AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

https://doi.org/10.1037/xap0000389

Exposure to Descriptions of Traumatic Events Narrows One's Concept of Trauma

Payton J. Jones[1], David E. Levari[2], Benjamin W. Bellet[1], and Richard J. McNally[1]
[1] Department of Psychology, Harvard University
[2] Harvard Business School, Harvard University

The concept of "trauma" was originally used by psychiatrists to describe horrific events such as rape and torture that characteristically provoke extreme emotional distress. Both colloquially and clinically, the concept of psychological trauma has broadened considerably. Although many clinical scientists have expressed concern about the broadening of the concept of trauma, it remains unclear how this concept expansion occurs. We present two experiments in which American adults ($N = 276$ and $N = 267$) sequentially classified descriptions of events (e.g., "broke a leg in a bicycle accident") as either "trauma" or "not trauma." In the first experiment, we manipulated the frequency of severe events (i.e., severe events became less and less common). In the second experiment, we manipulated the range of events (i.e., participants viewed only severe or only nonsevere events). Together, the findings suggest that an individual's frame of reference for the severity of events plays a role in narrowing or broadening the concept of trauma.

Public Significance Statement
This study suggests that when making decisions about whether an event qualifies as a "trauma," individuals rely on recent context. Specifically, when the context includes severe events (e.g., "was injured in an IED explosion") participants were less likely to classify any given event as traumatic. Accordingly, context may influence how individuals interpret potentially traumatic events in their lives.

Keywords: trauma, concept creep, PTSD

Supplemental materials: https://doi.org/10.1037/xap0000389.supp

Trauma centers treat injuries resulting from sudden physical insults to the body. Psychological trauma is a metaphorical extension of the medical term applied to emotional harm, formalized in the appearance of posttraumatic stress disorder (PTSD) in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III; American Psychiatric Association [APA], 1980). The diagnosis emerged in part due to discussions surrounding "post-Vietnam war syndrome" (Shatan, 1973). The psychological complications of war veterans were grouped with those of individuals who were traumatized by rape, natural disasters, or events such as the Holocaust to form the category of PTSD (McNally, 2003a). The original formulation presumed that PTSD could only arise following exposure to terrifying, presumably rare events falling outside the boundary of ordinary experience. Yet the concept of trauma has increasingly expanded to embrace a wider range of stressful events (McNally, 2016).

This raises an important question for diagnosis, treatment, and related policies: what is trauma? From a diagnostic standpoint, the current edition of the DSM (DSM-5) maintains the relatively strict definition of "exposure to actual or threatened death, serious injury, or sexual violence" (APA, 2013, p. 271). Yet among laypersons, "trauma" often has a much broader meaning. In linguistic analyses, "trauma" is shown to have expanded semantically (Vylomova et al., 2019) and is used with increasing relative frequency (Haslam & McGrath, 2020). Media outlets have applied it to include experiencing microaggressions (Williams, 2015), reading the news (Jacobs, 2018), or learning secondhand about "difficult or disturbing stories" (Lees, 2018, para. 3).

Expansions in the usage of the term "trauma" over time may indicate "conceptual bracket creep in the definition of trauma" (McNally, 2003b, p. 231). The term "creep" is here used to denote the expansion of a semantic boundary over time. Haslam (2016) suggests that a wide variety of harm-related concepts (such as bullying, abuse, and prejudice) have similarly crept. Psychologists vary in their views on the benefits and costs of expanding the definition of trauma and other harm-related concepts. On one hand, such expansions may reflect a development in deeper empathy toward individuals who suffer from various types of negative events (Haslam, 2016). In addition, expansions in the concept of trauma may reduce uncertainty about the (un)acceptability of certain behaviors, empowering victims and third-party allies to take more decisive action (Cikara, 2016).

Payton J. Jones https://orcid.org/0000-0001-6513-8498
David E. Levari https://orcid.org/0000-0002-7953-7999
Benjamin W. Bellet https://orcid.org/0000-0002-4338-3393
Richard J. McNally https://orcid.org/0000-0002-5228-8777

This research was supported by NSF grant DGE1745303 and by the Talley Fund.

Correspondence concerning this article should be addressed to Payton J. Jones, Department of Psychology, Harvard University, 1240 William James Hall, 33 Kirkland Street, Cambridge, MA 02138, United States. Email: paytonjjones@gmail.com

On the other hand, some scholars worry that expansions in the concept of trauma dilute the meaning of the term (Haslam, 2016; McNally, 2016). If even minor harms are traumas then nearly every adult should equally be considered a trauma survivor. From a sociocultural standpoint, Haslam (2016) worried that diluting the public's perception of the trauma concept might "[risk] reducing the range of people who see themselves as capable of moral agency," therefore increasing "a tendency for more and more people to see themselves as victims who are defined by their suffering, vulnerability, and innocence, and who have diminished agency to overcome their plight." From a clinical standpoint, McNally (2009, p. 598) wondered whether it might alter the etiological understanding of PTSD, "[undermining] the very rationale for having a diagnosis of PTSD in the first place." Interpreting adverse events and acute emotional reactions to events through the lens of trauma and PTSD may also cause individuals to generate more negative appraisals, increasing the risk of long-term harm (Beierl et al., 2020).

