10AUD ($n$ = 29) for their time.

Results

Statistical overview

We ran analyses using Null-Hypothesis Significance Tests (α = .05) in SPSS Version 25 and JASP for MacOS version 0.13.1. Since each of our analyses relates to a preregistered hypothesis, we have retained original p-values and have not corrected for multiple comparisons (Rubin, 2021).

Effectiveness of the analogue trauma task

Before turning to our main analyses, we examined how effective our analogue trauma film and image task was in inducing negative affect, as well as task compliance. Participants rated the film as distressing (M = 8.96, SD = 1.28), paid close attention (M = 8.83, SD = 1.25), and felt highly involved (M = 7.56, SD = 2.14). To examine state anxiety throughout the study, we ran a repeated measures ANOVA on state anxiety scores at baseline, post-film, and post-image task. The overall main effect of time was significant (F (1.49, 294.18) =

[2] See https://osf.io/tgk7n/ for descriptive statistics.

858.18, p <.001, partial eta = .81).[3] State anxiety scores increased significantly from baseline (pre-film; M = 10.57, SD = 2.84), to post-film (M = 19.84, SD = 3.37, p <.001), and decreased from post-film to post-image task (M = 18.67, SD = 3.74, p <.001).

Avoidance behavior

Reaction time data exclusions

As per our pre-registration, we excluded reaction time data if it fell below 200 ms or over 3SD above the mean—based on each participant's average reaction time for that task. However, likely because the first two trials were the first time participants viewed the warning and control screen messages (randomized as either the first or second trial participants saw), participants' first two trial responses (M = 4235.95, SD = 1288.14) were on average significantly slower than the remaining trial responses (M = 1666.08, SD = 985.69, paired samples t-test = t (198) = 27.33, p <.001). Likewise, for participants who chose to avoid images with the black screen during viewing time, the average time taken to cover the image was slower when it occurred during the first two trials (M = 3277.28, SD = 1068.04) than for the remaining trials (M = 2214.80, SD = 850.91). Only six participants chose to cover an image during the first two trials (M = 3097.00, SD = 1046.82) *and* in the remaining trials (M = 1770.13, SD = 1049.16, paired samples t-test = t (5) = 3.37, p = .020). Thus, for the first two trials, we only excluded responses < 200 ms. In total, we excluded 16 message screen responses < 200 ms, and 89 message screen and three cover time responses for falling higher than 3SD above the mean for that participant.[4]

We now turn to our main research aim—were participants more likely to avoid stimuli associated with an analogue trauma film if the content was preceded by a trigger warning message versus a control screen? Recall that avoidance behaviors were

[3] Mauchly's Test of Sphericity was violated (<.001), therefore a Greenhouse-Geisser (.743) correction was applied (Field, 2005).
[4] Only 10 participants had more than one data point excluded. See https://osf.io/tgk7n/ for a detailed breakdown.

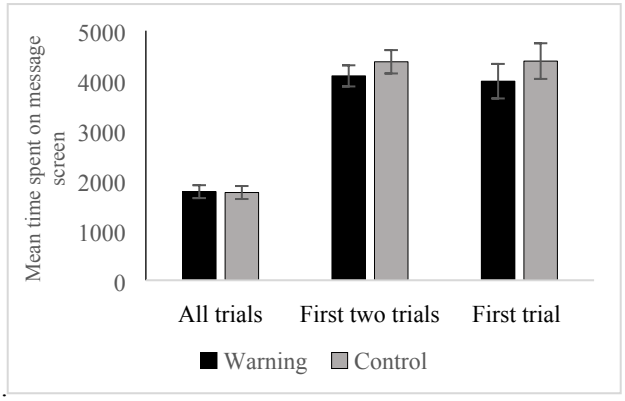
operationalized as a) time spent avoiding viewing an upcoming image by remaining on the instruction screen, b) number of images in each condition that participants covered, c) time spent viewing images. We now examine each of these avoidance behaviors in turn.

