AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Psychopathology and Clinical Science

© 2025 American Psychological Association
ISSN: 2769-7541

https://doi.org/10.1037/abn0000989

Past Suicide Attempt Is Associated With a Weaker Decision-Making Bias to Actively Escape From Suicide-Related Stimuli

Adam C. Jaroszewski[1, 2], Alexander J. Millner[3, 4], Samuel J. Gershman[3], Peter J. Franz[5, 6], Kate H. Bentley[1, 2], Evan M. Kleiman[7], and Matthew K. Nock[2, 3, 4]

[1] Department of Psychiatry, Harvard Medical School

[2] Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, United States

[3] Department of Psychology, Harvard University

[4] Mental Health Research, Franciscan Children's, Brighton, Massachusetts, United States

[5] Ferkauf Graduate School of Psychology, Yeshiva University

[6] Psychiatry Research Institute at Montefiore Einstein (PRIME), Albert Einstein College of Medicine/Montefiore Medical Center, Bronx, New York, United States

[7] Department of Psychology, Rutgers, The State University of New Jersey

Theory and evidence suggest that people attempt suicide to escape acute distress. However, little is known about why people select suicide instead of other ways to escape (e.g., alcohol/drug use). One possibility is that suicide-related stimuli in one's environment (e.g., suicide methods) bias this decision, particularly when such stimuli elicit little aversion. We tested whether suicide-related stimuli bias decisions to escape acute distress. We recruited 360 adults with past 3-month active suicidal thoughts and behaviors (STB; $n = 120$), elevated psychiatric symptoms without STB ($n = 152$), or no symptoms/STB ($n = 88$). Participants explicitly rated personalized suicide pictures (e.g., pointing a gun up at oneself) and positive contrasts and completed a behavioral task, where they made decisions to escape an acutely distressing noise in relation to these stimuli. We used a computational model of task performance to capture latent biases hypothetically influencing decision making. We assessed STB 3 months later. Results indicated that people with a past suicide attempt exhibited much lower suicide aversion than others. In the behavioral task, the suicidal group made more impulsive escape decisions in relation to suicide versus positive stimuli. The computational model helped explain this effect, capturing a latent bias driven by the suicide stimuli. Within the suicidal group, weaker biases mediated the association between lower suicide aversion and higher odds of past suicide attempt. These results provide evidence of novel, specific, incrementally valid, and objectively assessed suicide-attempt correlate and suggest that decision science is useful for understanding mechanisms increasing risk for suicide and other escape-related phenomena involving stimulus-driven processes (e.g., substance misuse, and anxiety).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Arielle Baskin-Sommers served as action editor.

Adam C. Jaroszewski https://orcid.org/0000-0003-4163-1741

Adam C. Jaroszewski was supported by grants from the National Institute of Mental Health (NIMH) while conducting this study and preparing the article (F31MH116649 and K23MH133876). The sponsor did not have a role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The contents are solely the responsibility of the authors and do not necessarily represent the views of the NIMH. Matthew K. Nock receives publication royalties from Macmillan, Pearson, and UpToDate. He has been a paid consultant in the past 3 years for Cambridge Health Alliance and for legal cases regarding a death by suicide. He has stock options in Cerebral, Inc. He is an unpaid scientific advisor for Empatica, Koko, TalkLife, and the JED Foundation.

Hypotheses and primary analyses were preregistered through the Center for Open Science (Jaroszewski, 2023). All study methods were approved by the Committee for Use of Human Subjects at Harvard University (Protocol: IRB19-0017). Data are not available because participants did not consent to data sharing. Computational and analytic code are available upon request. Findings related to group differences in explicit ratings of suicide pictures and computational model bias parameters as well as the related mediation effect were presented as part of conference symposia. No part of this article was written by artificial intelligence.

Adam C. Jaroszewski served as lead for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, visualization, writing–original draft, and writing–review and editing. Alexander J. Millner served as lead for software and served in a supporting role for conceptualization, formal analysis, methodology, supervision, and writing–review and editing. Samuel J. Gershman served in a supporting role for formal analysis, methodology, supervision, and writing–review and editing. Peter J. Franz served in a supporting role for conceptualization and writing–review and editing. Kate H. Bentley served in a supporting role for supervision and writing–review and editing. Evan M. Kleiman served in a supporting role for supervision and writing–review and editing. Matthew K. Nock served as lead for supervision and served in a supporting role for conceptualization, funding acquisition, resources, and writing–review and editing.

The preregistered design is available at https://osf.io/eap6x/?view_only=72152e5a01814c0ca58e5770e8ddce56.

Correspondence concerning this article should be addressed to Adam C. Jaroszewski, Department of Psychiatry, Massachusetts General Hospital, Simches Research Building, 185 Cambridge Street, Suite 2000, Boston, MA 02114, United States. Email: ajaroszewski@mgh.harvard.edu

General Scientific Summary

Research suggests that the primary reason people attempt suicide is to escape intense distress. Prior studies show that individuals with versus without a history of suicidal thoughts or behaviors tend to make different decisions when trying to escape distress. This study builds on prior findings by suggesting that, when trying to escape, suicide-related content influences decision making in those with past suicidal behavior differently than those with only suicidal thoughts. These differences might make people with past suicidal behavior less likely to actively limit their exposure to suicidal thoughts or methods during attempts to escape intense distress.

Keywords: suicide, behavioral task, computational model, reinforcement learning, decision making

Supplemental materials: https://doi.org/10.1037/abn0000989.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Prominent theories suggest that the primary reason people attempt suicide is to escape from acutely distressing states, such as intolerable emotional pain (e.g., Baumeister, 1990; Joiner, 2005; Shneidman, 1987). Empirical evidence also supports this idea (Bryan et al., 2013; Kidd, 2004; O'Brien et al., 2021; Nock et al., 2025). Yet little is known about the factors and processes influencing people to select suicide as a means of escape instead of alternative options. Research indicates that decision-making abnormalities may play a role (Dombrovski & Hallquist, 2022). For instance, people with prior suicidal thoughts and behavior (STB) make "suboptimal" decisions when they are required to learn or estimate the value of options to maximize reward over the long run (Dombrovski et al., 2010, 2011, 2019; Jollant et al., 2005, 2010; Millner et al., 2019). However, it is not known whether such value-learning deficits are directly related to the decision to attempt suicide specifically or, rather, decisions that increase the likelihood of attempting suicide by increasing the intensity and/or frequency of experiencing acute distress (i.e., via stress generation; Liu & Spirito, 2019), or both. Thus, although aberrant decision making appears to be associated with STB, the specificity of this relationship not yet well understood. With tens of thousands of Americans and nearly one million people worldwide dying by suicide each year (Naghavi, 2019), it is imperative that we uncover psychological processes leading certain people to engage in suicidal behavior. Understanding whether reduced aversion to suicide plays a role in the decision to attempt suicide could provide new targets to improve prediction and prevention of this tragic behavior (Jaroszewski et al., 2022).

Decades of social science theory and evidence indicate that people tend to select options that they perceive to be more positive and/or less aversive than available alternatives (Hofmann et al., 2010). Recent research has shown that people with prior STB perceive suicide-related stimuli (e.g., a picture of a person holding a gun to their head) to be less aversive than do people without prior STB, and that lower suicide aversion is associated with more recent and severe STB (Jaroszewski et al., 2020). Relatedly, reduced aversion to nonsuicidal self-injury (NSSI) plays a critical role in the maintenance of NSSI (Franklin et al., 2016; Nock & Banaji, 2007). Given that NSSI is strongly associated with STB (Franklin et al., 2017), it is possible that reduced suicide aversion could play a role in future STB; however, no research has directly investigated this possibility.

Perceiving suicide as less aversive may increase the risk of deciding to attempt suicide via a "stimulus-driven" process. Some decision-making mechanisms, such as Pavlovian control," rigidly bias animals toward selecting certain actions based on the valence of stimuli present in the environment (Guitart-Masip et al., 2014). For instance, positive stimuli typically elicit a bias to approach, whereas negative stimuli elicit a bias to avoid (Duckworth et al., 2002; Solarz, 1960). Most people regard suicide as extremely negative (Jaroszewski et al., 2020; Nazem et al., 2017) and, therefore, likely possess a strong bias to actively avoid suicide and related stimuli. However, for people with less suicide aversion (e.g., people with recent and frequent STB), suicide stimuli may elicit less bias to avoid suicide. This possibility is consistent with a recent theory of self-injury proposing that reduced aversion to self-injury stimuli reduces the motivation to avoid engaging in NSSI (Hooley & Franklin, 2018). Also, less bias to avoid suicide stimuli may make it more likely to consider or even simulate suicide as an option to escape distress (Dombrovski & Hallquist, 2022). Another reason people with lower suicide aversion might be at increased risk of attempting suicide is due to possessing a greater bias to approach suicide. Given that the primary reason people attempt suicide is to escape/gain relief, suicide stimuli, such as potential suicide methods in one's environment, may also represent or become associated with relief (Coppersmith et al., 2023; Kleiman et al., 2018), and therefore elicit motivation to approach suicide methods or simulate/imagine suicidal behavior. Indeed, qualitative and clinical data indicate that some people report feeling "drawn" or "pushed" toward attempting suicide after imagining their own suicidal behavior and/or seeing potential suicide methods in their environment (Crane et al., 2012; Holmes et al., 2007; O'Brien et al., 2021). Thus, there are at least two different, nonmutually exclusive stimulus-driven processes through which lower suicide aversion could increase the risk of attempting suicide: (a) Suicide-related stimuli might elicit less motivation to avoid suicide and/or (b) more motivation to approach it. Neither possibility has been tested.

In addition to Pavlovian control, "instrumental control" is an action selection mechanism that influences behavior based on prior actions that maximized beneficial outcomes (Guitart-Masip et al., 2014). Pavlovian and instrumental control systems typically favor the same behavioral choices (e.g., approach reward, avoid punishment), making it difficult to tease apart how each system ascribes value to actions. However, behavioral experiments that manipulate these two systems to be either congruent or incongruent with one another can reveal how stimulus-driven, Pavlovian control can either benefit or interfere with instrumental control (e.g., Guitart-Masip et al., 2012; Lindström et al., 2015; Millner et al., 2018). For example, prior studies have exposed participants to an acutely negative stimulus (i.e., an aversive noise that sounds like nails on a

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

chalkboard) and given them a choice between two response options, an active ("go") versus passive ("no-go") means of escaping the noise/gaining relief. Results indicated that the aversive noise consistently biased participants to do something active to escape instead of something passive (Millner et al., 2018)—that is, they consistently favored active escape (go) even when the passive option (no-go) was required to gain relief. Moreover, in a follow-up study, people with STB were more biased than psychiatric controls to escape via active versus passive means (Millner et al., 2019). Importantly, because instrumental control would readily select the passive response when doing so maximized escape/relief, the presence of a consistent bias for active escape when the required instrumental response was passive is an example of incongruence between Pavlovian and instrumental control systems, with Pavlovian control consistently ascribing value to the active response regardless of its instrumental value, presumably due to the negative valence of the noise stimulus eliciting an automatic, involuntary active responding. Although the term "Pavlovian" typically denotes instances where a previously neutral cue comes to elicit automatic, involuntary responding through learned stimulus–outcome associations (classical conditioning), here we use the term "Pavlovian" to refer to the theoretical system that mediates stimulus-driven, relatively automatic behavioral effects regardless of whether such effects are due to ontogenetic (e.g., associative learning) and/or phylogenetic (innate/inherited) factors. For instance, a bias to actively avoid a seldomly experienced aversive noise (Millner et al., 2018, 2019) may be more likely innate than learned, whereas a bias to actively avoid a gun may be more likely learned than innate, but, regardless, both such biases can be modeled as "Pavlovian" since they reflect relatively automatic, involuntary responding elicited by stimuli per se (Lindström et al., 2015).

