

# Correlates of “Coddling”: Cognitive distortions predict safetyism-inspired beliefs, belief that words can harm, and trigger warning endorsement in college students<sup>☆</sup>

Jared B. Celniker<sup>a,\*</sup>, Megan M. Ringel<sup>b</sup>, Karli Nelson<sup>a</sup>, Peter H. Ditto<sup>a</sup>

<sup>a</sup> Department of Psychological Science, University of California, Irvine, United States of America

<sup>b</sup> Department of Psychology and Behavioral Neuroscience, St. Edwards University, United States of America

## ARTICLE INFO

### Keywords:

Cognitive distortions  
College students  
Trigger warnings  
Open data

## ABSTRACT

In their book, *The Coddling of the American Mind*, Lukianoff and Haidt (2018) contended that the rise of “safetyism” within American society has inspired beliefs and practices that hinder college students’ socioemotional development. One of their most controversial claims was that college students’ safetyism-inspired beliefs (e.g., emotional pain or discomfort is dangerous) are rooted in and supported by cognitive distortions, or negatively biased patterns of thought (e.g., emotional reasoning). Citing evocative anecdotes, they argued that such distortions emerge in students’ perceptions of offensive or ideologically-challenging experiences as disproportionately harmful or traumatic. However, no empirical work has substantiated an association between cognitive distortions and safetyism-inspired beliefs or practices. In a large ( $N = 786$ ), ethnically and economically diverse sample of college students, we conducted the first examination of the relationship between these variables. Aligning with Lukianoff and Haidt’s assertions, we found that students’ self-reported prevalence of cognitive distortions positively predicted their endorsement of safetyism-inspired beliefs, the belief that words can harm, and support for the broad use of trigger warnings. Considering our exploratory results, we argue that greater empirical scrutiny of safetyism-inspired beliefs and practices is warranted before such customs become more widely adopted.

## 1. Introduction

In *The Coddling of the American Mind*, Lukianoff and Haidt (2018) highlighted the rise of “safetyism” on college campuses and forwarded explanatory threads of safetyism’s cultural and psychological antecedents. The authors define safetyism as a culture that treats safety – including emotional safety – as a sacred value, which results in adherents diminished willingness to sacrifice safety for other moral or practical considerations (Tetlock et al., 2000). Through a number of anecdotes, Lukianoff and Haidt highlighted a number of beliefs that they see as emerging from safetyism in universities, including the belief that *intentions don’t matter* in moral judgment, the belief that *emotional pain or discomfort is dangerous*, and the belief that *speech can be violence*. The authors argue that, in forming these beliefs, students have come to see

themselves as emotionally fragile and in need of protection from certain words, ideas, or individuals. For instance, in being taught that speech can be accurately perceived as violence (Barrett, 2017), rather than as a less extreme or intentional form of harm (Lilienfeld, 2017), students have come to endorse certain social practices (e.g., trigger warnings) that help them avoid offensive or merely counter attitudinal speech. Lukianoff and Haidt contended these beliefs, and the behavioral practices which they yield, may inadvertently hinder students’ intellectual and emotional growth by making them more anxious about, and avoidant of, minor stressors that can foster resiliency (though see Infurna & Jayawickreme, 2019 for a discussion of the limitations of resiliency and posttraumatic growth research).

One of Lukianoff and Haidt’s (2018) most controversial claims was that safetyism-inspired beliefs and behaviors stem, in part, from

<sup>☆</sup> Author note: this material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1839285 (to J.B. Celniker). The authors thank members of the Hot Cognition Lab, Melody M. Moore, Payton Jones, and Pamela Paresky for their constructive feedback on early versions of this work. The data and materials for this project are available on our OSF page: <https://osf.io/8rqcf/>.

\* Corresponding author.

E-mail address: [jcelnike@uci.edu](mailto:jcelnike@uci.edu) (J.B. Celniker).