Regardless of whether harm-related concept creep is helpful or unhelpful, little is understood about how and why it occurs. One possibility is that concept expansions occur as a result of decreasing frequencies of exposure. For example, if overt violence occurs at a low frequency, individuals may shift their concept boundaries to encompass additional examples (e.g., classifying hateful speech as violence). Indeed, in a series of experiments, Levari et al. (2018) found that shifting the prevalence of displayed categories reliably alters conceptual boundaries across a variety of stimuli. For example, when participants were shown decreasing amounts of threatening faces (relative to nonthreatening faces), they expanded the range of faces they classified as threatening. The researchers dubbed this effect "prevalence-induced concept change," referring to the semantic shift that occurs due to changes in the relative prevalence of certain kinds of stimuli. Prevalence-induced concept change is proposed as a broad within-person mechanism that could apply to the expansion or contraction of any kind of category (not just harm-related concepts). Harm-related concepts specifically (violence, trauma, and abuse) are of special interest because interpersonal violence and related forms of harm have decreased over time, in some cases quite drastically (for summaries see Jones, 2021; Pinker, 2011; see also Bureau of Justice Statistics, 2016; Eisner, 2003; Federal Bureau of Investigation [FBI], 2020).

Another possibility is that harm-related concept expansions may occur as a result of an altered range of reference. Parducci (1965) argued that when making decisions, humans are sensitive to both the relative *frequency* of different types of stimuli, as well as the absolute *range* of stimuli. For example, when gauging the severity of an injury such as a sprained ankle, a person's judgment may be influenced by their perception of how common it is relative to other injuries like broken bones and lacerations (the *frequency*) but also by the most and least severe injuries that come to mind (the *range*).

Recent research has suggested that this is a particularly computationally efficient way for the brain to make subjective evaluations (Bhui & Gershman, 2018), which may explain why it has been documented in domains as diverse as judgments of loudness (Jesteadt et al., 1977) and product prices (Niedrich et al., 2009). In cases where "true" signals must be differentiated from "false" noise, signal detection theory outlines how shifting one's criterion threshold (response bias, or β) in response to changes in the observed distribution optimizes the balance between false positives and false negatives (see Macmillan, 2002). Although subjective judgments of which faces are threatening or which events are traumatic lack the objective "ground truth" of many detection tasks, the logic of signal detection theory may nevertheless provide an evolutionary explanation for this aspect of human psychology.

It is important to distinguish between range and frequency as pieces of a cognitive puzzle existing in the minds of individuals (as mechanisms), and the ranges and frequencies of different stimuli that individuals encounter in real life or in an experiment (as environments). In an experimental setting, fully disentangling the cognitive mechanisms of range and frequency is difficult but doing so is not our aim. Instead, we hope to determine if manipulations of objective, environmental (rather than mental) ranges, and frequencies result in the expansion of the trauma concept. Specifically, we hope to determine whether decreasing the frequency of serious events results in expansions of the trauma concept and whether restricting ranges of seriousness affects judgments.

When it comes to establishing the recent context in terms of the range and frequency of events, two distinct types of experience might be influential: (a) experiencing events directly oneself or (b) hearing about, witnessing, or otherwise learning about events. In the lab, of course, we must rely exclusively on the latter. This provides an imperfect but meaningful analogue to real-world declines in potentially traumatic events such as violence (see Jones, 2021; Pinker, 2011).

We attempt to test the influence of displayed stimuli (descriptions of events) as they relate to binary judgments of events as either "trauma" or "not trauma." In the first study, we conducted a preregistered extension of Levari et al.'s (2018) experiments. Specifically, we asked participants to classify events as either trauma or not trauma while manipulating the relative frequency of severe and nonsevere events (compared to a control condition). In the second study, we repeated a similar experiment but manipulated the total range of event severity while maintaining the relative frequencies as a constant.

Experiment 1

Method

Stimuli Generation and Norming

We generated 600 descriptions of events covering the entire spectrum from "not at all traumatic" to "extremely traumatic." The descriptions ranged in length from 2 to 16 words and varied widely in their thematic and emotional content (e.g., "walked up a flight of stairs" and "killed a child pedestrian while driving").

To obtain initial objective ratings of the stimuli, we conducted a pretest. We randomly divided the stimuli into six equal sets of 110 items. In each set, 98 items were unique to the set, whereas the other 12 items appeared in all sets, serving as a consistency check. We presented each set of descriptions in random order to participants recruited from Amazon Mechanical Turk (MTurk) ($n_{total} = 250$ and $n_{set} \approx 42$). Participants were asked to rate each description on a 7-point Likert scale from *Not at all traumatic* to *Extremely traumatic*. Interrater reliability on the consistent set of 12 items was good when assessed across each of 250 participants as separate judges ($ICC_1 = 0.70$; Shrout & Fleiss, 1979), and excellent when considering the average value across each of the six sets ($ICC_1 = 0.99$). Inclusion and exclusion criteria for pilot raters were the same as for the main

experiments, as detailed below. Demographic information of raters is included in the Supplemental Materials (osf.io/3e2us). Pilot raters were ineligible to participate in the experiments.

Procedure

We preregistered Experiment 1 on the Open Science Framework (osf.io/tw92r). Participants were recruited from MTurk. They were allowed to participate in the study if they were adult United States residents and had an MTurk approval rate of 95% or greater. Participants first completed a CAPTCHA and U.S. residency screener (e.g., "What emergency number is most common in the United States?") and were immediately excluded if they failed either task. As preregistered, we recruited participants until a total of 300 had completed the study (which required passing this initial screener).

Participants were first given basic instructions regarding the survey.[1] They were then shown each event in sequence in a standardized window and were instructed to press one of two keys to indicate either "trauma" or "not trauma" for each description. They viewed each item for a minimum of 1.5 s before clicking and were instructed to take a break every 30 items (> 5 s). Each participant viewed a total of 300 items. Attention checks were interspersed throughout this task (i.e., "please press the p key on your keyboard"). After rating all items for their condition, participants completed a demographics and psychiatric history questionnaire and a human participant verifier (writing three sentences about the past weekend). They were then shown a debriefing form explaining the purpose of the experiment. As preregistered, participants were excluded from the analysis if they incorrectly answered attention checks or failed the human participant verifier during the experiment. A total of 24 participants were excluded, leaving a total of 276 participants.