Computational approaches can isolate the effect of a valenced stimulus on instrumental learning by modeling behavioral task performance (accuracy, response time [RT]) as the product of a Pavlovian–instrumental interaction (Guitart-Masip et al., 2012; Lindström et al., 2015; Millner et al., 2018, 2019). Experiments testing Pavlovian–instrumental interactions have been used to study escape-related biases among people with and without STB, but no study has investigated whether suicide stimuli per se bias decisions to escape aversive states. This may be especially relevant to better understanding, predicting, and preventing STB because people decide to attempt suicide while in the presence of suicide-related stimuli (e.g., available suicide methods), and some even report being "pushed" to attempt suicide upon encountering such stimuli. Therefore, a bias driven by suicide-related stimuli may influence those with low suicide aversion to attempt suicide. If so, this would represent a novel clinical target potentially mediated by a relatively well-understood, basic psychological process (i.e., Pavlovian–instrumental interaction).

Here we test whether suicide-related stimuli bias how people with recent STB decide to escape a distressing context. Building on prior work (Millner et al., 2018, 2019), we assessed the impact of personalized suicide pictures (e.g., looking down the barrel of a handgun that is being held in one's hand and pointed up at oneself) versus positive contrasts (e.g., looking down at a piece of cake being held in one's hand pointed up at oneself) on behavior (choice, RT) using a task where an aversive noise was used to create an acutely distressing context and correct choices terminated the noise, thereby providing escape/relief. To terminate the noise, participants chose between doing something active (i.e., pressing the spacebar on a keyboard, "go") or doing something passive (not pressing, "no-go"). Participants included people with recent (past 3 months) STB (suicidal group; $n = 120$), people with elevated recent-psychiatric symptoms but no STB (psychiatric group; $n = 152$), and people with neither STB nor elevated psychiatric symptoms (healthy group; $n = 88$). Including a psychiatric group helps determine which differences are specific to people with STB and not merely related elevated psychiatric symptoms. Like the prior escape-learning studies (Millner et al., 2018, 2019), we used a computational model based on task performance (choice, RT) to capture latent biases influencing behavior, including a hypothetical bias elicited by the suicide stimuli themselves (i.e., the stimulus-driven effect). Also, we assessed participants on the incidence of STB 3 months after baseline, allowing us to examine prospective associations between STB and picture ratings as well as computational model parameters. We hypothesized that the suicide-related (vs. positive) pictures would "push" or bias the suicidal group to escape the distressing context via active (go) versus passive means (no-go) relative to the control groups. Also, we expected that the computational model would reveal that, compared to both control groups, the suicidal group had a stronger latent bias elicited by the suicide pictures, explaining their hypothesized behavioral bias for active escape in the presence of suicide pictures.

Method

Procedure

Participants ($N = 360$) were adults (21–65 years old) recruited from a sociodemographically diverse university study pool from the greater Boston area, social media platforms (e.g., Facebook, Instagram), participant-initiated web searches (e.g., google), online study/crowdsourcing platforms (Amazon Mechanical Turk [MTurk], Prolific), and message boards (e.g., Reddit, Craigslist) from June 2019 through October 2020. Enrollees responded to an advertisement for a study on "thoughts, feelings, and decisions," completed an informed consent, which included multiple-choice questions to verify comprehension of study procedures. Participants completed the study online. There were two time points: (a) a brief ($M = 8.9$ min, $Mdn = 7.75$, $SD = 3.8$) screener to determine study eligibility and, for eligible enrollees, a baseline assessment ($M = 63.1$ min, $Mdn = 60.5$, $SD = 14.9$) consisting of self-report questions and two behavioral tasks (see screener and baseline assessment sections below); and (b) a 3-month follow-up assessment consisting of self-report questions ($M = 7.4$ min, $Mdn = 6.8$, $SD = 2.1$; see 3-month follow-up section below). Only suicidal group participants completed the follow-up because self-injurious thoughts/behaviors (SITB) were the primary outcomes of interest, which are rare among adults with no prior SITBs. All study methods were approved by the Committee for Use of Human Subjects at Harvard University.

Suicide Risk Mitigation

Safety resources were included at the bottom of each survey page (e.g., the suicide and crisis lifeline). At the end of each time point, enrollees answered questions about their past 24-hr suicidal thoughts and current suicide desire and intention to act on suicidal thoughts. Participants who reported that they had thought about suicide in the last 24 hr or rated their current suicide desire or intent as >0 automatically created a personalized safety plan, which they reviewed and received via email if desired. In total, 1,185 safety plans were

completed and emailed if desired. Given that the study procedures were fully automated and responses were not checked in real time, no further risk-mitigation steps were taken. Participants were made aware of these risk-mitigation procedures several times during the informed consent process.

Screener Questions

Enrollees were blind to study in/exclusion criteria and answered basic eligibility questions in an initial screener; 3,221 screeners were completed; 696 enrollees were excluded from participation because they (a) were younger than 21 or older than 65 years (excluded a priori to limit the impact of possible neurodevelopmental or degenerative processes on behavioral task performance); (b) had current cognitive impairment and/or auditory or visual hallucinations in the past week (excluded because these symptoms may interfere with the ability to provide informed consent and/or fully attend to all study procedures); (c) lacked access to headphones or a computer (excluded because these were required to complete the study procedures); (d) had hearing loss/difficulties (e.g., tinnitus; excluded because this could interfere with completing the behavioral task with the aversive noise stimulus); or (e) were located outside of the United States (excluded for legal reasons). Also, 1,918 enrollees were excluded if they reported psychiatric symptoms and/or SITB that were inconsistent with our predefined group eligibility criteria (e.g., experiencing suicidal thoughts before but not in the past 3 months; see Participants section). Enrollees were compensated $1.45 ($10/hr) for completing the screener via the recruitment platform (e.g., Prolific) or digital Amazon gift card emailed to them.

Baseline Assessment

In total, 607 enrollees were eligible and invited to complete the baseline assessment, 70 never initiated, and 59 were lost to attrition. In sum, 478 baseline assessments were completed. Baseline data from 44 enrollees were excluded due to fraudulent responding (e.g., identical internet protocol addresses), 18 for technological problems (e.g., behavioral tasks did not load/save data), 31 for failing >20% of behavioral task attention checks (see Suicide Escape Learning Task section below), and 25 for invalid behavioral task data (e.g., no RTs). Thus, data from 360 enrollees were analyzed. There were no differences on baseline characteristics between enrollees who never initiated or were lost to attrition and those who completed the baseline assessment ($ps > .05$). Participants were compensated $15 for completing the Baseline assessment.

Three-Month Follow-Up Survey

103 (85.8%) of the 120 eligible suicidal group members reported on their past-3-month SITB and symptoms, with $n = 3$ (3%) reported attempting suicide. Participants completed a risk assessment, safety plan, and received clinical resources if they reported any suicidal thoughts in the prior 24 hr. Participants were compensated $5 for completing the 3-month follow-up survey.

Participants

Summary statistics and group differences on baseline characteristics are reported in Table 1. Additional sociodemographic details are reported in Supplemental Method in the online supplemental materials.

Suicidal Group

We used the items from the Self-Injurious Thoughts and Behaviors Interview-Revised (SITBI-R; Fox et al., 2020) to assess death- and suicide-related thoughts, ranging from "passive" thoughts (e.g., "I wish I could disappear or not exist") to more "active" suicide thoughts (e.g., "I should kill myself"). Participants reporting experiencing any of the following thoughts "for longer than a few minutes" in the past 3 months were invited to participate in the baseline assessment: I've "seriously thought about killing myself"; "maybe I should kill myself"; and "I should kill myself." We chose these items to determine eligibility because they relate to suicide specifically, not just death, dying, or nonexistence/escape. The suicidal group consisted of 120 adults (79 identifying as female at birth, 1 nonidentifying) with an average age of 30.7 years ($Mdn = 28$, $SD = 9.8$). 67.5% identified as non-Hispanic White/Caucasian; 25.8% as either lesbian, gay, bisexual, transgender, queer, or another related identity (LGBTQ+); and 4.2% as "unsure."

Psychiatric Group

Participants were included in the psychiatric group if they: (a) met a validated threshold for experiencing a likely psychiatric disorder (i.e., endorsed greater than two symptoms) within the prior 3 months (via Global Appraisal of Individual Needs Short Screener [GAIN-SS 3.0.1]; Dennis et al., 2006) and (b) reported no SITB in their lifetime, including any and all suicidal and death-related thoughts or behavior and NSSI. The psychiatric group consisted of 152 adults (102 identifying as female at birth) with an average age of 31.0 years ($Mdn = 28$, $SD = 10.3$). 64.5% identified as non-Hispanic White/Caucasian; 8.6% as LGBTQ+; and 2.0% as "unsure."

Healthy Control Group

Participants were included in the healthy control group if they: (a) did not meet the GAIN-SS threshold for experiencing any likely psychiatric disorder in the past year (i.e., endorsed less than two past-year symptoms); (b) had low-to-moderate lifetime psychiatric symptom levels (i.e., less than six symptoms 1+ years prior); and (c) never experienced SITB, including no thoughts of death/dying, "passive" or "active" thoughts of suicide, NSSI, or suicidal behavior. The healthy control group consisted of 88 adults (53 identifying as female at birth, one nonidentifying) with an average age of 35.3 years ($Mdn = 31.5$, $SD = 12.5$). 55.7% identified as non-Hispanic White/Caucasian, 8.0% as LGBTQ+, 3.4% as "preferred not to say," and 1.1% as "unsure."

Baseline Assessment

Behavioral Tasks

Participants completed two behavioral tasks, the order of which was counterbalanced. To provide a buffer between tasks, participants rated affectively neutral pictures for approximately 2 min.