<https://doi.org/10.1016/j.paid.2021.111243>

Received 28 April 2021; Received in revised form 30 August 2021; Accepted 31 August 2021

Available online 15 September 2021

0191-8869/© 2021 Elsevier Ltd. All rights reserved.

cognitive distortions in students' psychology. Cognitive distortions are errors in reasoning resulting from negative intuitive thoughts that are not evidence-based (Burns, 1980; Covin et al., 2011). Common cognitive distortions include catastrophizing (perceiving a mildly negative event as a disaster), all-or-nothing thinking (viewing things as either-or, rather than perceiving nuance), and emotional reasoning (believing that one's feelings accurately represent reality). Clinical psychologists identify cognitive distortions as impediments to healthy psychological functioning and have developed cognitive-behavioral techniques to help people reduce such distorted thinking (Alford & Beck, 1997). While all people engage in cognitive distortions to some degree, Lukianoff and Haidt argued that students' justifications for safetyism reflect a more pronounced pattern of distorted thinking. Throughout their book, the authors draw connections between cognitive distortions and various examples of campus events and legal cases. For instance, they argued that the amount of harm students perceive offensive speech to cause, and their justifications of censorship because of it, demonstrates cognitively distorted thinking rather than reflective analyses of the costs and benefits of allowing offensive speech. They subsequently recommended students partake in empirically-supported, cognitive-behavioral exercises with the hope that reducing students' distorted thinking will improve their mental health and diminish their endorsement of safetyism.

However, critics of Lukianoff and Haidt (2018) have suggested that students' drive to regulate speech constitutes a rational rejection of outdated conceptions of harm (Sue et al., 2019). They contend that psychologists have historically downplayed the negative impacts of unintentional and institutional maltreatment (Barrett, 2017; Sue et al., 2007). Consequently, students may be justified in increasingly regulating speech because words can be more harmful than researchers previously recognized. Thus, students' newfound focus on "impact over intent" in speech judgments may indicate a reflective shift in moral psychology rather than a shift towards more intuitive and cognitively distorted thinking. Framed against the backdrop of cultural variation in the use of perceived intentions in moral judgment (Barrett et al., 2016), this possibility cannot be discounted outright. Indeed, Lukianoff and Haidt did not provide any data to substantiate the link between distorted thinking and safetyism-inspired beliefs or discount alternative explanations. Failing to observe a positive association between students' prevalence of cognitively distorted thinking and their endorsement of safetyism-inspired beliefs would be inconsistent with the assertions forwarded by Lukianoff and Haidt; alternatively, documenting this association would be consistent with Lukianoff and Haidt's descriptive claims about students' socioemotional functioning and the psychological underpinnings of students' safetyism endorsement.

In this paper, we provide the first empirical examination of the association between college students' self-reported prevalence of cognitive distortions and their endorsement of safetyism-inspired beliefs, the belief that words can harm, and the broad use of trigger warnings. We measure additional aspects of psychological health (e.g., loneliness, resilience) as well as other constructs identified by Lukianoff and Haidt (2018) as potentially relating to safetyism-inspired beliefs (e.g., analytic thinking, intellectual humility) to explore the unique variance in the dependent variables that may be explained by the prevalence of cognitive distortions. We also explore whether safetyism-inspired beliefs and the belief that words can harm statistically mediate the relationship between cognitive distortions and the endorsement of broad use of trigger warnings. Furthermore, by collecting a large, ethnically and economically diverse sample from a public university, we provide a conservative test of the Lukianoff and Haidt's hypothesis. Lukianoff and Haidt shared anecdotes from predominately elite, White, and affluent universities to exemplify safetyism's rise, so observing an association between cognitive distortions and safetyism-inspired beliefs outside of those environments would provide preliminary support for their claims about the prevalence of distorted thinking among college-aged young adults.

## 2. Method

### 2.1. Participants

From April 2019 through June 2019, student participants were recruited through the UC Irvine Human Subjects Lab Pool to participate in an online survey for course credit. A total of 812 participants began the study. After removing the data of participants who did not finish the study, the final sample consisted of 786 participants (653 female, 127 male, 6 other/unspecified). Participants ranged in age from 18 to 48 years ( $M = 21.03$ ,  $SD = 2.87$ ), were both ethnically and economically heterogeneous (approximately 39% Asian, 32% Latino/a/x, 13% White; approximately 40% with household income \$0 to \$39,999, 28% with income \$40,000 to \$80,000, 31% with income \$80,000 or higher). More detailed demographic information is available in the Supplementary Material.