Using the rating data from the pretest, we sorted items into categories depending on their mean rating on the 7-point Likert scale in the pretest: Nonsevere events (M = 1–2), ambiguous events (M = 3–4), and severe events (M = 5–7). We then selected 342 descriptions of events that corresponded to nonsevere events (142 descriptions, e.g., "walked up a flight of stairs"), ambiguous events (100 descriptions, e.g., "broke an ankle while running"), or severe events (100 descriptions, e.g., "killed a child pedestrian while driving") to be used in the main study. We intentionally selected items that had acceptably low standard deviations in the pretest (SD < 1.6) to avoid selecting items that were inconsistently interpreted by different participants. The exact number of items in each category was determined by a calculation that ensured we would have sufficient unique stimuli in each category given our experimental design.

We will refer to the probability that participants were shown an item from the severe category in each block as the signal prevalence. For participants in the control ("stable") condition (n = 138), participants were shown severe events with a signal prevalence of 33.3% throughout the experiment. For participants in the experimental ("decreasing") condition (n = 138), we modified the signal prevalence over time. The signal prevalence was set at 33.3% for the first 100 trials, 25% for the next 50 trials, 16.6% for the next 50 trials, 8.3% for the next 50 trials, and 4.12% for the last 50 trials.[2] The decrease in the probability of severe events was balanced by an increase in the probability of nonsevere events, whereas the frequency of ambiguous events remained constant.

Analysis

To analyze the data, we used the lme4 package (Bates et al., 2015; R Core Team, 2019). We fit a binomial generalized linear mixed-effects model to the data, estimated via bound optimization by quadratic approximation (BOBYQA; Powell, 2009). In each experiment, the dependent variable was the participants' identification of a stimulus as "trauma" or "not trauma." Random effects were added for the (a) intercepts for participants and (b) slopes for trial number if they improved model fit as determined by a χ^2 test.

Based on the results of Levari et al. (2018), we predicted that individuals in the stable condition would remain consistent in their ratings over time, whereas individuals in the decreasing condition would become more lenient in their concept of trauma (i.e., have a higher likelihood of rating nonsevere or ambiguous events as trauma in later trials). The predictor variables in our binomial generalized linear mixed-effects model were experimental condition, pilot ratings of descriptions (i.e., "objective severity" of each description), trial number, and their interactions. We expected a significant three-way interaction of the predictor variables, indicating that individuals in the decreasing condition would classify nonsevere or ambiguous items as "trauma" at a higher likelihood compared to individuals in the stable condition, but only at later trials.

Results

Our participants were predominantly male (56%), Caucasian/White (81%), Non-Hispanic (96%), not religious (56%), reported no history of serious trauma[3] (72%), and no history of diagnosed mental illness (85%). A table displaying full demographic information is provided in the Supplemental Materials.

A model including both random effects for the intercepts of participants and slopes for trial number showed superior fit compared to models excluding either of these effects, as determined by a χ^2 test (p < .001 and N = 276). Our prediction of a three-way interaction between condition, objective severity of each item, and trial number was not supported ($\beta = -0.21$, $z = -0.35$, p = .73, and model dispersion = 0.63). Instead, participants became stricter in their threshold of assigning the descriptor of "trauma" over time across *both* conditions. This effect is visualized in Figure 1. Over

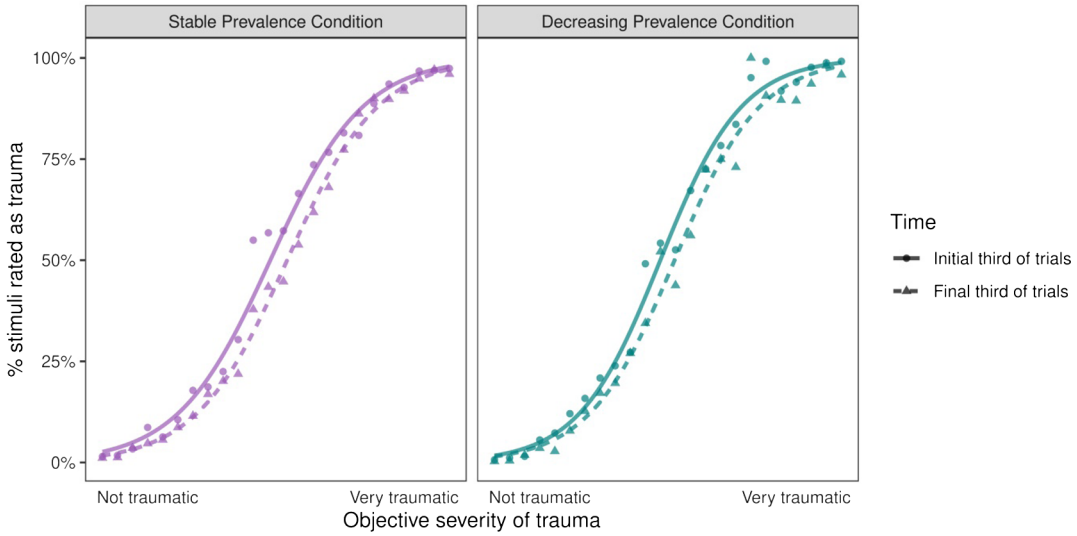
[1] "Some events that happen in life may be considered to be trauma. In this survey, you will read a series of [X] descriptions of events that may occur in a person's life. For each description, you will be asked to decide whether the event itself is "trauma" or "not trauma". There are no right or wrong answers."