Time spent avoiding viewing the image by remaining on the instruction screen

Recall that trigger warning advocates claim that warnings help people to pause and prepare to cope with content. For this claim to be true, we expected that participants would spend *more* time on the warning screens compared to control screens. Contrary to these claims, across all 32 trials there was no difference in the average time participants spent on the instruction screen when it was a trigger warning compared to when it was a control instruction (paired samples: t (198) = 0.67, p = .502, d = 0.05, 95% CI [-0.09, 0.19]; Figure 1). Because participants may have come to expect— over the course of the image task—that similar types of images were covered by warning and control screens, we also examined the first two trials only (which were always a warning and a control screen; randomized). Participants tended to spend *less* time on the trigger warning, versus the control screen (paired samples: t (198) = -2.11, p = .036, d = -0.15, [-0.29, -0.01]) within the first two trials. A similar pattern emerged when we compared participants who viewed a warning (n = 100) with those who viewed a neutral task instruction (n = 99) on the first trial, between-subjects; but this difference was not statistically significant (t (197) = -1.61, p = .110, d = -0.23, [-0.51, 0.05]).

Figure 1

Mean time (ms) spent on instruction screens (with 95% Confidence Intervals) by message screen type



Number of images that participants covered and time spent viewing images before they were covered

Recall that warning advocates and critics both claim that warnings encourage people to avoid potentially distressing stimuli. For this claim to be true, we expected that participants would be more likely to cover, and thus would spend less time looking at, images preceded by warning screens compared to the control screens. However, overall, participants rarely avoided the images by using the cover option: only 239 images were covered out of a total of 6,368 trials (3.75%), only 25 (12.56%) participants used the cover feature at all; of these, five participants covered only one image and 20 participants covered more than one. Therefore, our resulting sample sizes for the following analyses should be interpreted with extreme caution as they are drastically underpowered.

Among participants who covered both warning and control screens (n = 20), covering rates were similar for images preceded by a warning screen (M = 5.75, SD = 4.51) and images preceded by a control screen (M = 5.95, SD = 4.51, paired samples t-test: t (19) = 0.61, d = 0.14, 95% CI [-0.31, 0.57]). The average time spent viewing images preceded by a trigger warning (M = 2225.32, SD = 959.21) was also similar to images preceded by a control (M = 2116.70, SD = 945.70; paired samples: t (19) = -0.68, p = .505, d = -0.15, [-0.59, 0.29]). Participants covered twice as many images if they were preceded by a trigger warning message (3.01% of total participants, n = 6) within the first two trials, versus a neutral task

instruction (1.51% of participants, n = 3; $\chi^2 = 42.20$, $p < .001$, $\Phi = .461$. Only two participants used the cover feature for *both* of the first two image trials (trigger warning: $M = 3949.50$, $SD = 122.32$, control: $M = 2766.50$, $SD = 2130.51$; paired samples: $t(1) = -0.74$, $p = .593$, $d = -0.52$ [-1.96,1.07]). Only four participants who viewed a warning (average time before covering images: $M = 3581.50$, $SD = 1029.72$) on the first trial covered the image, while no participants who viewed a control screen on the first trial covered the image. Therefore, we could not run an independent samples t-test for these data. However, a one sample binomial test revealed that the degree to which participants avoided images (by choosing to cover the image), when a warning occurred in the first trial (estimate = 0.04, in the form of choosing to cover the image), was significantly greater than .001 ($p < .001$; n = 100; Clopper-Pearson 95% CI [.01, .09]).

Taken together then, contrary to the claims of advocates and critics, participants were extraordinarily reluctant to avoid viewing negative study stimuli when given the option to do so throughout the image viewing task, and warnings did not enhance avoidance behavior—apart from a minor increase in avoidance when a warning appeared in the first few trials.

Were avoidance behaviors associated with emotional responses and individual differences?

Recall our secondary aim: to assess if avoidance behaviors were related to emotional responses and to individual difference factors. We correlated our avoidance behaviors (time spent on message screens, number of images participants covered and time spent on images) with state anxiety, film ratings, trait anxiety, experiential avoidance, reported use of and perceived benefits of coping strategy types (approach vs. avoidance), anticipated symptoms, how others' would perceive the film/image content, perceptions that trigger warnings are protective (vs. coddling), and the perceived effectiveness of the image task (see Tables 1 and 2 and also https://osf.io/tgk7n/ for analyses pertaining to perceptions that trigger warnings are

protective vs. coddling). As per our preregistration, we also ran exploratory analyses to investigate whether splitting the data on various individual difference characteristics (i.e., having prior experience with the topic vs. no prior experience; belief that trigger warnings are coddling vs. of average protectiveness or more protective, or having experienced vs. not experienced a Criterion A event), revealed different relationships with our main avoidance variables (see Tables 1 and 2 and also https://osf.io/tgk7n/ for all analyses pertaining to having experienced vs. not experienced a Criterion A event).