### 2.2. Procedure and measures

After consenting, participants first completed a measure of safetyism-inspired beliefs and the Trigger Warning Attitudes Assessment (TWAA; Bellet et al., 2018) in randomized order. For the safetyism-inspired beliefs measure, we piloted six items that were intended to capture the three separate but interrelated sets of beliefs we previously outlined. Thus, two items measured the belief that *intentions don't matter* ("If I feel offended or oppressed by the actions of another person, then that person is guilty of an act of bigotry, no matter if they intended to offend me or not," and "Intentions don't matter; only the emotional impact of those words on the listener matters"), two items measured the belief that *emotional pain or discomfort is dangerous* ("Emotional pain is just as dangerous as physical pain," and "Emotional pain or discomfort is a sign that one is in danger"), and two items measured the belief that *speech can be violence* ("People expressing offensive political views are causing violence against those they offend," and "Offensive speech can be seen as an act of violence towards vulnerable groups"). Participants reported how much they agreed or disagreed with each of these statements on 7-point scales (1 = *Strongly disagree*, 4 = *Neither agree nor disagree*, 7 = *Strongly agree*). The TWAA provided participants with a brief definition of trigger warnings and then asked, "Do you think that trigger warnings should be used?" (0 = *No*, 1 = *Yes*). If participants selected "Yes" on this first item, they were then asked, "Why do you think that trigger warnings should be used?" and provided a list of reasons for supporting the use of trigger warnings (e.g., "It's not fair that vulnerable people, such as those with posttraumatic stress disorder, should be exposed to material that causes them distress without a warning," "Offensive material can cause psychological harm to anyone"). Participants were able to select as many of the six listed reasons as they liked, and an "Other" option with an open-ended text response was also included.

All of the remaining psychological measures were presented in randomized order. The Cognitive Distortions Scale (CDS, Covin et al., 2011) provides descriptions and examples of ten cognitive distortions: mind-reading, catastrophizing, all-or-nothing thinking, emotional reasoning, labeling, mental filtering, overgeneralization, personalization, should statements, and minimizing the positive. The CDS has been shown to be a valid and reliable measure of cognitive distortions in both clinical and non-clinical samples (Özdel et al., 2014) that does not correlate with demographic variables or social desirability (Covin et al., 2011). Participants are given two vivid examples of each distortion, one describing a social situation and one describing an achievement-related situation in a school or workplace context, and then report how frequently they

engage in each type of thinking (1 = *Never*, 4 = *Sometimes*, 7 = *All the time*).<sup>1</sup> For example, emotional reasoning is illustrated with the following examples:

People can believe something to be true because it “feels” that way. To illustrate, please read the following passages:

**A.** Kim’s friends told her that she could not come to the concert with them because they were unable to get enough tickets for everyone. Kim knows they probably didn’t exclude her on purpose, but she *feels rejected*. Therefore, part of her believes she was rejected.

**B.** Ted’s boss told him that his performance at the company has been good. Yet, Ted wonders if he could have done better. In fact, he *feels like a failure*. Consequently, he starts to believe he is a failure.

Participants indicated how often they “engage in Emotional Reasoning when in social situations (like when you’re with friends, partners or family)” and how often they “engage in Emotional Reasoning when in achievement situations (such as school or work).”

The Words-Can-Harm Scale (WCHS, Bellet et al., 2018) consisted of 10 items that capture the extent to which respondents believe that hurtful or offensive language can cause serious damage to themselves or others (e.g., “A person might develop posttraumatic stress disorder or at least some of its symptoms from something they read”). Each of these items was assessed on a 100-point scale (1 = *strongly disagree*, 100 = *strongly agree*). Conflict avoidance was measured using the 5-item version of the Conflict Approach/Avoidance subscale from Goldstein’s (1999) Conflict Communication Scale (as used in Mutz & Reeves, 2005) using 7-point scales (1 = *Strongly disagree*, 4 = *Neither agree nor disagree*, 7 = *Strongly agree*). Loneliness was measured with the 8-item UCLA Loneliness Scale (Hays & Dimatteo, 1987), resilience was measured with the Brief Resilience Scale (Smith et al., 2008), and intellectual humility was measured with the Intellectual Humility Scale (Leary et al., 2017). Analytic (versus intuitive) thinking was measured using the combined scores from the Cognitive Reflection Test (CRT, Frederick, 2005) and CRT-2 (Thomson & Oppenheimer, 2016). Finally, participants provided demographic information (e.g., age, sex) before concluding the study.