[2] Due to an error in the item selection algorithm, the last 25 trials showed items at an incorrect signal prevalence. These final 25 trials were therefore removed in all reported analyses for Experiment 1. When these faulty trials are included, the results do not change except that the interaction between trial and objective severity becomes nonsignificant, which is not relevant to our hypothesis.

[3] "Some people experience extremely distressing events that are outside the range of common experience. Examples of such an event include being in a life-threatening situation such as military combat, a natural disaster (floods, earthquakes), a man-made disaster (being in a car accident where someone was seriously injured or killed); being raped, being violently assaulted, or being tortured. At any time in your life, have you directly experienced any of these kinds of events?"

Figure 1

No Effect of Prevalence but an Effect of Trial Number on Trauma Ratings



Note. Each dot represents an event description, with lines drawn to indicate the threshold at which participants on average made the binary split between "not trauma" or "trauma" based on objective severity. Participants in both conditions rated descriptions more strictly in later trials compared to earlier ones. See Supplemental Figure S1 for a finer-grained breakdown across trials. See the online article for the color version of this figure.

increasing trial numbers, the threshold shifted to the right (increasingly strict) in both conditions. Because the three-way interaction was not supported, we subsequently tested for a two-way interaction, as preregistered. In the model without the three-way interaction, there was no two-way interaction between trial number and condition ($\beta = 0.10$, $z = 0.53$, and $p = .59$), indicating that any shifting of the thresholds across trial number did not differ by condition. We found significant two-way interactions between trial number and objective severity, and between condition and objective severity, neither relevant to our hypotheses ($\beta = 1.20$, -1.69, $z = 3.67$, -8.97, and $p_s < .001$). The first interaction indicated that across conditions, the slope was slightly steeper at earlier trial numbers. The second interaction indicated that across time points, the slope of the stable condition was less steep than the slope in the decreasing condition. This interaction was especially unexpected because the experimental manipulation was applied gradually, so any effect involving the conditions would presumably interact with the trial number (i.e., we expect differences in later trial numbers when the conditions are distinct but not in earlier ones when the conditions are identical). To ensure that this two-way interaction model was appropriate, we tested for differences in fit between it and the three-way interaction model. There was no significant difference in fit between this model and the model including the three-way interaction ($p = .73$).

In an exploratory model including only main effects, we indeed observed a main effect of trial number ($\beta = -0.81$, $z = 8.78$, and $p < .001$), which corresponds to the shift toward the right depicted in the graph. As noted earlier, the effect of trial number was not moderated by the condition when two-way interactions were included. We were surprised to see the movement of participants' rating threshold within the stable condition, as this effect was absent in all seven of Levari et al.'s (2018) experiments. Tables with complete details on each model are available in Supplemental Materials (osf.io/3e2us). Overall, these results did not provide support for prevalence-induced concept change in the rating of descriptions of traumatic events. Instead, they seemed to indicate the presence of another, unexpected effect of narrowing across both conditions.

Discussion

In Experiment 1, we did not find evidence for a prevalence-induced concept change. Participants in the decreasing condition did not expand their concept of trauma. Instead, we found an unexpected effect. Regardless of whether the prevalence of severe traumas changed over time, participants became increasingly strict in what they classified as trauma. After rating the first several items, participants became stricter, assigning the label of "trauma" only to relatively severe events. This effect appeared strongest for the first few trials, with a gradual but smaller shift continuing in later trials (see Supplemental Figure S1).

This diverges from previous findings such as when participants rated the threateningness of faces or ethicality of research proposals. Why might "trauma" not follow the same pattern? Perhaps, potentially dangerous events form a special mental category due to their relevance to survival. The mind may mark such events as important regardless of their perceived prevalence. For example, many people are afraid of serial killers despite no real-world exposure to them and a low risk of ever encountering one. The vividness of the imagined scenario is enough to cement persistent fear even when the event is perceived as rare.

Alternatively, our "trauma" stimuli might follow a unique pattern for an opposite reason—the concept of psychological "trauma" is relatively abstract, contemporary, and varies by culture (i.e., is

ontogenetic). In contrast, the threateningness of faces or the moral categories of "good versus bad" may have a more phylogenetic origin (Mikhail, 2007; Öhman et al., 2001). Ratings of psychological trauma may follow a process that diverges from the process participants use to rate threateningness, morality, or color.

Why might individuals' trauma concept narrow even in the decreasing condition? One possibility concerns the initial *range* in participants' working concepts of trauma, what Parducci called an "implicit frame of reference." Some of our more severe trauma items included "was tortured as a prisoner of war" and "was raped by a close friend." When our sample of relatively young American participants initially considered the term trauma, such events may not readily be brought to mind. In other words, it is possible that participants saw descriptions in the experiment that were more severe than they expected to see. We hypothesize that their initial implicit range encompassed low-to-moderate severity; as the experiment progressed, their implicit range expanded to encompass the full range of low-to-high severity.

We, therefore, tested whether the narrowing effect we observed in Study 1 was attributable to participants' frame of reference. We thus devised a new experiment with experimental conditions that directly addressed the range of events shown to participants.