Time spent avoiding viewing the image by remaining on the instruction screen and emotional responses and individual difference factors

The belief that participants needed warning messages about the images to prevent those images from causing distress was positively correlated with overall average time spent on message screens (n = 199, r =.18, p = .013), time spent on warning message screens (n = 199, r = .16, p = .028), and time spent on control message screens (n = 199, r = .19, p = .008). Therefore, it is possible that general beliefs about the usefulness of warning signals (i.e., as ways to prepare for upcoming content) extended to *both* screen types and led these participants to linger longer on the screen preceding each image as a way to emotionally prepare/brace themselves to view those images. However, the time participants spent on message screens—both overall and by screen type—was *not* associated with any reductions in reported distress (i.e., state anxiety, rs = -.01-.02, post image task, projected future symptoms, rs = .01, or perception of possible harm to others rs = -.01-.10, all ps > .05). Therefore, it seems unlikely that time spent waiting on each message screen—regardless of whether this screen contained a warning—helped participants to emotionally prepare to reduce the negative impact of each image.

Finally, generally using approach strategies when personally experiencing a stressful/negative event was positively associated with the average time spent across both

message screen types (n = 199, r = .14, p = .046) as well as the average time spent on the control message screen (n = 199, r = .15, p = .040), and the average time spent on the warning message screen (n = 199, r = .13, p = .065) analyzed individually, though the relationship with warning screen was not statistically significant. Perhaps participants who are generally more likely to try and "*engage with thoughts and feelings*"[5] lingered longer on the screens to try and think about/reappraise the upcoming content in a different way. However, the effect sizes of these relationships are small, with p-values close to .05 and thus should be interpreted with caution.

Advocates claim that trigger warnings help trauma survivors prepare to cope with upcoming content. For this claim to be true we expected that trauma survivors would spend *more* time on the warning screens compared to people who had not experienced a trauma. However, there was no difference for time spent on the trigger warning screen (experienced a Criterion A event = Yes: n = 151, M = 1830.95, SD = 950.91, No: n = 48, M = 1600.39, SD = 827.97, Welch's t (89.73) = 1.62, p = .109, d = 0.26, 95%CI [-0.07,0.58]). Curiously, the average time spent waiting on the control screen was higher for participants who had experienced a Criterion A event (n = 151, M = 1821.75, SD = 983.87) vs. who had not (n = 48, M = 1556.81, SD = 708.97, Welch's t (109.32) = 2.04, p = .044, d = 0.31, [-0.02, 0.64]).

Number of images that participants covered, time spent viewing images before they were covered, and emotional responses and individual difference factors

Because only 25 participants used the cover image option, the correlations between the number of images participants covered, time spent viewing images before they were covered, and emotional responses and individual difference factors are drastically

[5] e.g., from the questions participants were asked: *"finding personal meaning in the event"*, *"thinking of different ways to deal with the outcomes of the event"*,*"trying to see the good side of the situation"* etc.

underpowered. However, for completeness we report them in Tables 1 and 2 and describe them in detail here: https://osf.io/tgk7n/

Discussion

Advocates (e.g., Cripps, 2020) and critics (e.g., Lukianoff & Haidt, 2015) of trigger warnings both claim that trigger warnings promote avoidance behaviors. However, we found that apart from a minor increase in avoidance when a warning appeared in the first few trials, undergraduates did not passively (by remaining on an instruction screen) or actively (by covering images) avoid negative stimuli prefaced with a trigger warning any more than stimuli without a warning. In fact, participants were reluctant overall to avoid distressing images; only 12.56% (n = 25) participants used the option to cover such images when given the opportunity to do so in the image viewing task. Put differently, when warned about and given the chance to avoid material that reminds them of a negative experience (here operationalized as exposure to a trauma film), participants are extraordinarily reluctant to do so. Furthermore, we did not find any relationships between avoidance behaviors and emotional responses throughout the study (e.g., state anxiety) or with individual difference characteristics (e.g., trait anxiety)—including for participants with prior experience of the topic of the trauma film and/or exposure to a Criterion A trauma.