The Suicide Escape Learning Task (SELT; See Figure 1 for Details). The SELT is a task designed to measure how suicide

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Group Differences on Demographics and Psychological and Clinical Variables at Baseline (N = 360)

Variable	Group			Test statistic	Effect size (95% CI) for pairwise group comparisons		
	Suicidal ($n = 120$)	Psychiatric ($n = 152$)	Healthy control ($n = 88$)		S versus P	S versus HC	P versus HC
Age[a]	30.7 (9.8)	31 (10.3)	35.3 (12.5)	5.61*	0.04 [−0.27, 0.21]	0.42 [0.14, 0.69]	0.38 [0.11, 0.65]
Sex (female)[b]	79 (65.8%)	102 (67.1%)	53 (60.2%)	4.06	1.02 [0.67, 1.57]	1.41 [0.86, 2.32]	1.47 [0.92, 2.33]
BGLTQ+ (no)[b]	84 (70.0%)	135 (88.8%)	77 (87.5%)	21.31*	2.46 [1.57, 3.84]	2.34 [1.41, 3.89]	1.21 [0.76, 1.91]
Race (White)[b]	81 (67.5%)	98 (64.5%)	49 (55.7%)	14.69	1.46 [0.95, 2.25]	2.32 [1.40, 3.85]	2.09 [1.31, 3.35]
Death-IAT D score[a]	−0.27 (0.44)	−0.49 (.38)	−0.49 (0.37)	12.20*	0.53 [0.29, 0.77]	0.55 [0.27, 0.83]	0.02 [0.24, 0.28]
Int. disorder (likely)[b]	99 (82.5%)	75 (49.3%)	0 (0.0%)	30.57*	3.36 [2.29, 5.74]		
Ext. disorder (likely)[b]	86 (71.7%)	89 (58.6%)	0 (0.0%)	4.47*	1.60 [1.03, 2.47]		
SUD (likely)[b]	22 (18.3%)	7 (4.6%)	0 (0.0%)	11.86*	2.17 [1.39, 3.37]		
Aversive noise rating[a]	25.0 (26.1)	24.9 (26.2)	22.2 (25.5)	0.37	0.00 [−0.23, 0.24]	0.11 [−0.16, 0.38]	−0.10 [−0.36, 0.15]

Note. CI = confidence interval; S = suicidal group; P = psychiatric group; HC = healthy control group; BGLTQ = bisexual, gay, lesbian, transgender, queer, and questioning; IAT = Implicit Association Test; Int. = internalizing; Ext. = externalizing; SUD = substance use disorder; ANOVA = analysis of variance.
[a] Reported as mean (standard deviation), groups compared with an ANOVA test, pairwise Cohen's *d* effect size reported. [b] Reported as percentage (number), groups compared with a χ^2 test, odds ratio effect size reported.
* $p < .05$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

related-stimuli influence decisions to escape distress. While an aversive sound (i.e., metal fork scrapping across a piece of slate) is playing, participants must learn through trial and error which choice (go or no-go) provides the best chance of terminating/escaping the sound. Prior related work shows that the presence of just the aversive sound alone elicits a bias to make an active "go" (vs. to an inhibitory "no-go") response to terminate the sound. This bias has been termed an "active-escape bias" (Millner et al., 2018, 2019). For the current task, we added pictures depicting a suicide attempt from the participant's own perspective (e.g., see Figure 1B and SELT Images section below for details) to test whether suicide information would magnify the active-escape bias. Thus, each trial begins with the aversive noise and the presentation of one of four images, either a suicide-related or positively valenced contrast image. Participants must learn which choice, either go (pressing a spacebar) or no-go (not pressing), will most often (i.e., 80% of the time) allow them to escape the aversive noise given the image presented. The aversive noise is presented over participants' own headphones at approximately 75 dB. The task uses a 2 (picture type) × 2 (response type) factorial design and consists of 30 trials of each condition (120 in total). Each trial lasting 5 s (∼11 min in total). We also included 10 trials to check if the participant was still wearing their headphones at the prescribed volume.

Validity of Administering the SELT Online. To deter participants from removing their headphones or reducing the noise volume, we semirandomly interspersed 10 "headphone check" trials throughout the task, where participants were directed to indicate whether simple mathematical statements (e.g., "two plus two equals four"), which were audible only at prespecified volume levels, were true or false (see Figure 1 in the online supplemental materials). We excluded data from participants failing >20% of headphone checks ($n = 31$; 7.9%).

SELT Images. The suicide images used in the present study depict what it could look like to the participant, from their own eyes/perspective, to begin attempting suicide (e.g., looking down the barrel of a gun; see Figure 2 in the online supplemental materials). Positively valenced contrast images matched the suicide images in terms of body position and action, but contained positive- (cake, candy bar) instead of suicide-related content (gun, knife). For realism and construct validity, we attempted to personalize the stimuli by matching the demographics of the model in the pictures to each participant's own demographics. We created eight different sets of suicide and positive contrast images that varied by skin tone (light, medium, dark, and darkest) and arm shape (more muscular/masculine, less muscular/feminine). Participants were presented with the eight different arm options (Figure 3 in the online supplemental materials) and selected the arm that best resembled their own. They then viewed the corresponding set of suicide and positive contrast pictures while completing the SELT and providing explicit affect ratings (see self-report measures).

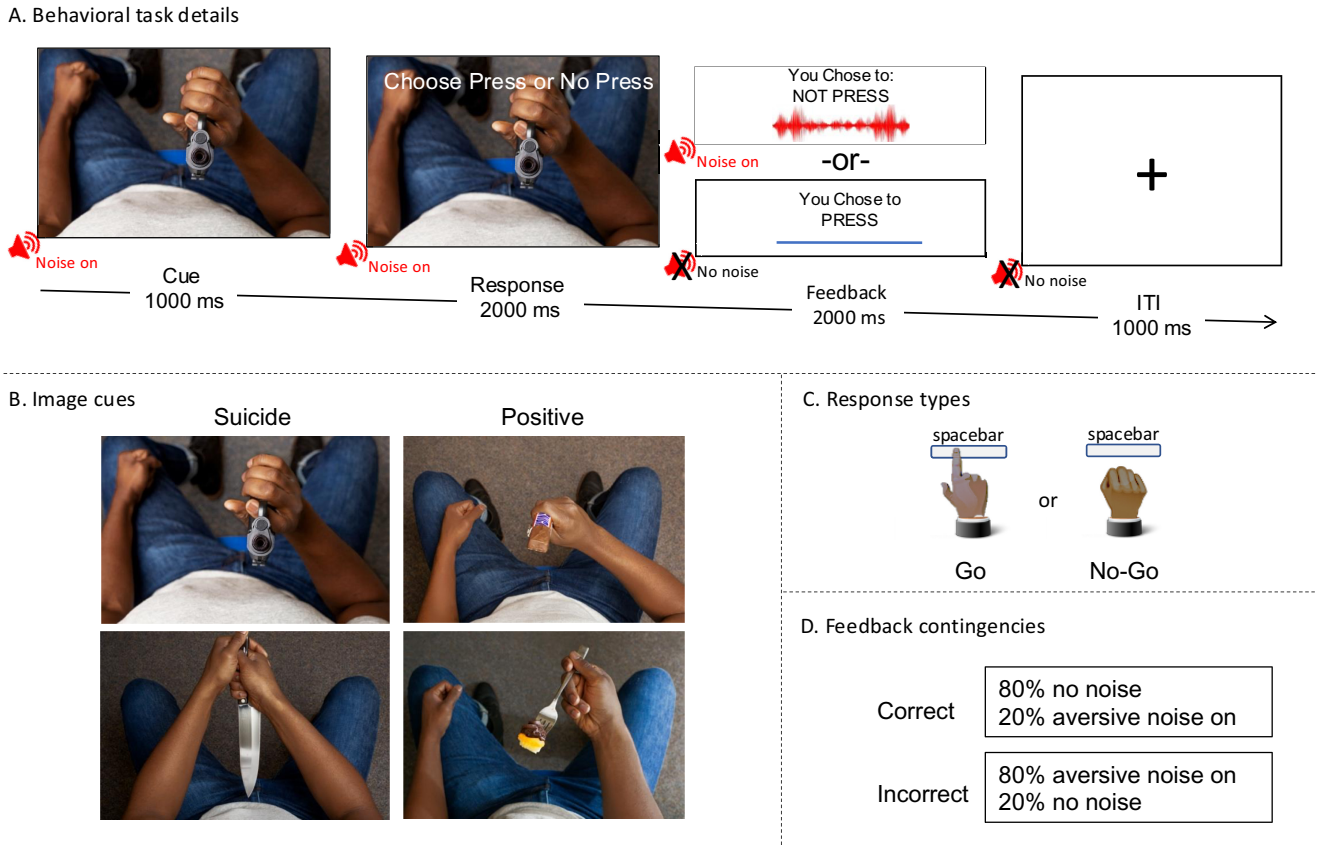
Death Implicit Association Test (Death-IAT; Nock et al., 2010). The Death-IAT is a computerized behavioral task designed to indirectly assess the degree to which participants associate death- and self-related constructs (see Supplemental Method in the online supplemental materials for details). We followed the standard seven-block IAT structure and scoring procedure (Greenwald et al., 2003), calculating a *D* score for each participant, with positive scores representing a stronger association between death and self (e.g., faster responding when death and me are paired relative to when life and me are paired), and negative scores representing a stronger association between life and self. Based on Nosek et al., (2014) we excluded participants with more than 10% of trial RTs <300 or >10,000 ms.

Self-Report Measures

Demographics. Participants completed a brief questionnaire with sociodemographic items measuring: age, sex at birth, race/ethnicity, and bisexual, gay, lesbian, transgender, queer, and questioning (BGLTQ+) status.

Likely Psychiatric Disorder. The Global Appraisal of Individual Needs-Short Screener (GAIN-SS 3.0.1.; Dennis et al., 2006) assessed the likelihood of an internalizing (e.g., depression, anxiety), externalizing (e.g., conduct problems), and substance use disorder within the past 3 months. The GAIN-SS has strong internal consistency and excellent sensitivity and specificity for identifying people with and without a psychiatric disorder (Dennis et al., 2006) and has been successfully used in online studies

Figure 1
Behavioral Task Details



Note. (A) At the start of each trial, participants see a visual cue (e.g., suicide image) and hear a high-pitched, aversive noise of a metal fork scraping on a piece of slate rock. Next, with the aversive noise still on, participants choose between pressing a spacebar (go) or withholding a press (no-go). Participants then experience feedback: either continuation of the aversive noise ("noise on") or silence ("no noise"). Participants see a fixation cross during the ITI before the next trial starts. (B) Two types of visual cues were used: suicide and positively valenced contrast images. Participants' goal was to learn which response to make in relation to each image cue. (C) There are two response-types participants could choose: pressing a keyboard spacebar (go) and withholding a press (no-go). (D) Rewarding feedback (no noise/silence) was offered probabilistically, with correct responses resulting in no noise 80% of the time, and incorrect responses resulting in continuation of the noise 80% of the time. ITI = intertrial interval. See the online article for the color version of this figure.

with suicidal populations (Mortier et al., 2017). In the present study, participants indicated whether they experienced each symptom item during the past month, 2–3 months ago, 4–12 months ago, 1+ years ago, or never. We used the recommended cutoff scores to maximize the accurate classification of experiencing a psychiatric disorder.

Self-Injurious Thoughts and Behavior. Items from the SITBI-R (Fox et al., 2020), a valid and reliable measure, were used to assess frequency, recency, and severity of STB, NSSI, and thoughts about death/dying (e.g., "I'd be better off dead") and nonexistence (e.g., "I wish I could disappear or not exist"), which often co-occur with STB but do not pertain to suicide specifically. Participants were shown a table to help them accurately estimate the frequency of SITB thoughts (e.g., "1 week = 7 days") to 20 years ("20 years = 7,300 days").

Perceived Stress. The Perceived Stress Scale (Cohen et al., 1983) is a 10-ietm self-report measure designed to assess global perceptions of stress. At follow-up, participants rated items relating to stress since the baseline assessment (3 months prior), using a 0 (*never*) to 4 (*very often*). We computed a sum score of items.