Experiment 2

Method

Our participant recruitment and data collection methods were the same as in Experiment 1, but we modified the conditions to test our hypothesis about the range of events. After exclusion, 267 participants remained (33 participants were excluded). In the first condition (hereafter "nonsevere range" condition; $n = 135$), participants were shown only items decisively judged to be nontraumatic in the stimuli norming pretest (range of mean: 1–3).[4] To illustrate, the least severe event in the nonsevere range was "walked up a flight of stairs" and the most severe was "was not hired after a job interview." In the second condition (hereafter "severe range" condition; $n = 132$), participants were shown only items decisively judged to be in the range of severe traumatic events (range of mean: 5–7). The least severe event in the severe range was "received chemotherapy" and the most severe was "was raped by a family member." Participants viewed 90 items for which the relative frequencies within each condition remained constant throughout (no frequency manipulation). If our hypothesis concerning the effect of the range of events was correct, we expected to see significantly stricter ratings in the severe condition compared to the nonsevere condition.

To analyze the data, we used a binomial generalized linear mixed-effects model. Our dependent variable was participants' binary rating of each stimulus as "trauma" or "not trauma." The predictor variables were experimental condition, pilot ratings of descriptions (i.e., "objective severity" of each description), and trial number. In this case, we did not model a three-way interaction, as we did not expect the objective severity of trauma to interact with the trial number and the condition. Instead, we hypothesized a significant two-way interaction of the condition and trial number, indicating that individuals in the severe range condition would classify items as "trauma" at a lower likelihood compared to individuals in the nonsevere condition, but primarily at later trials. Our model included the main effects and the hypothesized interaction term.

Results

Participants in Experiment 2 were predominantly male (58%), Caucasian/White (75%), non-Hispanic (90%), not religious (58%), trauma-naïve (64%), and with no history of diagnosed mental illness (82%). A table displaying complete demographic information is provided in the Supplemental Materials.

Our prediction of a significant two-way interaction of the variables was not supported ($\beta = 0.22, z = 0.84, p = .40, N = 267$, and model dispersion = 0.68). Instead, our prediction that individuals in the severe condition would classify items as "trauma" at a lower likelihood was supported, but this effect was not moderated by trial number. This is reflected by the main effect of condition on participant ratings, which remained significant regardless of whether the nonsignificant interaction term was in the model ($\beta = 5.58.\ 5.66$, $z = 17.79$, 19.243, and $p_s < .001$; see Figure 2). We suspect the moderation by trial number was not found because the effect manifested quickly rather than linearly over time. Indeed, in an exploratory model in which time was defined by a front-loaded binary indicator (first 10 trials vs. all later trials) rather than a linear one, the expected interaction was found ($p < .001$)."

To interpret the results, it is useful to consider them in relation to the Experiment 1 data. In the stable condition of Experiment 1, participants were shown events from the full range in random order. Because there was no manipulation of frequency, this condition is conceptually identical to the two Experiment 2 conditions. We can therefore use these data to visualize a hypothetical "full range" condition. Keep in mind that this visualization is intended for explanatory purposes, not for inference.

If the range hypothesis were correct, we would expect participants in the "nonsevere range" condition to be the most lenient, followed by participants in the proxy "full range" condition, with individuals in the "severe range" condition behaving the most strictly. Indeed, this pattern applies, clearly visible in Figure 3. Individuals in the nonsevere range condition have the most leftward curve, indicating that even mild items were frequently classified as traumatic. In contrast, individuals in the severe range condition had the curve farthest to the right, indicating that even severe descriptions were often classified as nontraumatic. The proxy full range condition falls between these two. Recall that when separated by time, the early trials of the full range condition produced a curve farther to the left, whereas the later trials produced a curve farther to the right.

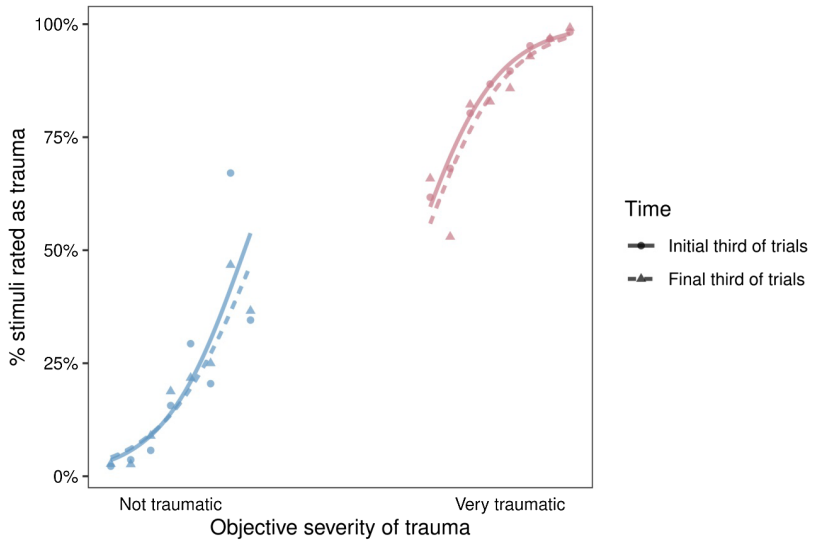
General Discussion

Concepts can expand over time to include events hitherto deemed to fall outside their original boundaries. In our first experiment, we sought to determine whether the concept of trauma would expand as the most severe instances of trauma became increasingly rare. We did not find evidence of prevalence-induced concept change for trauma. Instead, we found an unexpected effect whereby individuals became stricter in their trauma ratings over time, regardless of altered frequencies.

[4] To provide at least some diversity in event severity and mitigate demand characteristics, we slightly expanded the range in the non-severe category (defined as 1–2 in Experiment 1). For reference, the events "was sick with the common cold" and "overslept and arrived late to work" were tied for the most severe events in this category in Experiment 1.