### 3. Results

#### 3.1. Descriptive and bivariate analyses

As in past research using the TWAA (Bellet et al., 2018; Bellet et al., 2020), an overwhelming majority of participants (94.3%,  $n = 739$ ) indicated support for the use of trigger warnings in our sample. Of these individuals, 87.5% ( $n = 688$ ) believed that trigger warnings help protect vulnerable populations (e.g., those with posttraumatic stress disorder), 66.8% ( $n = 525$ ) believed that trigger warnings help protect minority populations (e.g., women and ethnic minorities), and 59.7% ( $n = 469$ ) believed that trigger warnings help protect all individuals. We created a composite measure of trigger warning endorsement by summing the number of reasons participants selected for supporting trigger warnings on the TWAA (0 = did not endorse the use of trigger warnings, 6 = endorsed all of the provided reasons for supporting the use of trigger warnings). This TWAA composite measure displayed suitable internal reliability. Additionally, the six safetyism-inspired belief items displayed adequate internal reliability and were thus averaged into a composite measure. Table 1 presents the descriptive statistics, internal reliabilities, and bivariate correlations of the measured variables.

Supportive of theorizing by Lukianoff and Haidt (2018), higher CDS scores were positively correlated with endorsement of the safetyism-inspired belief items. In other words, students who reported more

<sup>1</sup> We used participants’ total CDS scores, rather than separating into the social and achievement subscales, given the high level of correlation between the subscales in these data ( $r = 0.81$ ).

**Table 1**  
Descriptive statistics and bivariate correlations.

Dependent variable	Mean (SD)	Internal reliability (α)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Cognitive distortions	4.29 (1.02)	0.91	–													
2. Safetyism-inspired beliefs	4.72 (0.91)	0.73	0.25***	–												
3. Words can harm	57.94 (19.05)	0.89	0.30***	0.39***	–											
4. TWAA composite	3.92 (2.06)	0.84	0.12***	0.31***	0.35***	–										
5. Conflict avoidance	4.69 (1.09)	0.75	0.13***	0.18***	0.20***	0.22***	–									
6. Loneliness	2.19 (0.61)	0.86	0.47***	0.09*	0.13***	0.00	0.01	–								
7. Resilience	3.03 (0.77)	0.85	–0.45***	–0.22***	–0.25***	–0.11**	–0.25***	–0.46***	–							
8. Analytic thinking	3.18 (2.04)	0.75	–0.13***	–0.15***	–0.05	–0.08*	–0.08*	–0.02	0.09*	–						
9. Intellectual humility	3.79 (0.69)	0.85	0.02	0.07*	0.18***	0.15***	–0.00	–0.04	0.13***	0.07*	–					
10. Political orientation (social)	5.41 (1.39)	–	0.09*	0.28***	0.18***	0.18***	0.10**	0.07	–0.13**	0.07	0.18***	–				
11. Political orientation (economic)	4.98 (1.43)	–	0.06	0.30***	0.19***	0.12**	0.05	0.07	–0.12**	–0.03	0.11**	0.71***	–			
12. Household income	3.51 (2.09)	–	–0.05	–0.09*	0.02	–0.01	–0.04	–0.09*	0.08*	0.09**	0.05	–0.01	–0.11**	–		
13. Age	21.03 (2.87)	–	–0.08*	–0.08*	–0.14***	–0.10**	–0.00	–0.03	0.08*	–0.10**	0.02	–0.06	–0.03	–0.14***	–	
14. Sex	0.16 (0.37)	–	–0.15***	–0.15***	–0.13***	–0.11**	–0.26***	–0.02	0.15***	0.07	0.02	–0.15***	–0.11**	–0.02	0.05	–

Note: The political orientation items are coded 1 = *Very Conservative*, 7 = *Very Liberal*. Household income is coded in \$20,000 increments from 1 = \$0–\$19,999, 7 = \$120,000 or higher. Sex is coded 0 = *Female*, 1 = *Male*.  
\*  $p < .05$ .  
\*\*  $p < .01$ .  
\*\*\*  $p < .001$

frequent cognitive distortions generally reported stronger safetyism-inspired beliefs. Given Lukianoff and Haidt’s focus on emotional reasoning in their book, it is worth noting that, when looking at each of the ten cognitive distortions on the CDS separately, the cognitive distortion of *emotional reasoning* had the strongest correlation with safetyism-inspired beliefs ( $r = 0.24, p < .001$ ; the full set of these results is presented in the Supplementary Material). Additionally, CDS scores were positively associated with WCHS and the TWAA composite measure, indicating that those who reported more frequent cognitive distortions tended to have a stronger belief that words can harm and endorsed more reasons for using trigger warnings. The safetyism-inspired belief measure, WCHS, and TWAA composite were all positively associated with one another, and all three of these measures were positively associated with conflict avoidance, intellectual humility, both social and economic liberalism, and being female (compared to male). Safetyism-inspired belief, WCHS, and TWAA composite scores were each negatively associated with resilience and age.