Our finding that trigger warnings do not seem to *enhance* avoidance behaviors towards distressing stimuli fits with emerging research (Bruce & Roberts, 2020; Bridgland, Barnard, & Takarangi, 2021; Bridgland, Bellet, & Takarangi, 2021; Gainsburg & Earl, 2018; Kimble et al., 2021). Our results build on prior research, by showing this pattern for content directly related to a negative experience. Previous trigger warning research has examined what happens when someone views novel distressing stimuli, with no way of knowing if this material reminds the person of their past experience. Here we addressed this limitation and found that participants *did not* avoid more stimuli related to an analogue trauma event when

they saw a trigger warning message versus a control message. In fact, despite reporting that they found the film highly distressing, participants *rarely* avoided images at all in the image viewing phase—only *one* participant used the "cover photo" feature for all 32 images. Our results may also indicate that trigger warnings are linked to avoidance behaviors, but only in a minority of people already prone to avoidance. That is, it seems likely that people who use avoidance as a way to cope with distressing material may be as likely to do so when seeing content presented with a warning as when seeing content without a warning.

However, there are at least two alternative possibilities for the low rate of avoidance behaviors we observed. First, recall that our cover story was that participants were completing a juror decision making task and that we falsely told them that they would complete a memory test for the images. It is possible that the desire to perform well on a memory test drove participants to want to view the images for the full time those images were displayed on screen. However, providing participants with a reason to view the images mimics people's motivation for consuming negative or potentially triggering media in a real-world setting (e.g., lecture material required for a course, entertainment via watching a television show). Therefore, avoidance behaviors across the study may have been low because of a motivation to view the images, but this is likely reflective of real-world conditions where a desire to avoid distressing content competes with other reasons to consume it. Of course, this cover story was also integral to justify the task to participants and prevent participants from responding differently to warning versus control trials based on what they may have guessed about the expectations of the experimenter or study aims. Second, recall that state anxiety decreased significantly from before to after the image viewing task. It is possible that the image viewing task was not as distressing as viewing the film itself and therefore participants felt that they did not need to avoid the images. Along the same lines, since participants were already in a negative mood state (due to watching the

film) it is possible that they did not perceive any further emotional harm could come to them by viewing film image stills. This possibility is supported anecdotally by participant comments during debriefing and carries some worrying implications. Trigger warnings are typically for "vulnerable" people such as people with poor mental health (e.g., low-wellbeing, depression etc.; Mosseri, 2019) who are likely in negative mood states. These people may therefore be *even less* likely (versus non-vulnerable people) to use trigger warnings as a tool for avoidance if they believe that what they view is unlikely to have any further dampening effect on their mood. Future research could address the possibility by allowing sufficient time for mood to return to normal before participants complete the photo viewing task or by manipulating mood to examine if it affects avoidance rates.

Aside from complete avoidance, trigger warning advocates also claim that warnings help people *pause* and emotionally *prepare* to cope with material. However, participants did not spend longer on the trigger warning screens versus the control screens. In fact, within the first two trials (which were always a warning and control screen) we found that participants spent more time waiting on the *control* screen. One potential explanation for this finding is our warning screens created a "Forbidden Fruit effect" (Ringold, 2002), encouraging participants' morbid curiosity about distressing content (Oosterwijk, 2017), and therefore leading them to engage with the covered content more quickly.

However, we did find that participants' belief that warning messages about the images would prevent them from causing distress was positively associated with spending more time, on average, on the message screens (both overall and by screen type). Thus, it is possible that general beliefs about the usefulness of warning messages (i.e., as ways to prepare for upcoming content) extended to *both* screen types and led these participants to linger longer on each screen as a way to emotionally prepare/brace themselves to view the next image. Yet there was no evidence that time spent on the message screens was associated with reduced

distress throughout the study—which is what we would have expected if participants were spending their time "pausing" on message screens to emotionally prepare themselves to view content. Therefore, even if participants prepare in a physical *pausing* sense, that preparation appears unhelpful.

Overall, one implication of these results—consistent with prior work on the effects of trigger warnings on emotional reactions—could be that trigger warnings are inert or trivial (Sanson et al., 2019). That is, trigger warnings do not do any overt good, but they also do not seem to lead to any obvious harm either. Therefore, if people use warnings because they believe warnings are helpful, they will not experience increased harm as a result. Alternatively, trigger warnings *could* lead to harm if they are the only mental health safeguard or support policy employed in various domains. Indeed, there is a continued push to promote the use of trigger warnings despite mounting evidence that they are ineffective in reducing emotional reactions or promoting avoidance behaviors. For instance, Strothman (2021) states that even though research suggests warnings are not helpful, she *"feel[s] strongly that we need to keep using them*." This continued belief in the illusory of trigger warnings could result in two potential harms. First, for individuals trigger warnings may become a "box-ticking exercise" (Hay, 2019) or a "sticker-fix" (Fagan, 2019). That is, some people may think that adding a trigger warning to their content—whether that be a university lecture or a social media post—might absolve them of making any other efforts to present distressing material in a conscientious way. On a more macro level, the continued beliefs about the benefits of trigger warnings could result in reduced efforts by policy makers or institutions to find efficacious mental health support strategies, because trigger warnings may be considered one such approach already in use.