Figure 2

Range Restrictions in Ratings of Trauma



Note. Each dot represents an event description, with lines drawn to indicate the threshold at which participants on average made the binary split between "not trauma" or "trauma" based on objective severity. Controlling for the objective severity of the descriptions, participants who saw only severe events (red, right) were much stricter than expected compared to participants who saw only nonsevere events (blue, left). See Supplemental Figure S2 for a finer-grained breakdown across trials. See the online article for the color version of this figure.

Perhaps people underweight the role of frequency when evaluating potentially dangerous events. For example, many people dread shark attacks despite their rarity. Thus, exposure to severe events may affect trauma concepts even when exceedingly rare. We hypothesized that exposure to our most extreme examples broadened participants' implicit range of events.

We tested this range hypothesis in our second experiment. Indeed, we found that altering the range of events shown to participants influenced their ratings of trauma. Participants who saw only nonsevere events were lenient in classifying events as trauma compared to those who saw severe events. A shift in the working range of events thus provides one plausible explanation for the narrowing effect we observed in the first experiment. This finding is reminiscent of context effects that have been observed for decades across diverse arenas of decision making (e.g., range-frequency, Parducci, 1965; anchoring, Tversky & Kahneman, 1974; and signal detection, Macmillan, 2002).

What implications does this have for the concept creep of "trauma"? It seems that an individual's frame of reference may play a greater role than frequencies of events within that frame of reference. This suggests that the perceived *absence* of certain threats (genocide does not happen in Boston) may play a special role beyond the perceived *rarity* of threats (assault happens rarely in Boston). That is, harm-related concepts may take especially large leaps forward when the most extreme events are eliminated from the public consciousness. Pinker (2011, Chapter 1) provides dramatic coverage of such examples, reminding us of truly horrific events that were once common but have since disappeared from our concerns: being broken on the wheel, forced to fight to the death for others' entertainment, crucified, burned at the stake, and a litany of other terrible fates. More recently, perhaps young people in developed nations can mostly discard worries about being paralyzed by polio, captured and tortured by their government, extorted by the mafia, or drafted to the front lines of an interstate war. That is not to say that the public ever becomes *unaware* of such events, but that the events cease to be relevant concerns that quickly spring to mind when considering the concept of trauma (Phillips et al., 2019).

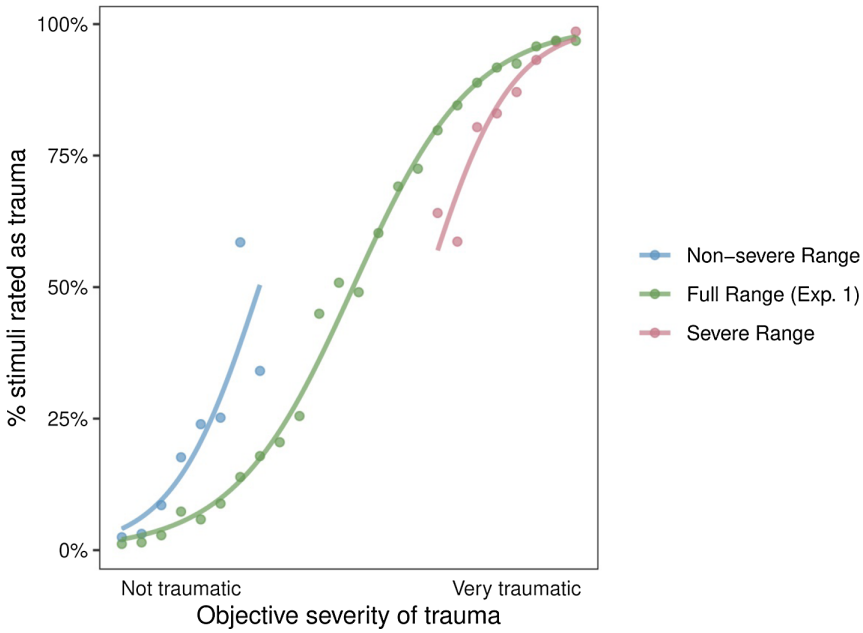
Our results suggest that implicit frames of reference shape individual judgments about the breadth of the trauma concept. This may explain seemingly paradoxical trends in the prevalence of trauma and PTSD. Epidemiological studies often indicate very high self-reported rates of exposure to trauma in first-world countries, despite relatively low rates of violence and disaster in those same countries. In one remarkable example, the lifetime rate of exposure to trauma in Canada was recorded as slightly higher than the rate in South Africa (76% vs. 74%; Dückers et al., 2016).[5]

Furthermore, countries with high vulnerability indices (e.g., high-income inequality and political corruption), such as Mexico and South Africa, have *lower* rates of PTSD following trauma compared to countries with lower vulnerability indices, such as Canada and the United States (Dückers et al., 2016). In the United

[5] Although these studies typically limit reported traumatic events to DSM Criterion A, flexibility remains in how participants classify their own experiences (e.g., perhaps more participants in South Africa fail to report domestic disputes as physical assaults). An alternative explanation is that trauma is more evenly distributed in Canada, with individuals in South Africa experiencing a much greater number of traumas per person, but with both countries having a similar number of individuals never experiencing trauma.

Figure 3

The Influence of Range on Trauma Ratings



Note. Each dot represents an event description, with lines drawn to indicate the threshold at which participants on average made the binary split between "not trauma" or "trauma" based on objective severity. Participants had the strictest threshold (e.g., only very severe events are "trauma") when they saw only severe events, and the most lenient threshold (e.g., even some minor events are "trauma") when they saw only nonsevere events. See the online article for the color version of this figure.