### 3.2. Regression analysis

To further assess the robustness of these associations and control for shared variance, we constructed multiple linear regression models. We composed two separate models that used the safetyism-inspired belief measure and the WCHS measure as outcome variables, respectively.<sup>2</sup> In each of these models, the outcome measures were regressed on CDS, conflict avoidance, loneliness, resilience, analytic thinking, intellectual humility, social and economic political orientation, household income, age, sex (dummy-coded), and ethnicity (dummy-coded).<sup>3</sup>

Starting with the model predicting safetyism-inspired beliefs, CDS scores significantly predicted higher safetyism-inspired belief scores,  $b = 0.15, 95\% \text{ CI} = [0.07, 0.23], p < .001$ . In other words, reporting more frequent experiences of cognitive distortions was predictive of greater safetyism-inspired beliefs. The other positive predictors of safetyism-inspired beliefs were conflict avoidance,  $b = 0.09, 95\% \text{ CI} = [0.02, 0.16], p = .008$ , social liberalism,  $b = 0.08, 95\% \text{ CI} = [0.01, 0.15], p = .026$ , and economic liberalism,  $b = 0.11, 95\% \text{ CI} = [0.04, 0.18], p = .002$ . The negative predictors of safetyism-inspired beliefs were resilience,  $b = -0.14, 95\% \text{ CI} = [-0.25, -0.03], p = .012$ , and analytic thinking,  $b = -0.07, 95\% \text{ CI} = [-0.10, -0.03], p < .001$ .

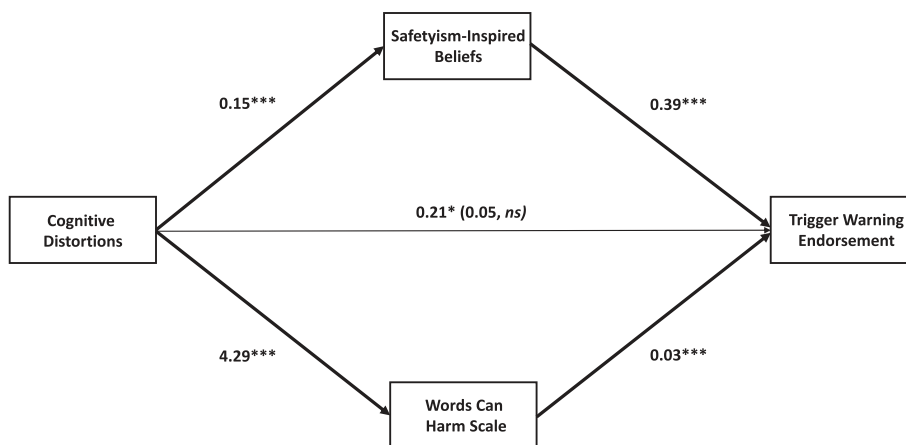
Higher CDS scores were also positively predictive of WCHS scores,  $b$

$= 4.29, 95\% \text{ CI} = [2.71, 5.86], p < .001$ , meaning that reporting more cognitive distortions was associated with a stronger belief that words can harm. Conflict avoidance was another positive predictor of WCHS scores,  $b = 2.14, 95\% \text{ CI} = [0.82, 3.46], p = .002$ , as was intellectual humility,  $b = 5.88, 95\% \text{ CI} = [3.85, 7.92], p < .001$ . Resilience was negatively associated with WCHS scores,  $b = -2.35, 95\% \text{ CI} = [-4.51, -0.19], p = .033$ , and age was negatively predicted of WCHS scores as well,  $b = -0.62, 95\% \text{ CI} = [-1.08, -0.16], p = .009$ . In sum, while cognitive distortions, conflict avoidance, and resilience were significant predictors across outcome variables, we also found unique predictors of safetyism-inspired beliefs (social liberalism, economic liberalism, and analytic thinking) and WCHS scores (intellectual humility and age).