Explicit Picture Ratings. Participants explicitly rated how pleasant (valence), threatening, and arousing they found a variety of digital images. We followed the same procedures described in Jaroszewski et al. (2020), where all dimensions were rated on a 9-point scale—for example, pleasantness (−4 [*extremely unpleasant*] to 0 [*neutral*] to 4 [*extremely pleasant*]). Here we report ratings of three different picture types: (a) the two self-relevant suicide pictures used in the SELT, which depicted (1) looking down the barrel of a gun and (2) stabbing oneself in the abdomen; (b) six pictures from the self-directed violence picture system (SDVPS; Nazem et al., 2017), each depicting the suicidal behavior of another person (e.g., a man hanging himself); and (c) 11 pictures from the open affective standardized image set (Kurdi et al., 2017), depicting highly negative valenced images (e.g., an emaciated man, cockroach), none of which contained suicidal/self-injurious content or blood/tissue damage.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Computational Model

Model Description

We fit a computational model to capture hypothetical latent biases (parameters) that potentially generated observed task performance (choice, RT). We used identical modeling tools described in prior related work (Millner et al., 2018, 2019). To be consistent with prior studies, the computational model integrated a reinforcement learning (RL) model and drift-diffusion model (DDM; see Figure 2). The RL model operationalized how, given each trial's state (suicide/positive image), the value of a response (go/no-go) was updated based on reward feedback (successfully [silencing] vs. not escaping the noise). These response values were then used in the DDM to estimate choice and RT probabilities. DDMs incorporate a decision variable specifying the "starting point" of a decision process. The decision process is modeled as a spatial–temporal trajectory that starts at a location on a vertical axis (akin to an intercept) and progresses over time, vertically and horizontally, until it reaches one of two decision boundaries (go vs. no-go). Reaching a boundary represents a decision and, therefore, a response (Millner et al., 2019). Vertically higher starting point values start the decision process closer to the go boundary, making it more likely that the decision process reaches the go versus no-go boundary, thereby representing the go/"active-escape" bias. We specified a DDM with two decision starting point parameters: one for suicide images and the other for positive-contrast images, both free to vary. Thus, for each participant we estimated the initial or starting bias for the relative value of escape actions (go vs. no-go) given the state (suicide/positive image). Importantly, starting point parameters are not dynamically updated based on reward feedback but, instead, are modeled as fixed values. In so doing, the starting point parameters model the preexisting latent "Pavlovian"/innate biases participants possess which are elicited by the suicide and positive cues. We fit this computational model on two different data structures: (a) all data and (b) data stratified by required response (go, no-go), with the model fit separately on each stratified subset. We used this approach because our goal was to isolate the most accurate parameter values possible, thus maximizing model fit. Modeling all data simultaneously regularizes parameter estimates (averaging over go and no-go conditions), whereas modeling stratified data separately allows the model to flexibly isolate parameter values for each required-response condition. Thus, these two models represent the two ends of a spectrum of model flexibility. Modeling all data simultaneously is least flexible. Modeling stratified data separately is most flexible. We provide additional details in the Supplemental Method in the online supplemental materials. As planned, the "active-escape suicide bias" included in main analyses is the decision starting point parameter for suicide images derived from the computational model fit on no-go trials. We decided a priori to estimate this bias from no-go trials because, theoretically, on no-go trials the latent bias for active escape (go) that is mediated by the Pavlovian control system would be incongruent with the required instrumental response (no-go) and, therefore, may be most readily captured.

Model Fitting

Similar to prior studies (Guitart-masip et al., 2012; Millner et al., 2018, 2019; Mkrtchian et al., 2017), we used a hierarchical model fitting procedure. We provide additional details and a formal explanation of model fitting in the Supplemental Method in the online supplemental materials.

Data Analyses

Baseline Characteristics

Group differences in baseline characteristics were assessed with analysis of variance and pairwise *t* tests for normally distributed continuous variables (e.g., age), and χ^2 tests for nominal variables (e.g., sex).

Behavioral Data Group Differences

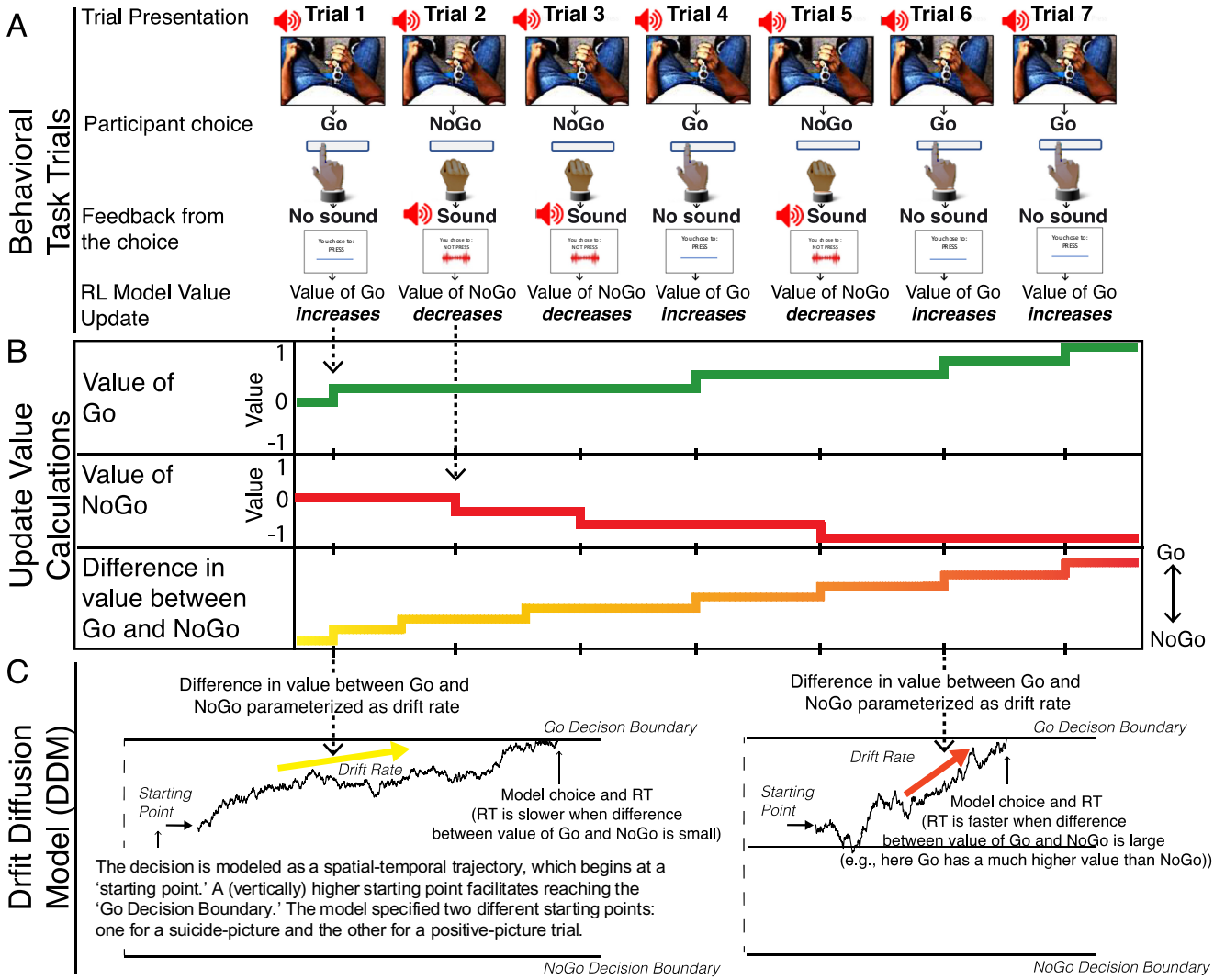
All analyses were preregistered unless otherwise noted. Behavioral data were analyzed using either Bayesian linear or generalized linear mixed-effects regression (BLMER/BGLMER) with the rstanarm package (Goodrich et al., 2020) in R using a Markov chain Monte Carlo (MCMC) sampling algorithm with four chains and 5,000 iterations, with the first 1,000 iterations discarded as burn-ins. We used preset, weakly informative priors. We used a Bayesian instead of frequentist modeling approach because Bayesian models estimate fixed-effect parameter values averaged over the uncertainty in the random effects parameters, whereas frequentist models use point estimation (e.g., maximum likelihood estimation) to estimate fixed-effect parameters, which is prone to issues with singularities, local minima, and boundary points when models are complex, include many factors, or lack enough data (Gelman & Hill, 2007). All mixed-effects models included a by-subject random intercept to account for within-subject correlations in performance (accuracy, RT), and random intercepts for the skin tone and arm shape image set used in the task, counterbalancing order of images used in the task, and counterbalancing order of behavioral tasks. For coefficients of interest, we report the mean (M) of the posterior distribution, 95% Bayesian credible intervals (CIs), and the probability of direction (pd), which is the proportion of the posterior distribution that shares the median's sign (+/−) and reflects the certainty that the effect is meaningfully different from 0 (Makowski, Ben-Shachar, Chen, et al., 2019). We consider coefficients from planned tests with a pd $\geq$ 95% to be "highly likely" and pd $\geq$ 90% as "probable," the latter providing greater confidence in the CI bounds' stability (vs. pd 95%), given the relatively small sample size and potential for skewed posterior distributions (McElreath, 2020). Given that there is no consensus on how or whether to correct for multiple comparisons within a Bayesian framework (Berry & Hochberg, 1999; Gelman et al., 2012; Neath et al., 2018), we caution against interpreting CI and pd values from unplanned tests with the same credence as planned tests. We also report two sampling diagnostic statistics: Rhat, an indicator of MCMC algorithm convergence; and effective sample size (ESS), an indicator of the amount of independent information in autocorrelated sampling chains (Kruschke, 2014). An Rhat $\leq$ 1.01 indicates likely convergence (Vehtari et al., 2021), and for most applications, ESS $\geq$ 10,000 is sufficient for stable estimates of the highest density interval (HDI) bounds (Kruschke, 2021) but can be lower for estimating central tendencies within HDIs (Gong & Flegal, 2016).