Our research has several limitations. First, of course the trauma analogue paradigm differs from autobiographical trauma. Yet research examining how participants respond to

stimuli that match an autobiographical trauma has also found trigger warnings do not prompt avoidance behaviors (i.e., Bruce & Roberts, 2020; Bridgland, Barnard, & Takarangi, 2021) or ameliorate negative emotional reactions (i.e., Jones, Bellet, & McNally, 2019) in response to the stimuli. Moreover, it is important to triangulate findings (i.e., use multimethod approaches) to reduce the possibility that a set of results can be attributed to specific methodological approaches or data sources (Lin, Werner, & Inzlicht, 2021). That is, our results only contribute one piece to the puzzle; future research is necessary to expand our knowledge on the ways that trigger warnings may or may not affect avoidance behaviors and thus allow us to reach convergent and valid conclusions. For instance, one future direction could be to include a 'no video' or 'neutral video' control condition. Such a design would test whether experiencing a trauma changed how people respond to trigger warnings about that type of trauma (i.e., video condition), versus people who had not experienced that event (i.e., no/neutral video condition).

Second, because the rates of avoidance in our study were extraordinarily low, our analyses for some avoidance variables (i.e., rates of covering photos and also time spent on photos before they were covered) were underpowered. Therefore, we were not able to comprehensively test some of our hypotheses related to these avoidance variables. However, if the overall rate of avoidance is so low, it seems unlikely that avoidance is a widespread behavior in the general population. Perhaps we should be more concerned with finding out why people seem so eager to approach negative content and the implications of this behavior, rather than on the minority of people who seem to want to avoid this material.

Third, we used a university undergraduate sample—meaning that our results may not generalize to other populations that vary in sociodemographic variables (e.g., age, education, socioeconomic status etc.). Indeed, because trigger warnings are primarily intended for people with clinical levels of symptoms, it is possible that avoidance behaviors might change

if we specifically recruited specific clinical populations (e.g., people with a clinical diagnosis of PTSD).

Taken together, contrary to the claims of advocates and the concerns of critics, we did not find evidence that trigger warnings significantly enhance avoidance behaviors in undergraduates—apart from a *minor* increase when a warning appeared in the first few trials. Furthermore, we did not find any indication that trigger warning messages help people to pause and emotionally prepare themselves to view negative content. Future work should focus on developing strategies that fulfil the purported aims of trigger warning messages.

References

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cui- jpers, P., . . . Kessler, R. C. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. Journal of Abnormal Psychology, 127, 623–638. http://dx.doi.org/10.1037/abn0000362

Badour, C. L., Blonigen, D. M., Boden, M. T., Feldner, M. T., & Bonn-Miller, M. O. (2012). A longitudinal test of the bi-directional relations between avoidance coping and PTSD severity during and after PTSD treatment. *Behaviour Research and Therapy*, *50*(10), 610–616. https://doi.org/10.1016/j.brat.2012.06.006

Bellet, B.W., Jones, P.W.., & 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

Berntsen, D., & Rubin, D. C. (2014). Pretraumatic Stress Reactions in Soldiers Deployed to Afghanistan. *Clinical Psychological Science*, *3*(5), 663–674. https://doi.org/10.1177/2167702614551766

Boysen, G. A., Isaacs, R. A., Tretter, L., & Markowski, S. (2021). Trigger warning efficacy: The impact of warnings on affect, attitudes, and learning. *Scholarship of Teaching and Learning in Psychology*. https://doi.org/10.1037/stl0000150

Bridgland, V. M. E., Barnard, & Takarangi, M. K. T. (2021). Unprepared: Thinking of a trigger warning does not prompt preparation for trauma-related content. Under review; see https://osf.io/7n85z/ for information.

Bridgland, V. M. E., Bellet, W. B., & Takarangi, K.T. M. (2021). Curiosity disturbed the cat: Instagram's Sensitive Content Screens do not deter vulnerable users from viewing distressing content. Under review. See https://osf.io/rj987/ for information.