### 3.3. Mediation analyses

Finally, we conducted exploratory mediation analyses to model one of the causal chains proposed by Lukianoff and Haidt: cognitive distortions in students’ thinking lead them to hold beliefs that influence their endorsement of policies like trigger warnings. In our models, the CDS measure was the predictor variable, the safetyism-inspired belief measure and WCHS were separate mediators, and the TWAA composite measure was the outcome variable. We modelled both the safetyism-inspired belief measure and WCHS as mediators to assess whether our pilot measure provided incremental validity in explaining TWAA endorsement above and beyond the WCHS. Using the jAMM package in jamovi, we constructed mediation models with and without including conflict avoidance, loneliness, resilience, analytic thinking, intellectual humility, social and economic political orientation, household income, age, sex, and ethnicity as covariates. Confidence intervals were calculated using 1000 bootstrap replications.

Across model specifications, safetyism-inspired beliefs and WCHS scores mediated the association between CDS scores and the TWAA composite measure. In the model without covariates, the indirect effect through safetyism-inspired beliefs was significant,  $b = 0.10, 95\% \text{ CI} = [0.06, 0.15], p < .001$ , as was the indirect effect through belief that words can harm,  $b = 0.17, 95\% \text{ CI} = [0.11, 0.23], p < .001$ . In the model including covariates, the indirect effect through safetyism-inspired beliefs,  $b = 0.06, 95\% \text{ CI} = [0.02, 0.10], p = .005$ , and the indirect effect through the WCHS,  $b = 0.11, 95\% \text{ CI} = [0.05, 0.17], p < .001$ , remained significant. The key path estimates for the model including covariates



**Fig. 1.** Mediation model showing the association between CDS scores and the TWAA composite measure, as mediated by the safetyism-inspired beliefs measure and the WCHS. The presented path estimates control for conflict avoidance, loneliness, resilience, analytic thinking, intellectual humility, social and economic political orientation, household income, age, sex, and ethnicity. Unstandardized coefficients are displayed. On the center path, the coefficient outside the parentheses is the total effect, and the coefficient inside the parentheses is the direct effect. Asterisks indicate significant paths (\*  $p < .05$ , \*\*\*  $p < .001$ ).

<sup>2</sup> Please see the Supplementary Materials for full model statistics and an additional models with the TWAA composite and resilience as the outcome variables.

<sup>3</sup> Missing data in the demographic items, particularly political orientation, reduced the degrees of freedom in the regression analyses to 577.



#### 4. Discussion

This is the first empirical study to explore the association between college students' prevalence of cognitive distortions and their endorsement of safetyism-inspired beliefs and practices. Supportive of theorizing by Lukianoff and Haidt (2018), we found that cognitive distortions and safetyism-inspired beliefs were positively correlated in a large, economically and ethnically diverse, public university sample. The association between cognitive distortions and safetyism-inspired beliefs remained significant when accounting for other relevant psychological and demographic variables, like resiliency and analytic thinking. Furthermore, this study revealed that cognitive distortions were a robust predictor of students' belief that words can harm and the number of reasons they selected for endorsing the use of trigger warnings. Finally, we found that safetyism-inspired beliefs had predictive utility beyond the belief that words can harm in explaining the statistical relationship between cognitive distortions and the number of reasons selected for endorsing trigger warnings. Overall, these results support Lukianoff and Haidt's claim that distorted reasoning is related to particular safetyism-inspired beliefs and practices that they highlighted in *The Coddling of the American Mind*.