Accuracy and RT

To examine the effect of image type on overall accuracy we computed a logistic BGLMER with trial accuracy (yes/no; i.e., whether a

Figure 2

Schematic of RL-DDM Model



Note. (A) Example of suicide-image trials where go is the correct response. Information about each trial (condition, image-type, response, latency, feedback) is used to train the computational model. (B) When a response results in "no sound" feedback, the value associated with that response on that trial-type is increased, whereas value decreases when responses result in the aversive sound feedback. (C) After a response value is updated on each trial, the difference in value between go and no-go is parameterized as the drift rate in the model. On early trials, when participants are uncertain which response is better, the value difference between go versus no-go is small, resulting in a lower drift rate and longer RTs (left panel). Later, when the value difference is greater, the drift rate is higher and RTs are faster (right panel). Another model parameter depicted is the "starting point," which was the only parameter allowed to vary by image-type (suicide vs. positive). A higher starting point starts the decision process closer to the go decision boundary, thus facilitating reaching this boundary faster. This is instantiated behaviorally by a go response/decision. RL-DDM = reinforcement-learning drift diffusion model; RT = response time. Adapted from "Suicidal Thoughts and Behaviors Are Associated With an Increased Decision-Making Bias for Active Responses to Escape Aversive States," by A. J. Millner, H. E. M. den Ouden, S. J. Gershman, C. R. Glenn, J. C. Kearns, A. M. Bornstein, B. P. Marx, T. M. Keane, and M. K. Nock, 2019, *Journal of Abnormal Psychology*, *128*(2), p. 111 (https://doi.org/10.1037/abn0000395). Copyright 2019 by the American Psychological Association. See the online article for the color version of this figure.

participant was correct on a given trial) as the dependent variable (DV) and the three-way interaction between group (suicidal, psychiatric, healthy), required response (go, no-go) and image-type (suicide, positive) as independent variables (IVs). We included the random intercepts mentioned above. When three-way interactions were highly likely (pd 95%) we conducted follow-up analyses with separate Image Type × Group interactions for go and no-go trials, respectively. Group difference contrasts (akin to *t* tests) on estimated marginal means (EMMs) for conditions (e.g., suicide-image no-go trials) were examined using the emmeans package (Lenth et al., 2019) in R. To examine whether effects showing a stronger active-escape bias among suicidal versus psychiatric participants (Millner et al., 2019) replicated, we computed a BGLMER regressing trial accuracy onto a two-way Group × Required

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Response interaction. To examine the effect of image-type on RT we stratified data by required response and computed a BGLMER (gamma likelihood function) with RT (DV) regressed onto an Image Type × Group interaction and random intercepts.

Because the effects of an experimental manipulation can vary between people, significant "causal effect heterogeneity" can lead to misestimating the size, direction, and/or confidence of fixed effects, such as group differences (Bolger et al., 2019). To address this, for each preregistered accuracy and RT analysis, we also ran an unplanned BGLMER, which included the same Group × Picture Type interaction fixed-effect and random intercepts, but also included a correlated random intercept and random slope across picture type for each participant. Given the significant computational time associated with running these more complex "random slope" models on large data via MCMC-based sampling, we used a Bayesian modeling approach that mathematically derives the joint posterior distribution via integrated nested Laplace approximation (INLA; Rue et al., 2009; we note that INLA and MCMC sampling produced nearly identical results for all preregistered random intercept models). First, using INLA, we ran the preregistered "random intercept" models. Next, we ran the "random slope" models. Then we compared the corresponding random intercept to the random-slope model on widely applicable information criterion (WAIC) values, which assess goodness of model fit after adjusting for effective number of model parameters. If a random-slope model's WAIC value was substantially less and/or the fixed-effect coefficient values or pds meaningfully changed relative to the random intercept models, then we interpreted this as evidence of substantial between-person variability (i.e., causal effect heterogeneity) in the effect of picture type.

Computational Model

Group Differences on Parameter Estimates

To examine group differences on model parameter estimates, we computed a series of BGLMs with each parameter value (DV) regressed onto group (IV). Parameter values used were derived from fitting the computational model on go and no-go data separately as described above. Our a priori hypothesis was that, compared to both control groups, the suicidal group would have greater decision starting point parameters for suicide pictures derived from the computational model fit to both go and to no-go data (note that the "active escape suicide bias" is the decision starting point parameter for suicide images derived from the computational model fit on no-go trials). Also, we predicted that the groups would not differ on other parameters (e.g., learning rate). In unplanned follow-up analyses, we examined within group differences between positive and suicide picture bias parameters derived from no-go data by computing: (a) a separate BGLM for each group (akin to dependent samples *t* tests), with parameter values (DV) regressed onto bias type (positive/suicide; IV) and (b) computing a BGLMER, regressing parameter values onto a Bias Type × Group interaction (fixed effects) and a by-subject random intercept, and subsequent contrasts computed on EMMs from this model.

Predicting STB Within the Suicidal Group With the Active Escape Suicide Bias Parameter

Unless otherwise noted, the IV for all models was the "active escape suicide bias," which indexes the latent bias to escape actively (go) elicited by a suicide image when the required instrumental response was to do nothing (no-go). The DVs for the bivariate and incremental concurrent validity analyses were STBs collected at baseline (i.e., lifetime suicidal thought frequency, suicide attempt history) among suicidal participants ($n = 120$). The DVs for the bivariate and incremental prospective validity analyses were suicide thought incidence (yes/no) and suicide thought frequency (i.e., number of days) among respondent suicidal participants ($n = 103$; 85.8% response rate). Covariates added to these models include the following documented STB risk factors collected at baseline: Death-IAT D score, BGLTQ+ status, age, birth sex, race, likely (yes/no) internalizing, externalizing, and substance use disorder (assessed via GAIN-SS). Prospective models also included baseline suicide attempt as a covariate. Our a priori hypotheses were that the active escape suicide bias would be associated with all STBs at baseline and follow-up and would demonstrate incremental validity by capturing unique variance above and beyond other risk factor covariates. We used BGLMs for all analyses. Due to low incidence of suicide attempt at follow-up ($n = 3$, 3%), statistical models prospectively predicting suicide attempt were not reliable and thus excluded. We also conducted planned bivariate models estimating the association between the suicide bias derived from go trials and the STB DVs described above.

Lastly, we preregistered analyses relating to an additional potential behavioral marker of the active escape suicide bias. This additional marker was calculated from the behavioral task data itself (i.e., not the computational model). For clarity and concision, we have elected to present these results in a separate article. Please see the Supplemental Method in the online supplemental materials for additional details on this additional behavioral marker and the rational for reporting separately.

Association Between Active Escape Suicide Bias and Explicit Picture Ratings

A series of unplanned BGLMs were used to test the within-group association between the bias parameter (DV) and average perceived valence, arousal, and threat (IVs) of the two suicide pictures used in the SELT. A whole-sample BGLM with a Group × Valence Rating interaction tested whether groups differed on the association between the bias parameter (DV) and valence ratings of the SELT suicide pictures. Also, within-group BGLMs tested whether the bias parameter (DV) was associated with two other types of pictures (IVs): (a) highly negative but nonsuicidal (e.g., an emaciated man) or (b) the suicidal behavior of a different person (i.e., a person asphyxiating themselves).

Active Escape Suicide Bias Mediating the Relationship Between Suicide Picture Valence and Suicide Attempt

As an unplanned analyses, we used Bayesian mediation analysis (bayestestR package; Makowski, Ben-Shachar, & Lüdecke, 2019) to test whether the bias parameter mediated the direct relationships between perceived valence of the suicide pictures used in the SELT and past suicide attempt reported at baseline. This model controlled for perceived arousal and threat ratings of the suicide pictures.

Transparency and Openness

Hypotheses and primary analyses were preregistered through the Center for Open Science (Jaroszewski, 2023). Data are not available

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

because participants did not consent to data sharing. Computational and analytic code are available upon request. No part of this article was generated by artificial intelligence.

Results

Baseline Characteristics

Group comparisons on all baseline characteristics are reported in Table 1. Also, the suicidal group, M (SD) = −2.61 (1.82), Mdn = −3.5, explicitly rated suicide pictures as more neutral (i.e., less aversive) than both healthy, M (SD) = −3.58 (0.79), Mdn = −4.0, and psychiatric controls, M (SD) = −3.58 (0.69), Mdn = −4.0, ORs = 0.59, CI = [0.51, 0.69], pds = 100%, Rhats = 1.00, ESSs > 14,000. Follow-up analyses among participants in the suicidal group revealed that this group effect was largely driven by those with a past suicide attempt, M (SD) = −1.58 (2.27), Mdn = −1.5, versus those with recent suicidal thoughts but no attempt, M (SD) = −3.05 (1.38), Mdn = −3.5. Within the suicidal group, lower suicide aversion was associated with higher odds of past suicide attempt (n = 36, 30%; OR = 1.56 [1.23, 2.01], pd = 99.9%, Rhat = 1.00, ESS = 10,305; see Figure 4 in the online supplemental materials). Explicit arousal (ORs = 0.99–1.06, pds = 51.9%–56.9%) and threat ratings of the suicide pictures were not related to baseline suicide attempt (OR = 0.89, pds = 51.83%–85.21%, Rhats = 1.00, ESSs > 10,000). These results indicate that people with recent suicidal thoughts and a prior suicide attempt have much lower suicide aversion than others, including people with recent suicidal thoughts alone.

Behavioral Data Group Differences

Accuracy

Overall Accuracy. A BGLMER revealed two probable three-way interactions (Group [Suicidal/Psychiatric/Healthy] × Required Response [Go/No-Go] × Picture Type [Suicide/Positive]), where the suicidal group's accuracy differed from both psychiatric ($M_{posterior}$ = 0.65, OR = 1.91, CI = [0.47, 0.83], pd = 100%; see Figure 3A) and healthy groups ($M_{posterior}$ = 0.35, OR = 1.42, CI = [0.14, 0.56], pd = 99.9%, Rhats = 1, ESSs > 22,600). For greater interpretability, we stratified the data by required response and computed follow-up BGLMERs.

Notably, prior findings (Millner et al., 2019) showing a stronger active-escape bias among suicidal versus psychiatric controls was replicated here, as a BGLMER with a Group × Required Response interaction revealed that, when averaging across picture type, the difference in the suicidal group's accuracy on go versus no-go trials (13.8%) likely differed from both psychiatric (12.1%, $M_{posterior}$ = 0.10, OR = 1.11, CI = [0.01, 0.19], pd = 98.7%) and healthy groups (9.6%, $M_{posterior}$ = 0.22, OR = 1.25, CI = [0.11, 0.32], pd = 99.99%, Rhats = 1.00, ESSs > 20,000).

Accuracy on Go Trials. A BGLMER (see Figure 3A) revealed the suicidal group displayed a greater difference in accuracy on trials requiring a go response in relation to suicide (M = 79.1%) versus positive-picture targets (M = 74.6%) relative to the psychiatric group (M = 74.9% vs. M = 77.6%, $M_{posterior}$ = −0.43, OR = 0.65, CI = [−0.57, −0.29], pd = 100%); however, suicidal and healthy groups responded similarly (M = 77.3% vs. M = 73.7%, $M_{posterior}$ = −0.06, OR = 0.94, CI = [−0.23, 0.10], pd = 76.1%, Rhats = 1.00, ESSs > 18,500). Contrasts computed on EMMs indicate that, on go trials, all three groups displayed similar accuracy to both suicide (ORs = 0.78–1.21, CI = [0.38, 1.75]) and positive pictures (ORs = 0.74–0.78, CI = [0.35, 1.31]). An exploratory test helping estimate the between-person variability on the effect of picture type ("causal effect heterogeneity") revealed nearly identical coefficient sizes, CIs, pds and WAIC values (26,136.73 vs. 26,136.34), indicating that between-person variability was likely not substantial. Together, these results indicate that on trials requiring participants to make an active (go) response to escape/gain relief, the suicide pictures influenced both the suicide and healthy control groups (but not the psychiatric group) to respond actively more often than the positive pictures.