Bridgland, V. M. E., Green, D. M., Oulton, J. M., & Takarangi, M. K. T. (2019). Expecting the worst: Investigating the effects of trigger warnings on reactions to ambiguously themed photos. *Journal of Experimental Psychology: Applied*, *25*(4), 602–617. https://doi.org/10.1037/xap0000215

Bruce, M. & Roberts D. (2020). Trigger warnings for abuse impact reading comprehension in students with histories of abuse. *College Student Journal*, *54*(2): 157-168.

Cripps, C. (2020, November 16). The Crown was right to warn viewers of Princess Diana's bulimia. *The Independent*. Retrieved from https://www.independent.co.uk/arts-entertainment/netflix/the-crown/princess-diana-bulimia-the-crown-b1721340.html

Ehlers, A., Hackmann, A., & Michael, T. (2004). Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory*, *12*(4), 403–415. https://doi.org/10.1080/09658210444000025

Fagan, A. (2019, April 17). Do Trigger Warnings Actually Work?. *Psychology Today*. Retrieved from https://www.psychologytoday.com/au/blog/brainstorm/201904/do-trigger-warnings- actually-work

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. http://doi.org/10.3758/BF03193146

Field, A. (2005). *Discovering statistics using SPSS*. London: SAGE.

Flaherty, C. (2019, March 21). New study says trigger warnings are useless. *Inside Higher Education*. Retrieved from https://www.insidehighered.com/news/2019/03/21/new-study-says-trigger-warnings-are-useless-does-mean-they-should-be-abandoned

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. http://doi.org/10.1016/j.jesp.2018.08.006

George, E. & Hovey, A. (2020). Deciphering the trigger warning debate: a qualitative analysis of online comments, *Teaching in Higher Education*, *25*(7), 825–841, https://doi.org/10.1080/13562517.2019.1603142

Green, D. M., Strange, D., Lindsay, D. S., & Takarangi, M. K. T. (2016). Trauma-related versus positive involuntary thoughts with and without meta-awareness. *Consciousness and Cognition*, *46*, 163–172. https://doi.org/10.1016/j.concog.2016.09.019

Hay, M. (2019, May 10). Do Trigger Warnings Actually Work?. *Vice.* Retrieved from https://www.vice.com/en/article/wj9ba4/do-trigger-warnings-actually-work

Hayes, S. C., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., … McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *The Psychological Record*, *54*(4), 553–578. https://doi.org/10.1007/bf03395492

Harris, S. (2016, August 31). Think trigger warnings are never mandatory on campus? Think again. *FIRE*. Retrieved from https://www.thefire.org/ think-trigger-warnings-are-never-mandatory-on-campus-think-again/

Holmes, E. A., & Bourne, C. (2008). Inducing and modulating intrusive emotional memories: A review of the trauma film paradigm. *Acta Psychologica*, *127*(3), 553–566. https://doi.org/10.1016/j.actpsy.2007.11.002

Holmes, E. A., Brewin, C. R., & Hennessy, R. G. (2004). Trauma Films, Information Processing, and Intrusive Memory Development. *Journal of Experimental Psychology: General*, *133*(1), 3–22. https://doi.org/10.1037/0096-3445.133.1.3

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. https://doi.org/10.1016/j.cpr.2016.04.010

Jean-Philippe, M. (2020, December 29). Bridgerton's Controversial Sex Scene Needs a Trigger Warning. *Oprah Daily*. Retrieved from https://www.oprahdaily.com/entertainment/tv-movies/a35090027/bridgertons-controversial-sex-scene-episode-6/

Jones, P. J., Bellet, B. W., & McNally, R. J. (2019). Helping or harming? The effect of trigger warnings on individuals with trauma histories. *Clinical Psychological Science*, *8*(5), 905-917. https://doi.org/10.31219/osf.io/axn6z

Kimble, M., Flack, W., Koide, J., Bennion, K., Brenneman, M., & Meyersburg, C. (2021). Student reactions to traumatic material in literature: Implications for trigger warnings. *PLOS ONE*, *16*(3), e0247579. https://doi.org/10.1371/journal.pone.0247579

Krypotos, A. M., Effting, M., Kindt, M., & Beckers, T. (2015). Avoidance learning: a review of theoretical models and recent developments. *Frontiers in Behavioral Neuroscience*, *9*, 189. http://doi.org/10.3389/fnbeh.2015.00189