While there is merit in adapting our definitions of harm to encapsulate experiences that may be invisible to dominant cultural groups (Sue et al., 2019), we must also be cautious of adverse effects that may result from expanding our conceptions of violence, prejudice, and trauma too broadly (Haslam, 2016). Overextending these concepts may inadvertently engender "looping effects" whereby students come to interpret actions that they would have otherwise deemed minimally harmful as more aversive (Haslam, 2016; Lilienfeld, 2017). For instance, Asian American students who interpret ambiguous situations as evidence of racial discrimination experience more anxiety than those who do not perceive discrimination in the same situations (e.g., being asked "Where are you from?" in an airport food court; Wang et al., 2011). Such emotional reasoning may, in turn, cause students to endorse safetyism practices that reinforce the notion that speech can cause irrevocable harm (Lukianoff & Haidt, 2018). This purported cycle, which begins with conceptual expansions that can yield positive outcomes, may unintentionally condition patterns of distorted thinking that undermine students' socioemotional functioning. Nonetheless, despite establishing an association between cognitive distortions and safetyism-inspired beliefs, our data cannot speak to the causal structure of this relationship. While these variables may causally influence one another, they may merely cooccur as a function of other causal forces. It is also possible that cognitive distortions are associated with extreme beliefs in general rather than safetyism-inspired beliefs specifically. Future research should assess the boundaries and causal mechanisms of the reported associations in more politically diverse student populations. Furthermore, while the safetyism-inspired beliefs measure created for this study demonstrated acceptable psychometric properties, future investigations would benefit from the development of more precise measures of safetyism-inspired beliefs and direct measurement of safety as a sacred value.

Importantly, this study does not address whether safetyism-inspired beliefs or practices may be functional in some environments. For instance, although trigger warnings appear to be unhelpful for reducing anxiety when one is exposed to distressing content (Bellet et al., 2020; Bridgland et al., 2019; Jones et al., 2020; though see Gainsburg & Earl, 2018 for more mixed results), limited research has examined whether such warnings are effective in helping people avoid distressing content. Although content warnings can increase the attractiveness of negative content to youth and young adults (Bijvank et al., 2009), and avoidance coping (but not approach coping) is generally associated with psychological distress (Littleton et al., 2007), more research is needed to understand when, and for whom, trigger warnings may yield healthy avoidance behaviors (Boysen, 2017). Practices like trigger warnings may be beneficial for some populations in some contexts (e.g., those

with PTSD avoiding trauma-related content) yet be unhelpful or counterproductive at other times or for other populations (e.g., healthy college students avoiding ideologically-inconsistent content). Indeed, researchers must assess the generalizability of the present effects in various populations and behavioral contexts before condemnations of safetyism writ large can be empirically justified.

Furthermore, our data cannot resolve normative disagreements regarding speech regulation on college campuses. While the present results suggest that students' endorsement of safetyism-inspired practices relates to cognitively distorted thinking, this does not necessarily bear on the validity of normative justifications for increasing speech regulations. However, we did find that safetyism-inspired beliefs were negatively associated with resiliency and analytic thinking. Given that two core aims of higher education are to foster resiliency and analytic thinking, our results suggest that there may be tension between fulfilling those missions and satisfying students' desires for emotional safety. More broadly, if university stakeholders aspire to develop campus cultures and evidence-based policies that better prepare students for the conflict-ridden world they inhabit, then they must be more willing to scrutinize the psychological antecedents of safetyism-inspired beliefs and the consequences of safetyism-inspired practices. According with this goal, our exploratory findings lend credence to those who are skeptical of safetyism and its inspired beliefs as unmitigated goods for college students' socioemotional development.

#### CRedit authorship contribution statement

**Jared B. Celniker:** Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Megan M. Ringel:** Conceptualization, Methodology, Investigation, Writing – review & editing. **Karli Nelson:** Conceptualization, Methodology, Investigation, Writing – review & editing. **Peter H. Ditto:** Writing – review & editing, Supervision.

#### Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.paid.2021.111243>.