Accuracy on No-Go Trials. A BGLMER revealed the suicidal group displayed a greater difference in accuracy in relation to suicide (M = 61.1%) versus positive pictures (M = 64.9%) compared to both the psychiatric (M = 64.1% vs. M = 63.1%, $M_{posterior}$ = 0.27, OR = 1.31, CI = [0.14, 0.40], pd = 100%; see Figure 3A) and healthy groups (M = 66.7% vs. M = 65.1%, $M_{posterior}$ = 0.31, OR = 1.36, CI = [0.16, 0.45], pd = 99.9%, Rhats = 1.00, ESS > 16,000). EMM contrasts indicate that, on no-go trials, the suicidal group was less accurate to suicide pictures than both the healthy (OR = 0.40, CI = [0.20, 0.67]) and psychiatric groups (OR = 0.72, CI = [0.48, 1.01]). Also, the suicidal group was less accurate than the healthy group to positive pictures (OR = 0.54, CI = [0.26, 0.90]), but similar to the psychiatric group (OR = 0.94, CI = [0.62, 1.33]). An exploratory causal effect heterogeneity analysis revealed nearly identical coefficient sizes, CIs, pds, and WAIC values (31,253.97 vs. 31,254.07), indicating that between-person variability was likely not substantial. Together, these results indicate that on trials requiring participants to make a passive (no-go) response to escape/gain relief, the suicide pictures influenced the suicidal group (but not psychiatric or health controls) to respond impulsively, evinced by making more frequent active (go) responses in relation to suicide pictures compared to positive pictures, even though a passive response (no-go) was required.

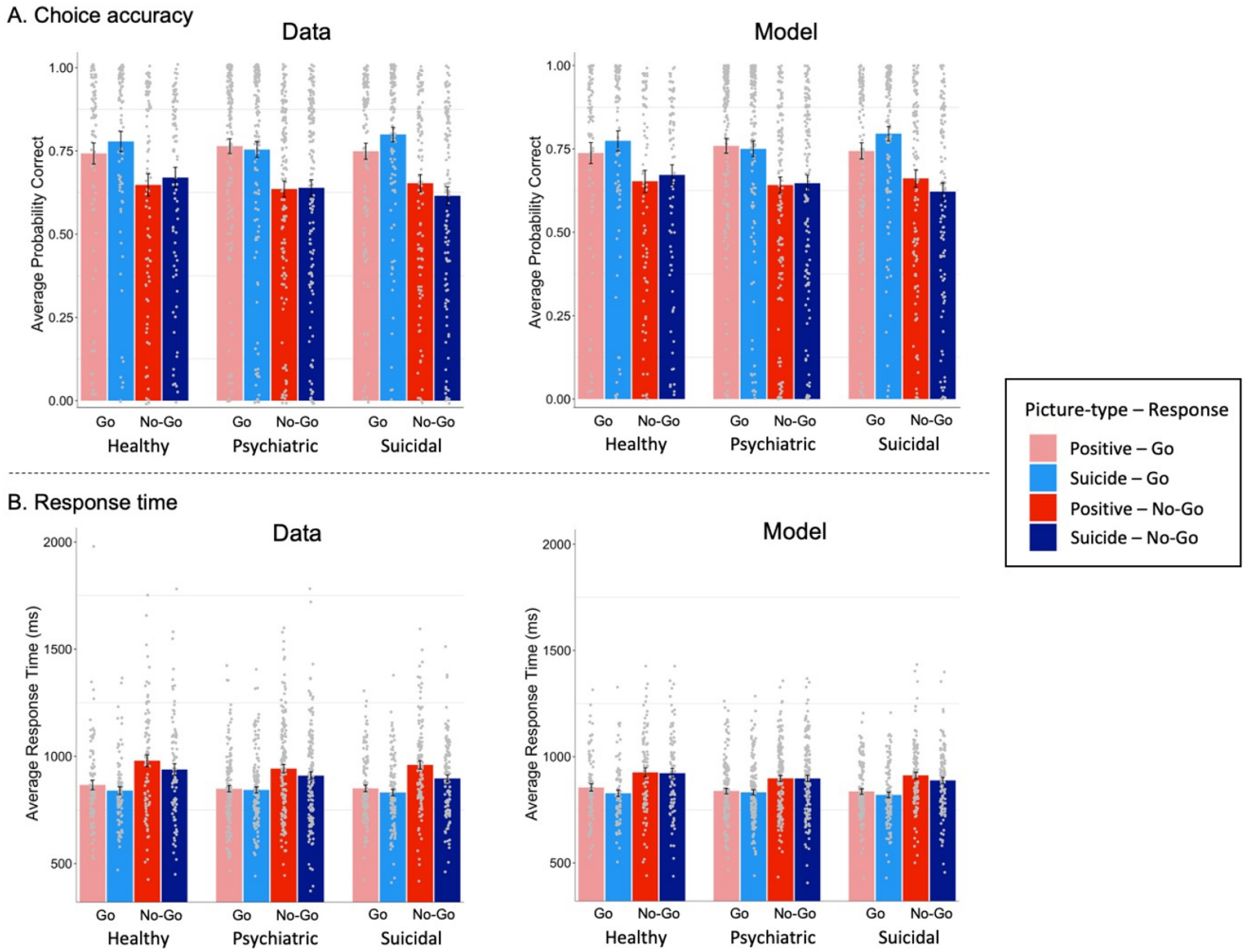
RT

RT on Go Trials. A BGLMER revealed the suicidal group, which responded faster to suicide- versus positive pictures (M = 815 ms vs. M = 835 ms), possibly differed from the psychiatric group, which responded with similar speed to both suicide and positive pictures, M = 824 ms vs. M = 833 ms, $M_{posterior}$ = 0.02, CI = [0.00, 0.03], exp($M_{posterior}$) = 1.02, pd = 94.6% (see Figure 3B). The healthy group also responded slightly faster to suicide pictures (M = 838 ms vs. M = 845 ms), but with less variability than the psychiatric group (SD = 301 ms vs. SD = 312), such that the healthy and suicidal groups did not differ on this interaction, $M_{posterior}$ = 0.00, CI = [−0.03, −0.02], exp($M_{posterior}$) = 1.00, pd = 58.5%, Rhats = 1.00, ESSs > 26,000. EMM contrasts indicate that all three groups responded with similar RT to suicide pictures, exp($M_{posterior}$)s = 1.01–1.02, CI = [0.93, 1.09], and positive pictures, exp($M_{posterior}$)s = 0.98–1.01, CI = [0.92, 1.09]. An exploratory causal effect heterogeneity analysis modeling the between-person variability in the effect of picture type on RT revealed Group × Picture Type interaction coefficient sizes that were similar to the preregistered random intercept model, but larger CIs and thus smaller pds (e.g., 94.6% vs. 83.7%); however, the random-slope model (WAIC = 13,126.45) did not

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 3

Average Accuracy and RT From (Observed) Data and Computational Model (Simulation)



Note. (A) The observed accuracy data (upper left plot) indicate that all three groups displayed a (general) active-escape bias, where accuracy was highest when an active (go) relative to a passive (no-go) response was required to escape the aversive noise. However, the suicidal group in particular displayed a suicide-specific active escape bias, where suicide pictures induced more active responding (relative to positive picture) when no-go was required, thus resulting in lower accuracy. (B) The observed RT data (lower left) indicate participants responded slightly faster on trials requiring an active (go) response with a suicide picture cue, but this difference was larger for the suicidal participants compared to the psychiatric, but not healthy, group. On no-go trials, where participants should not respond at all, all three groups responded slightly faster to the suicide- versus positive pictures, but this difference was slightly larger among the suicidal group relative to both the psychiatric and healthy groups. Qualitatively, RL-DDM model captures both the within- and between-group patterns of accuracy and RT data. Error bars represent standard error of the mean. RL-DDM = reinforcement-learning drift diffusion model; RT = response time. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

provide better fit relative to the random intercept model (WAIC = 10,990.46), suggesting between-person variability may not be substantial. Together, these results indicate that on go trials, the suicide pictures may have influenced both the suicidal and healthy (but not psychiatric) groups to respond faster than the positive pictures.

No-Go Trials. All RTs on no-go trials indicate erroneous responding. An unplanned BGLMER revealed one likely interactions. All three groups responded slightly faster to the suicide- versus positive pictures, with the suicidal group ($M = 868$ ms vs. $M = 921$ ms) responding slightly faster, but not meaningfully so, relative to both the psychiatric, $M = 850$ ms vs. $M = 873$ ms, $M_{\text{posterior}} = 0.02$, CI = [−0.01, 0.06], $\exp(M_{\text{posterior}}) = 1.03$, pd = 90.1% (see Figure 3A) and healthy groups, $M = 856$ ms vs. $M = 864$ ms, $M_{\text{posterior}} = 0.02$, CI = [−0.01, 0.06], $\exp(M_{\text{posterior}}) = 1.02$, pd = 81.3%, Rhats = 1.00, ESSs > 14,000. EMMs indicate that all three groups responded with similar speed on both suicide-picture target, $\exp(M_{\text{posterior}})$s = 1.02–1.04, CI = [0.94, 1.15], and positive-picture targets, $\exp(M_{\text{posterior}})$s = 0.97–0.99, CI = [0.89, 1.10]. An exploratory causal effect heterogeneity analysis revealed nearly identical coefficient sizes, CIs, pds, and larger/worse WAIC values (8,764.88 vs. 8,562.28), indicating that between-person variability was likely not substantial.

In sum, behavioral task accuracy and RT results indicate that the presence of suicide-related stimuli biased the way that people with

recent suicidal thoughts made decisions, leading them to consistently favor an active versus passive means of escaping an acutely distressing experimental context, whereas suicide-related stimuli did not consistently bias either of the control groups.

Computational Model

Group Differences on Parameter Estimates

The computational model fit on stratified data qualitatively captured the observed behavioral data (see Figure 3). We hypothesized that the computational model would reveal that the suicidal group had a stronger latent active (go) bias elicited by the suicide pictures, thus explaining their more frequent and faster active responses to the suicide stimuli. BGLMs revealed that on no-go trials the suicidal group had a higher active escape suicide bias, M (SD) = 0.17 (0.21) (see Figure 5 in the online supplemental materials) than healthy group, M (SD) = 0.13 (0.17), $M_{posterior} = -0.30$, $OR = 0.74$, CI = [−0.59, −0.02], pd = 98.0%, but not psychiatric group, M (SD) = 0.16 (0.18), $M_{posterior} = -0.05$, $OR = 0.95$, CI = [−0.30, 0.19], pd = 65.6%, Rhats = 1.00, ESSs > 35,000; however, on go trials the suicidal group's suicide bias, M (SD) = 0.26 (0.19) did not differ from either healthy, M (SD) = 0.26 (0.22), $M_{posterior} = -0.12$, $OR = 0.88$, CI = [−0.41, 0.16], pd = 79.4%, or psychiatric groups, M (SD) = 0.25 (0.20), $M_{posterior} = -0.15$, $OR = 0.86$, CI = [−0.40, 0.09], pd = 89.4%, Rhats = 1.00, ESSs > 15,000.