States, rates of violent crimes such as rape steadily declined between the early 1990s and 2010s[6] (FBI, 2020; Pinker, 2011). Yet during the same timeframe, rates of PTSD have remained relatively stable (Kessler et al., 1995, 2005; Kilpatrick et al., 2013).

Our results help in explaining why the same stressors may be perceived as being more traumatic in an advantaged society. This is not only an epidemiological issue but potentially a clinical one. Negatively appraising an adverse event and its sequelae forms the initial basis for developing PTSD in the cognitive model (Ehlers & Clark, 2000). Negative appraisals of an event incrementally predict long-term PTSD rates above and beyond other risk factors (e.g., Bryant & Guthrie, 2005; Ponnamperuma & Nicolson, 2016).

It is unclear whether classifying an event as a "trauma" necessarily leads to a more negative appraisal of that event. However, it does seem plausible that the two might be linked. For instance, the individual who classifies an event as a "trauma" may be more likely to see the event through the lens of permanent harm, importance to life narrative, and risk for PTSD. The term *trauma* implies lasting damage, whereas other terms (such as adversity, hardship, or negative event) do not. Applying a relatively severe label to an event may promote rumination about the event and increase vulnerability (Berntsen & Rubin, 2007).

Trauma concepts are also relevant to therapists. A therapist who works in the criminal justice system may be less likely to categorize a distressing event as a Criterion A Trauma compared to a therapist who works primarily with the worried-well. Indeed, manipulations of immediate context have previously been shown to alter standardized ratings of psychopathology, a problem insufficiently addressed in clinical practice (Wedell et al., 1990). In the broader political and educational landscape, our results suggest that reminders of very severe events may curb the categorization of relatively minor events as "trauma."

Our research has several limitations. Our sample was restricted to American adults on MTurk, most of whom had no trauma history. We expect that base rates of judgments about trauma might differ in other demographics. We attempted to reduce demand characteristics (e.g., by emphasizing in the instructions that there were no right or wrong answers), but we cannot be certain that demand characteristics were fully absent. For instance, a participant who saw primarily nonsevere events might have rated some event as "trauma" that they did not truly believe fits the category merely to please the researcher.[7] Importantly, our experiments cannot fully disentangle range from frequency in the mechanistic sense of range-frequency

[6] Various complexities apply to tracking rape and other violent crimes. The FBI Uniform Crime Reporting program aggregates reported crimes, underestimating the true prevalence. However, since the methods of data collection were kept consistent during this period, changes over time can be asserted with confidence. Between 1992 and 2014, rates of rape dropped 38%, aggravated assault dropped 48%, and homicide dropped 53%.

[7] Notably, in Levari et al.'s (2018) studies, participants' ratings shifted even when participants were told ahead of time that the prevalence would change and even when they were paid to remain consistent in their ratings over time. This somewhat reduces concerns that the effects of stimuli range/frequency are primarily due to demand characteristics.

theory. Instead, they provide information on specific manipulations. Although manipulating the range seemed to have the most prominent effect in our experiments, it remains unclear whether there exist conditions under which manipulation of the frequency would affect ratings of trauma. It is possible that changes in frequency do affect rating thresholds, but only over a much longer period. We also only tested a decreasing frequency over time compared to a stable frequency, having no condition for increasing frequency.

It is possible that the range effects we obtained in this short timeframe are not durable. Indeed, the fact that ratings can be so easily manipulated suggests that immediate context is relatively powerful, and thus any interventions that influence trauma concepts may be overridden in future contexts. This study only provides information about exposure to *descriptions* of events, not direct exposure to events. It seems likely that direct exposure to events might influence trauma concepts more drastically or more durably.[8] Future studies might explore the effect of exposure to events through audio, video, or virtual reality.

The concept of psychological trauma has expanded rapidly in the sociopolitical arena, as well as in psychiatric diagnosis. We investigated whether a paradigm assessing prevalence-induced concept change could help explain this concept creep. Our results indicated that unlike stimuli tested in previous experiments, such as ratings of color, threateningness, or ethicality (Levari et al., 2018), the categorization of trauma was not significantly affected by changes in prevalence. Instead, when individuals were shown the full range of events, they became stricter over time in their willingness to categorize events as trauma. A second experiment revealed that manipulating the range of events altered trauma ratings: individuals shown mostly benign events were lenient in categorizing events as trauma, whereas individuals shown mostly severe events were more restrained. Thus, expansions in the concept of trauma may occur primarily in frames of reference in which very severe events are absent.

[8] At the suggestion of a reviewer, we tested whether prior exposure to trauma affected ratings of events. In Experiment 1, those with prior trauma exposure were *more* likely to rate events as trauma, holding other variables constant ($p < .01$). In Experiment 2, there was no significant effect. Because prior trauma exposure was based on subjective self-reports, it is unclear whether the effect is due to differences in trauma exposure or differences in self-reporting (e.g., Hardt & Rutter, 2004).