#### References

- Alford, B. A., & Beck, A. T. (1997). *The integrative power of cognitive therapy*. New York, NY: Guilford Press.
- Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., ... Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 113(17), 4688–4693. <https://doi.org/10.1073/pnas.1522070113>.
- Barrett, L. F. (2017, July 14). When is speech violence?. *The New York Times* [https://www.nytimes.com/2017/07/14/opinion/sunday/when-is-speech-violence.html?\\_r=0](https://www.nytimes.com/2017/07/14/opinion/sunday/when-is-speech-violence.html?_r=0).
- Bellet, B. W., Jones, P. J., & McNally, R. J. (2018). Trigger warning: Empirical evidence ahead. *Journal of Behavior Therapy and Experimental Psychiatry*, 61(April), 134–141. <https://doi.org/10.1016/j.jbtep.2018.07.002>.
- Bellet, B. W., Jones, P. J., Meyersburg, C. A., Brenneman, M. M., Morehead, K. E., & McNally, R. J. (2020). Trigger warnings and resilience in college students: A preregistered replication and extension. *Journal of Experimental Psychology: Applied*. <https://doi.org/10.1037/xap0000270>. Advance online publication.
- Bijvank, M. N., Konijn, E. A., Bushman, B. J., & Roelofsma, P. H. M. P. (2009). Age and violent-content labels make video games forbidden fruits for youth. *Pediatrics*, 123(3), 870–876. <https://doi.org/10.1542/peds.2008-0601>.
- Boysen, G. A. (2017). Evidence-based answers to questions about trigger warnings for clinically-based distress: A review for teachers. *Scholarship of Teaching and Learning in Psychology*, 3(2), 163–177. <https://doi.org/10.1037/stl0000084>.
- 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>.
- Burns, D. D. (1980). *Feeling good: The new mood therapy*. New York, NY: Avon Books.
- Covin, R., Dozois, D. J. A., Ogniewicz, A., & Seeds, P. M. (2011). Measuring cognitive error: Initial development of the Cognitive Distortions Scale (CDS). *International Journal of Cognitive Therapy*, 4(3), 297–322.

- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4), 25–42.
- 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(August), 252–263. <https://doi.org/10.1016/j.jesp.2018.08.006>.
- Goldstein, S. B. (1999). Construction and validation of a conflict communication scale. *Journal of Applied Social Psychology*, 29(9), 1803–1832. <https://doi.org/10.1111/j.1559-1816.1999.tb00153.x>.
- Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, 27(1), 1–17. <https://doi.org/10.1080/1047840X.2016.1082418>.
- Hays, R. D., & Dimatteo, M. R. (1987). A short-form measure of loneliness. *Journal of Personality Assessment*, 51(1), 69–81. [https://doi.org/10.1207/s15327752jpa5101\\_6](https://doi.org/10.1207/s15327752jpa5101_6).
- Infurna, F. J., & Jayawickreme, E. (2019). Fixing the growth illusion: New directions for research in resilience and posttraumatic growth. *Current Directions in Psychological Science*, 28(2), 152–158. <https://doi.org/10.1177/0963721419827017>.
- Jones, P. J., Bellet, B. W., & McNally, R. J. (2020). Helping or harming? The effect of trigger warnings on individuals with trauma histories. *Clinical Psychological Science*, 8(5), 905–917. <https://doi.org/10.1177/2167702620921341>.
- Leary, M. R., Diebels, K. J., Davison, E. K., Isherwood, J. C., Jongman-Sereno, K. P., Raimi, K. T., ... Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality & Social Psychology Bulletin*. <https://doi.org/10.1177/0146167217697695>.
- Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on Psychological Science*, 12(1), 138–169. <https://doi.org/10.1177/1745691616659391>.
- Littleton, H., Horsley, S., John, S., & Nelson, D. V. (2007). Trauma coping strategies and psychological distress: A meta-analysis. *Journal of Traumatic Stress*, 20(6), 977–988. <https://doi.org/10.1002/jts>.
- Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. New York: Penguin Books.
- Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, 99(1), 1–15.
- Özdel, K., Taymur, I., Guriz, S. O., Tulaci, R. G., Kuru, E., & Turkcapar, M. H. (2014). Measuring cognitive errors using the Cognitive Distortions Scale (CDS): Psychometric properties in clinical and non-clinical samples. *PLoS One*, 9(8). <https://doi.org/10.1371/journal.pone.0105956>.
- Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine*, 15(3), 194–200. <https://doi.org/10.1080/10705500802222972>.
- Sue, D. W., Alsaidi, S., Awad, M. N., Glaeser, E., Calle, C. Z., & Mendez, N. (2019). Disarming racial microaggressions: Microintervention strategies for targets, white allies, and bystanders. *American Psychologist*, 74(1), 128–142. <https://doi.org/10.1037/amp0000296>.
- Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, 62(4), 271–286. <https://doi.org/10.1037/0003-066X.62.4.271>.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853–870. <https://doi.org/10.1037/0022-3514.78.5.853>.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11(1), 99–113.
- Wang, J., Leu, J., & Shoda, Y. (2011). When the seemingly innocuous “stings”: Racial microaggressions and their emotional consequences. *Personality and Social Psychology Bulletin*, 37(12), 1666–1678. <https://doi.org/10.1177/0146167211416130>.