Unplanned tests analyzed the relative difference between positive- and suicide-picture biases derived from no-go trials. First, bivariate BGLMs run on each group separately revealed that the suicidal group's active escape suicide bias was higher than their positive bias ($M = 0.17$ vs. 0.14, $OR = 1.19$, CI = [0.94, 1.53], pd = 93.2%, Rhat = 1.00, ESS = 23,028), whereas suicide and positive picture biases did not differ in psychiatric ($M = 0.16$ vs. 0.16, $OR = 0.97$, CI = [0.77, 1.21], pd = 60.27%, Rhat = 1.00, ESS = 26,350) or healthy groups ($M = 0.12$ vs. 0.11, $OR = 0.98$, CI = [0.73, 1.32], pd = 54.6%, Rhat = 1.00, ESS = 27,152). Second, a whole-sample BGLMER revealed a likely Group × Bias-Type interaction, such that the suicidal group displayed a greater difference in suicide versus positive biases compared to the psychiatric group ($OR = 1.39$, CI = [0.89, 2.18], pd = 92.4%) but not the healthy group ($OR = 1.35$, CI = [0.80, 2.24], pd = 87.0%, Rhats = 1.00, ESSs > 30,000). EMM contrasts indicated that suicidal group's active escape suicide bias was higher than their positive bias ($OR = 1.23$, CI = [0.92, 1.58]), but biases did not differ in psychiatric and healthy groups (ORs = 0.98–1.00, CI = [0.69, 1.96]). Together, these results indicate that only the suicidal group had a stronger bias elicited by suicide relative to positive stimuli, leading to increased active (go) escape-responding even when passive responding (no-go) was required to escape the noise. We note here that groups did not differ on no-go or go computational model learning rate parameters (ORs = 0.99–1.00, CI = [0.97, 1.02], pds = 51.5%–73.8%; see Tables 1 and 2 in the online supplemental materials for model parameter descriptives and group comparisons and Tables 4–6 in the online supplemental materials for spearman correlations of computational model parameters).

Predicting STB Within the Suicidal Group With the Active Escape Suicide Bias

Theoretically, on no-go trials, a latent "Pavlovian" or innate bias to escape actively (go) from a suicide stimulus would be incongruent with the passive instrumental response required. Thus, we mainly focus analyses on the "active escape suicide bias," which was derived from no-go trials.

Bivariate concurrent and prospective validity. Contrary to our hypotheses, bivariate BGLMs revealed that stronger active escape suicide biases were associated with lower odds of past suicide attempt ($n = 36$, 30.0% of suicidal group, $OR = 0.04$, CI = [0.00, 0.41], pd = 99.8%, Rhat = 1.00, ESS = 7,353). Again, we do not report models of follow-up suicide attempt due to low incidence ($n = 3$, 3%) and thus low reliability. The active escape suicide bias was not associated with thinking about suicide in terms of lifetime suicidal thought frequency assessed at baseline ($Mdn = 30$, $M = 788.1$, $SD = 1{,}652.2$, $OR = 1.99$, CI = [0.53, 7.38], pd = 86.0%, Rhat = 1.00, ESS = 13,700), or follow-up suicide thought frequency ($Mdn = 2$, $M = 9.8$, $SD = 17.3$, $OR = 0.78$, CI = [0.18, 3.78], pd = 62.3%, Rhat = 1.00, ESS = 13,418), or dichotomous (yes/no) follow-up suicide thought incidence ($n = 69$, 66.9%, $OR = 0.78$, CI = [0.36, 1.34], pd = 78.8%, Rhat = 1.00, ESS = 13,418). Effect sizes and pd values of the active escape suicide bias remained consistent with the above reported results when, in an unplanned analysis, we included as an additional covariate the positive-picture go bias derived from no-go trials to serve as a within subject control for a general go bias. Also, planned bivariate models estimating the association of these outcomes with the suicide bias derived from go trials, where theoretically the Pavlovian and instrumental systems should be congruent with one another, were generally consistent, albeit weaker in coefficient magnitude, with the above results: The go/active-escape suicide bias derived from go trials was associated with past suicide attempt ($OR = 0.23$, CI= [0.02, 11.90], pd = 91.1%, but not past ($OR = 1.36$, CI = [0.43, 14.29], pd = 83.3%) or future suicide thought frequency ($OR = 1.97$, CI = [0.31, 13.9], pd = 76.8%) or incidence ($OR = 1.23$, CI = [0.62, 2.40], pd = 73.2%, Rhats = 1.00, ESSs > 10,000). Overall, these results indicate that the bias to escape actively from suicide stimuli is associated with prior suicidal behavior, with stronger active-escape biases relating to lower odds of past suicide attempt.

Incremental Concurrent and Prospective Validity. A BGLM revealed that the active escape suicide bias is associated with suicide attempt at baseline ($OR = 0.03$, CI = [0.00, 0.79], pd = 98.6%, Rhats = 1.00, ESSs > 10,000) above and beyond other robust STB risk-factor covariates (e.g., Death-IAT D score, BGLTQ + status, birth sex, likely internalizing, externalizing, and substance use disorders; see Table 2). However, similar to the bivariate analyses above, when including additional STB risk factors covariates, the active escape suicide bias was not associated with (a) baseline lifetime frequency of suicide thoughts ($OR = 1.48$, CI = [0.27, 8.32], pd = 67.0%), (b) follow-up incidence (yes/no; $OR = 0.88$, CI = [0.05, 16.04], pd = 54.1%) or (c) frequency of suicide thoughts ($OR = 1.61$, CI = [0.49, 5.60], pd = 78.3%). Interestingly, however, prior suicide attempt reported at baseline (ORs = 2.77–5.07, CI = [1.07, 27.99], pds = 98%–100%) and IAT D score (ORs = 2.18–4.55, CI = [0.83, 27.18], pds = 92%–99%) did display incremental validity in predicting all three suicide thought outcomes (see Table 7 in the online supplemental materials). When, in an unplanned test, including the positive-picture go bias as a covariate in the above models, effect sizes and pd values of the active escape suicide bias did not meaningfully change. Together, these results indicate that the active escape suicide bias is associated

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2

Bayesian Logistic Regression Modeling Prior Suicide Attempt as a Function of the Active Escape Suicide Bias and Covariates Within the Suicidal Group at Baseline ($n = 120$)

Variable	Past suicide attempt reported at baseline ($n = 36$)		
	OR [95% CI]	pd	R^2 (*SD*)
Male sex	0.52 [0.11, 2.28]	.80	0.46 (0.06)
Age	0.95 [0.88, 1.02]	.92	
Likely internalizing disorder: yes	5.02 [0.61, 59.86]	.93	
Likely externalizing disorder: yes	3.04 [0.66, 16.07]	.92	
Likely substance use disorder: yes	2.26 [0.46, 11.54]	.84	
Race: Black/African American	14.63 [1.1, 235.23]	.98	
Race: Hispanic/Latino/a/x	0.00 [0.00, 3.12]	.95	
Race: "mixed"	6.53 [0.52, 89.03]	.93	
Race: White/Caucasian	3.41 [0.54, 23.5]	.90	
BGLTQ status: "other"	26.64 [1.18, 746.83]	.98	
BGLTQ status: yes	2.63 [0.73, 9.7]	.93	
Death-IAT D score	3.86 [0.71, 21.69]	.94	
Active escape suicide bias[a]	0.03 [0.00, 0.74]	.98	

Note. Bold indicates pd ≥ .95. CI = credible interval; pd = probability of direction; BGLTQ = bisexual, gay, lesbian, transgender, queer, and questioning; IAT = Implicit Association Test.

[a] The active escape suicide bias is derived from the computational model that was fit to no-go trials. It indexes the tendency to favor active (go) escape when the required instrumental response is to do nothing (no-go).

with prior suicide attempt above and beyond several robust STB risk factors, including death IAT D score, which itself displays incremental validity in predicting both past and future STB (Nock et al., 2010; Sohn et al., 2021).

Association Between Active Escape Suicide Bias and Explicit Picture Ratings

We hypothesized that explicit ratings of the suicide pictures used in the behavioral task would be related to the active escape suicide bias. A BGLM among suicidal participants revealed that a stronger bias was associated with perceiving suicide as less pleasant/more aversive (valence; $OR = 0.90$, CI = [0.78, 1.05], pd = 91.7%) and less arousing ($OR = 0.89$, CI = [0.77, 1.01], pd = 96.0%), but not related to perceived threat ($OR = 1.01$, CI = [0.87, 1.17], pd = 56.1%, Rhats = 1.00, ESSs > 9,000). Among psychiatric controls, stronger biases were related to perceiving suicide as more aversive ($OR = 0.79$, CI = [0.57, 1.12], pd = 91.0%), more arousing ($OR = 1.09$, CI = [0.98, 1.20], pd = 95.2%) and less threating ($OR = 0.88$, CI = [0.72, 1.06], pd = 90.0%, Rhats = 1.00, ESSs > 10,000), whereas among healthy controls, stronger biases were related to finding suicide more arousing ($OR = 1.08$, CI = [0.97, 1.21], pd = 92.0%), but not related to suicide aversion (valence; $OR = 1.09$, CI = [0.82, 1.57], pd = 68.2%) or threat ($OR = 0.88$, CI = [0.68, 1.08], pd = 86.6%, Rhats = 1.00, ESSs > 14,000). Table 8 in the online supplemental materials displays robust correlations between the active escape suicide bias and each rating dimension for each group; we note here that the bias was related to valence ratings in the suicidal group only ($r = -.17$ [−.34 to .01], $p = .05$). An unplanned whole-sample BGLM revealed a Group × Valence interaction indicating that the suicidal group likely had a stronger association between explicit valence ratings and the active-escape suicide bias relative to the healthy ($OR =$ 1.25, CI = [0.92, 1.77], pd = 92.0%) but not the psychiatric group ($OR = 0.91$, CI = [0.65, 1.29], pd = 71.2%). Together, these results indicate that suicidal participants who perceive the suicide pictures as highly aversive have stronger active escape suicide biases, suggesting that the bias may index a relatively reflexive motivation to avoid suicide.

Notably, unplanned BGLMs revealed that explicit valence of two other types of pictures, which depicted (a) highly negative (but nonsuicidal) content (e.g., an emaciated man), or (b) the suicidal behavior of a different person (i.e., non-self-relevant suicide, e.g., a man hanging himself), are not associated with the bias (ORs = 0.95–1.05, CI = [0.79, 1.46], pds = 70.7%–73.7%). This indicates the active escape suicide bias is associated with self-relevant suicide information specifically (i.e., pictures depicting one's own suicidal behavior), not just highly negative or other-oriented suicide information.

Active Escape Suicide Bias Mediating the Relationship Between Suicide Picture Valence and Suicide Attempt

We hypothesized that explicit aversion to suicide stimuli (i.e., valence ratings) would be associated with STB and that the active escape suicide bias would mediate this relationship. Bayesian mediation analysis revealed that the bias likely mediates the direct relationship between suicide aversion and past suicide attempt (indirect effect: $OR = 1.33$, CI = [0.94, 3.27], pd = 95.2%, 33.5% mediated; direct effect: $OR = 1.75$, CI = [1.30, 2.41], pd = 100%, Rhats = 1.00, ESSs = 24,481–27,042), indicating that people perceiving self-relevant pictures of suicidal behavior with low aversion tend to have a weaker decision-making bias elicited by this suicide-related information, which, in turn, is associated with higher odds of having attempted suicide in the past.