Levy, H. C., & Radomsky, A. S. (2013). Safety Behaviour Enhances the Acceptability of Exposure. *Cognitive Behaviour Therapy*, *43*(1), 83–92. http://doi.org/10.1080/16506073.2013.819376

Lin, H., Werner, K. M., & Inzlicht, M. (2020). Promises and perils of experimentation: The mutual internal validity problem. *Perspectives on Psychological Science, 16*(4), 854–863. http://doi.org/10.31234/osf.io/hwubj

Logan, E. (2020, December 26). This Controversial 'Bridgerton' Scene Is Causing a Lot of Frustration on Twitter. *Glamour*. Retrieved from https://www.glamour.com/story/controversial-bridgerton-scene-sexual-assault-twitter-response

Lukianoff, G & Haidt, J. (2015, September 1). How Trigger Warnings Are Hurting Mental Health on Campus. *The Atlantic*. Retrieved from https://www.theatlantic.com/magazine/archive/2015/09/the-coddling-of-the-american-mind/399356/

Marteau, T. M. & Bekker, H. (1992) The development of a six-item short-form of the State Scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*, 301-306. http://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Medhora, S. (2021, Jan 11). Does Netflix have a responsibility to provide trigger warnings on its content?. *Hack.* Retrieved from https://www.abc.net.au/triplej/programs/hack/netflix-trigger-warning-content-sexual-assault/13048904

Millgram, Y., Joormann, J., Huppert, J. D., & Tamir, M. (2015). Sad as a Matter of Choice? Emotion-Regulation Goals in Depression. *Psychological Science*, *26*(8), 1216–1228. https://doi.org/10.1177/0956797615583295

Mosseri, A. (2019, October 27). Taking More Steps To Keep The People Who Use Instagram Safe. *Instagram Blog*. Retrieved from https://about.instagram.com/blog/announcements/more-steps-to-keep-instagram-users-safe

Oosterwijk, S. (2017). Choosing the negative: A behavioral demonstration of morbid curiosity. *PLOS ONE*, *12*(7), e0178399. https://doi.org/10.1371/journal.pone.0178399

Palmer, T. (2017, March 29). Monash University trigger warning policy fires up free speech debate. *ABC News*. Retrieved from https://www. abc.net.au/news/2017-03-28/monash-university-adopts-trigger-warning- policy/8390264

Psychology Software Tools, Inc. [E-Prime Go]. (2020). Retrieved from https://support.pstnet.com/.

Rachman, S., Radomsky, A. S., & Shafran, R. (2008). Safety behaviour: A reconsideration. *Behaviour Research and Therapy*, *46*(2), 163–173. http://doi.org/10.1016/j.brat.2007.11.008

Ringold, D. J. (2002). Boomerang effects in response to public health interventions: Some unintended consequences in the alcoholic beverage market. *Journal of Consumer Policy*, *25*(1), 27–63. https://doi.org/10.1023/a:1014588126336

Rubin, M. (2021). When to adjust alpha during multiple testing: A consideration of disjunction, conjunction, and individual testing. *Synthese*, 1-32. https://doi.org/10.22541/au.162569018.85105015/v1

Sanson, M., Strange, D., & Garry, M. (2019). Trigger warnings are trivially helpful at reducing negative affect, intrusive thoughts, and avoidance. *Clinical Psychological Science, 7*(4), 778–793. https://doi.org/10.1177/2167702619827018

Simister, E. T., Bridgland, V. M. E., Takarangi M. K. T. (2021). An investigation of sensitive screens: Understanding when and why people approach sensitive content. See: https://osf.io/2fdr7

Speisman, J. C., Lazarus, R. S., Davison, L., & Mordkoff, A. M. (1964). Experimental analysis of a film used as a threatening stimulus. *Journal of Consulting Psychology*, *28*(1), 23–33. https://doi.org/10.1037/h0047028

Spielberger, C. D. (1983). State-Trait Anxiety Inventory for Adults. *PsycTESTS Dataset*. https://doi.org/10.1037/t06496-000

Strothman, L. (2021). Trigger Warnings: How to protect yourself from online triggers. *Dr Lisa Strohman*. Retrieved from https://drlisastrohman.com/how-to-protect-yourself-from-online-triggers/

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

Woodward, M. J., & Beck, J. G. (2017). Using the trauma film paradigm to explore interpersonal processes after trauma exposure. *Psychological Trauma: Theory, Research, Practice, and Policy*, *9*(4), 445–452. https://doi.org/10.1037/tra0000169

Highlights:

- Avoidance is one of the purported benefits and harms of trigger warnings.
- Here we aimed to directly measure avoidance following trigger warnings.
- Participants watched a traumatic film and viewed film images.
- Participants had the option to "cover" and avoid each image.
- Participants did not avoid more images preceded by trigger warnings (vs. control).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Table 1

Welch's t-test and descriptive statistics for participants who said they have personal experience with the topic of the film on key avoidance variables

Avoidance type	*t*	*df*	*p*	Cohen's *d*	95% CI for Cohen's *d* Lower	95% CI for Cohen's *d* Upper	Personal experience	*n*	*M*	*SD*
Average time spent on message screens across all trials	0.69	50.40	.496	0.13	-0.22	0.48	Yes	40	1871.74	1135.86
							No	159	1740.14	845.88
Average time spent on control screens	0.91	47.38	.365	0.18	-0.17	0.53	Yes	40	1913.06	1278.27
							No	159	1718.79	820.81
Average time spent on warning screens	0.39	55.50	.697	0.07	-0.28	0.42	Yes	40	1830.41	1017.73
							No	159	1761.48	904.43
Average time spent before covering images across all trials	-1.98	19.64	.062	-0.79	-1.66	0.09	Yes	8	1835.05	677.84
							No	17	2509.63	998.54
Average time spent before covering control images	-1.53	17.82	.144	-0.64	-1.52	0.25	Yes	8	1822.11	761.39
							No	15	2385.34	976.95
Average time spent before covering warning images	-2.63	19.95	.016	-1.04	-1.98	-0.08	Yes	7	1685.98	469.36
							No	15	2573.70	1112.35
Total times images covered across trials	0.61	13.57	.552	0.26	-0.59	1.10	Yes	8	11.25	9.56
							No	17	8.76	9.38
Total times images control images covered	0.52	14.05	.608	0.23	-0.64	1.09	Yes	8	6.00	4.69
							No	15	4.93	4.56
Total times images warning images covered	0.45	11.84	.664	0.20	-0.70	1.10	Yes	7	6.00	4.90
							No	15	5.00	4.91

Table 2

Correlations between distress variables and individual differences, and key avoidance variables

	n	Baseline STAI	Post Film STAI	Post Photo Task STAI	Film Distressing	PCL Modified	AAQ	Trait Anxiety	TW Protect	Photo task beneficial	Cover option beneficial	Warnings needed for images like these	Should always receive warnings for similar content	Anxiety average other person	Anxiety other personal experience	Avoid General average	Approach general average	Avoid Beneficial	Approach beneficial
Average time spent on message screens across all trials	199	-.04	-.01	-.02	-.06	.01	<.001	-.01	.08	.06	.06	.18*	.02	-.06	-.07	-.01	.14*	.01	.03
Average time spent on warning screens	199	-.03	.01	<.01	-.04	.01	<.01	.01	.09	.05	.05	.16*	<.01	-.01	-.07	-.02	.13	.02	.04
Average time spent on control screens	199	-.04	-.02	-.03	-.07	.01	<.001	-.02	.06	.06	.06	.19*	.05	-.10	-.06	<.01	.15*	<.01	.01
Total times images covered across trials	25	.25	.12	.21	.18	.31	.11	-.02	<.01	-.31	.19	-.10	.14	.05	.24	-.03	.04	.23	-.33
Total times images warning images covered	22	.31	.03	.19	.17	.24	.01	-.06	-.04	-.21	-.06	-.21	-.03	-.05	.15	-.25	.15	.20	-.29
Total times images control images covered	23	.16	.11	.20	.22	.40	.10	<.001	.04	-.27	.28	-.03	.25	.07	.30	.07	-.07	.21	-.30
Average time spent before covering images across all trials	25	-.33	.08	-.06	-.05	-.33	-.17	-.03	-.03	.47*	-.25	.08	-.02	-.08	<.01	-.06	-.17	-.27	.04
Average time spent before covering warning images	22	-.35	.19	-.08	-.05	-.27	-.14	-.03	-.39	.62**	-.31	.14	-.13	-.12	<.01	-.06	-.10	.34	.97
Average time spent before covering control images	23	-.33	-.16	-.15	.13	-.49*	-.12	-.19	-.23	.34	-.34	-.06	.03	-.11	.08	-.12	-.08	-.23	-.08

Note * correlation significant at the 0.05 level (2-tailed) ** correlation significant at the 0.01 level (2-tailed)