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beierl, E. T., Böllinghaus, I., Clark, D. M., Glucksman, E., & Ehlers, A. (2020). Cognitive paths from trauma to posttraumatic stress disorder: A prospective study of Ehlers and Clark's model in survivors of assaults or road traffic collisions. *Psychological Medicine*, *50*(13), 2172–2181. https://doi.org/10.1017/S0033291719002253

Berntsen, D., & Rubin, D. C. (2007). When a trauma becomes a key to identity: Enhanced integration of trauma memories predicts posttraumatic stress disorder symptoms. *Applied Cognitive Psychology*, *21*, 417–431. https://doi.org/10.1002/acp.1290

Bhui, R., & Gershman, S. J. (2018). Decision by sampling implements efficient coding of psychoeconomic functions. *Psychological Review*, *125*, 985–1001. https://doi.org/10.1037/rev0000123

Bryant, R. A., & Guthrie, R. M. (2005). Maladaptive appraisals as a risk factor for posttraumatic stress: A study of trainee firefighters. *Psychological Science*, *16*, 749–752. https://doi.org/10.1111/j.1467-9280.2005.01608.x

Bureau of Justice Statistics. (2016). *US violent victimization at school*. NCVS Victimization Analysis Tool. https://www.bjs.gov/index.cfm?ty=nvat

Cikara, M. (2016). Concept expansion as a source of empowerment. *Psychological Inquiry*, *27*, 29–33. https://doi.org/10.1080/1047840X.2016.1111830

Dückers, M. L., Alisic, E., & Brewin, C. R. (2016). A vulnerability paradox in the cross-national prevalence of post-traumatic stress disorder. *The British Journal of Psychiatry*, *209*, 300–305. https://doi.org/10.1192/bjp.bp.115.176628

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*, 319–345. https://doi.org/10.1016/S0005-7967(99)00123-0

Eisner, M. (2003). Long-term historical trends in violent crime. *Crime and Justice*, *30*, 83–142. https://doi.org/10.1086/652229

Federal Bureau of Investigation. (2020). *FBI uniform crime reporting (UCR) program data explorer*. https://crime-data-explorer.fr.cloud.gov/

Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(2), 260–273. https://doi.org/10.1111/j.1469-7610.2004.00218.x

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, *27*, 1–17. https://doi.org/10.1080/1047840X.2016.1082418

Haslam, N., & McGrath, M. J. (2020). The creeping concept of trauma. *Social Research*, *87*(3), 509–531. https://muse.jhu.edu/article/773952

Jacobs, J. (2018). When the news itself is a form of trauma. *New York Times*. https://www.nytimes.com/2018/09/26/us/metoo-survivors-kavanaugh-cosby.html

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 92–104. https://doi.org/10.1037/0096-1523.3.1.92

Jones, P. J. (2021). *The neurotic treadmill: Decreasing adversity, increasing vulnerability?* [Dissertation]. ProQuest Dissertations and Theses database.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602. https://doi.org/10.1001/archpsyc.62.6.593

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. B. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *52*, 1048–1060. https://doi.org/10.1001/archpsyc.1995.03950240066012

Kilpatrick, D. G., Resnick, H. S., Milanak, M. E., Miller, M. W., Keyes, K. M., & Friedman, M. J. (2013). National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. *Journal of Traumatic Stress*, *26*, 537–547. https://doi.org/10.1002/jts.21848

Lees, A. B. (2018). Yes, you can be traumatized by the media! *Psychology Today*. https://www.psychologytoday.com/us/blog/surviving-thriving/201810/yes-you-can-be-traumatized-the-media

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*, 1465–1467. https://doi.org/10.1126/science.aap8731

Macmillan, N. A. (2002). Signal detection theory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (pp. 43–90). Wiley. https://doi.org/10.1002/0471214426.pas0402

McNally, R. J. (2003a). Progress and controversy in the study of posttraumatic stress disorder. *Annual Review of Psychology*, *54*, 229–252. https://doi.org/10.1146/annurev.psych.54.101601.145112

McNally, R. J. (2003b). *Remembering trauma*. Harvard University Press.

McNally, R. J. (2009). Can we fix PTSD in DSM-V? *Depression and Anxiety*, *26*, 597–600. https://doi.org/10.1002/da.20586

McNally, R. J. (2016). The expanding empire of psychopathology: The case of PTSD. *Psychological Inquiry*, *27*, 46–49. https://doi.org/10.1080/1047840X.2016.1108168

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152. https://doi.org/10.1016/j.tics.2006.12.007

Niedrich, R. W., Weathers, D., Hill, R. C., & Bell, D. R. (2009). Specifying price judgments with range–frequency theory in models of brand choice. *Journal of Marketing Research*, *46*, 693–702. https://doi.org/10.1509/jmkr.46.5.693

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407–418. https://doi.org/10.1037/h0022602

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007

Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. Viking.

Ponnamperuma, T., & Nicolson, N. A. (2016). Negative trauma appraisals and PTSD symptoms in Sri Lankan adolescents. *Journal of Abnormal Child Psychology*, *44*, 245–255. https://doi.org/10.1007/s10802-015-9985-y

Powell, M. J. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives* (Cambridge NA Report NA2009/06). University of Cambridge. http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Shatan, C. F. (1973). The grief of soldiers: Vietnam combat veterans' self-help movement. *American Journal of Orthopsychiatry*, *43*, 640–653. https://doi.org/10.1111/j.1939-0025.1973.tb00834.x

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vylomova, E., Murphy, S., & Haslam, N. (2019). *Evaluation of semantic change of harm-related concepts in psychology* [Conference session]. Proceedings of the 1st international workshop on computational approaches to historical language change, Florence, Italy. https://doi.org/10.18653/v1/W19-4704

Wedell, D. H., Parducci, A., & Lane, M. (1990). Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and type of anchors. *Journal of Personality and Social Psychology*, *58*, 319–329. https://doi.org/10.1037/0022-3514.58.2.319

Williams, M. T. (2015). The link between racism and PTSD. *Psychology Today*. https://www.psychologytoday.com/us/blog/culturally-speaking/201509/the-link-between-racism-and-ptsd

Received January 19, 2021
Revision received June 23, 2021
Accepted June 30, 2021 ■