Discussion

The purpose of this study was to test whether suicide-related stimuli can bias the decisions that people with and without recent STB make when they are trying to escape acute distress. There are four main findings. First, people with a past suicide attempt and recent suicidal thoughts show much less explicit aversion to suicide than others, including people with recent suicidal thoughts alone. Second, suicide-related stimuli consistently biased the suicidal group's decisions (but not healthy or psychiatric controls), leading them to favor an active versus passive means of escaping an acutely distressing experimental context. Third, a computational model of behavioral task performance revealed that, relative to both control groups, the suicidal group had a stronger latent decision-making bias to escape actively (go) elicited by the suicide stimuli, which was most evident when the required instrumental response was passive (no-go). This "active escape suicide bias" accounted for variance in choice behavior, helping explain why the suicidal group favored the active escape option in the presence of suicide stimuli. Fourth, within the suicidal group, lifetime history of suicide attempt was associated with lower suicide aversion and weaker active escape suicide biases above and beyond other robust predictors; separately, the active escape suicide bias mediated the association between past suicide attempt and low suicide aversion. Together, these findings suggest that people with low suicide aversion are less biased to actively escape from suicide-related stimuli and, consequently,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

may be less likely to actively limit their exposure to suicide thoughts or methods when trying to escape acute distress. Each finding warrants additional comment.

The suicidal group explicitly rated pictures of suicidal behavior as much less aversive/more pleasant than both control groups. This difference was largely driven by people in the suicidal group with a past suicide attempt. These results replicate and extend prior research (Jaroszewski et al., 2020) showing that, not only do people with past suicidal thoughts have lower suicide aversion than those without prior STB, but that people with past suicidal behavior have even lower suicide aversion. We cannot determine from these cross-sectional data alone whether reduced suicide aversion is a risk factor of suicidal behavior, a consequence of it, or both; however, prior work indicates that people engaging in NSSI also display reduced aversion to self-injury stimuli (Fox et al., 2018; Franklin et al., 2016; Franklin, Lee, et al., 2014; Franklin, Puzia, et al., 2014), and that lower suicide aversion predicts future suicidal attempt (Ribeiro et al., 2020).

Prior research by Millner et al. (2018, 2019) demonstrated that people possess a "Pavlovian" or innate bias to escape an aversive stimulus by doing something active (go) versus doing nothing (no-go). Further, this "active-escape bias" is even stronger among people with STB compared to psychiatric controls. The present study replicates and builds on this prior work by adding suicide picture stimuli to a similar escape-learning task. We hypothesized that the suicide stimuli would magnify the active escape bias among people with recent suicidal thoughts, "pushing" them to make more active responses in relation to a suicide versus positive picture. Our findings supported this hypothesis. The suicidal group displaying higher accuracy and faster RTs on suicide- versus positive-picture trials that required an active (go) response, yet lower accuracy (and faster erroneous RTs) on suicide-picture trials requiring a passive (no-go) response.

The computational model of task behavior revealed that this seemingly discrepant pattern of results is explained by the same mechanism: a relatively stronger latent bias elicited by suicide stimuli that favors active (go) escape regardless of the required instrumental response. When the required response was to do something active, this active-escape bias facilitated the suicidal group's learning/accuracy; however, when the required response was to do nothing, the bias hindered accuracy. These results align with prior work on Pavlovian–instrumental interactions, which indicates that when Pavlovian and instrumental systems are congruent with one another (i.e., ascribe greater relative value to the same action) learning/performance may benefit but, when incongruent, learning can suffer (e.g., Guitart-Masip et al., 2012; Millner et al., 2018). Importantly, a key instrumental-value-learning parameter (i.e., learning rate) did not differ across groups. Together, these results suggest that suicide stimuli per se consistently biased the suicidal group's decision making through a relatively reflexive, stimulus-driven process. These results extend findings demonstrating that not only can aversive stimuli elicit rigid biases that influence instrumental behavior (Lindström et al., 2015; Millner et al., 2018, 2019), but also that the presence of some such biases depend on individual differences, such as recent suicidal thoughts.

Contrary to our hypotheses, within the suicidal group, participants with stronger active escape suicide biases were much less likely to have attempted suicide prior to starting the study. Importantly, this association held even after statistically controlling for a variety of other robust STB risk factors, including LGBTQ+ status, likely psychiatric disorder(s), and D-IAT score (Nock et al., 2010; Tello et al., 2020). We do not yet know whether the active escape suicide bias is associated with future suicidal behavior, since we were not able to reliably estimate this effect due to low suicide attempt incidence ($n = 3, 3\%$) at 3-month follow-up. Notably, the active escape suicide bias parameter was not associated with concurrent or prospective suicidal thought frequency/incidence, suggesting the bias relates to suicidal behavior specifically. Likewise, the bias was not associated with explicit valence ratings of other aversive pictures, which depicted highly negative but nonsuicidal content or the suicidal behavior of a different person (e.g., a man hanging himself), suggesting the active escape suicide bias relates to self-relevant suicide-information specifically. Together, these results indicate that the active escape suicide bias parameter, which indexes the behavioral tendency for self-relevant suicide information to elicit active (vs. passive) attempts to escape distress when the required instrumental response is to wait and do nothing, is a novel, specific, incrementally valid, and objectively assessed suicide-attempt correlate.

Lastly, the active escape suicide bias parameter mediated the relationship between higher suicide aversion and lower odds of past suicide attempt. The direction of this mediation cannot be determined because these data are cross-sectional; however, to date, more evidence supports the possibility that lower suicide aversion is a risk factor (not a mere consequence [cf. Joiner, 2005; Van Orden et al., 2010]) of suicide attempt (Franklin et al., 2016; Ribeiro et al., 2020). This mediational association could play out in real-life contexts in several ways. For example, when someone with recent suicidal thoughts and high suicide aversion is in distress and encounters a suicide-related stimulus (e.g., potential suicide method), they might experience a strong reflexive bias or "push" to actively escape from this stimulus, which could be protective. For instance, when the suicide stimulus is endogenous, like a suicidal thought/mental image, a stronger bias may push someone to actively suppress the thought and/or think about different ways to gain relief/escape, thereby decreasing the chances that suicide enters into one's "consideration set" of potential ways to gain relief (Dombrovski & Hallquist, 2022). Also, when the suicide-related stimulus is exogenous, like an actual gun in one's home, a stronger bias may push someone to avert their gaze or physically turn their body away from it, thereby possibly decreasing the chances of either simulating/imagining or actually engaging in suicidal behavior. By the same token, those with low suicide aversion may experience less of a reflexive push to actively escape from a suicide-related stimulus, thus increasing the chances of actually attempting suicide. This possibility is consistent with recent theories that people possess a natural aversive barrier to self-injurious-related stimuli that ordinarily inhibits self-harm (Hooley & Franklin, 2018; Joiner, 2005) and that risk for suicidal behavior increases when one is only able to generate and consider relatively few and unelaborated ways to gain relief due to cognitive impairments, mood congruent Pavlovian biases, and impaired value-learning and outcome simulation (Dombrovski & Hallquist, 2022). Interestingly, the mediation finding also raises the possibility that increasing aversion to suicide might reduce risk for suicidal behavior (e.g., via counter-conditioning procedures) just as increasing NSSI aversion reduces NSSI (Franklin et al., 2016). Notably, the healthy and psychiatric controls groups displayed high suicide aversion, but within these groups suicide aversion was less strongly

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

associated with the active escape suicide bias. This pattern of results suggests that the relationship between suicide aversion and the bias is somewhat specific to participants with recent STB. Prior work has shown that stimuli vary on their capacity to elicit direct, automatic action tendencies (e.g., approach, avoid), likely because some stimuli are perceived as more/less ambiguous and, thus, require more context and interpretation (Hans Phaf et al., 2014). It is possible that, compared to people with STB, people who have never thought about suicide before (i.e., controls) likely perceive pictures depicting their own hypothetical suicidal behavior as less relevant and/or more ambiguous and, therefore, are less consistently influenced by such stimuli (Jaroszewski et al., 2022).

The findings from this study must be interpreted in the context of several limitations. First, this study used an unpleasant noise stimulus to induce acute distress. This stimulus was effective and afforded high experimental control, but has limited ecological validity because people likely rarely if ever attempt suicide to escape an aversive noise. Future studies could address this by using other, more commonly experienced distressing stimuli (e.g., emotional distress, monetary loss). Second, participants were recruited online. Although prior research shows that online samples are representative of the broader population, it is possible that certain aspects of this sample (e.g., high technological literacy, openness to/familiarity with research) may not generalize to some individuals with STB. Future studies could address this by recruiting from in-person settings (e.g., emergency department, inpatient unit). Third, group in/exclusion was determined via self-report. Although participant's report is necessary when assessing constructs like STB (Jaroszewski et al., 2020), this approach has many well-known limitations. We took a number of steps to mitigate these concerns, including not disclosing eligibility criteria; using valid and reliable self-report assessments; embedding in/exclusion items in larger surveys to prevent selective misreporting, self-presentation bias, and/or demand characteristics; and assessing some constructs (e.g., suicidal thoughts) multiple times to detect inconsistent responding. Fourth, although the suicidal group was relatively large ($n = 120$) and $>85\%$ responded at follow-up, few participants reported attempting suicide at follow-up ($n = 3$, 3%), restricting variance in this key DV. Future studies could address this by recruiting samples enriched for suicide attempt incidence (e.g., people receiving inpatient psychiatric care). Fifth, although nonsuicidal negatively valenced pictures were not correlated to the active escape suicide bias, it is possible negative nonsuicidal information exerts a similar influence on escape-related decision making that suicidal information does. Future studies should directly test different kinds of valenced stimuli in an escape learning task to better isolate the effect of suicide-related information specifically. Sixth, we did not attempt to match or misalign participants' observable demographics (e.g., skin tone) with the non-self-relevant aversive pictures they rated. This may have differentially impacted some pictures' perceived self-relevance, thereby influencing the strength of association with the active escape suicide bias. Future research could test this possibility. Seventh, consistent with prior work (Millner et al., 2019), we fit variations of only one type of computational model to the data (RL-DDM with separate starting points) and, therefore, do not show that this approach fits the data better than reasonable alternatives, which future work could investigate.

Despite these limitations, our findings contribute to a growing literature that uses decision science to investigate factors and processes related to STB. We showed that lower aversion to suicide is associated with greater odds of past suicidal behavior. Using an experimental task and computational model, we discovered that suicide information elicits a decision-making bias that "pushes" some people with recent suicidal thoughts to actively escape from suicide-related stimuli. Those regarding suicide as less aversive possess a weaker bias and, consequently, may be less likely to actively limit their exposure to suicide stimuli (e.g., suicide methods), thus increasing risk of attempting suicide. This active escape suicide bias represents an objectively assessed, cognitive factor linking individual differences in perception and affect toward suicide-specific stimuli to prior suicidal behavior. This line of research may provide additional insight into processes that influence people to select suicide to escape as well as help identify novel clinical targets and precision methods to help prevent this tragic decision. It is also possible that this approach could be useful in studying other clinical phenomena (e.g., substance misuse, anxiety) that involve escape and are influenced by stimulus-driven processes (Carter & Tiffany, 1999).

References

Baumeister, R. F. (1990). Suicide as escape from self. *Psychological Review*, *97*(1), 90–113. https://doi.org/10.1037/0033-295X.97.1.90

Berry, D. A., & Hochberg, Y. (1999). Bayesian perspectives on multiple comparisons. *Journal of Statistical Planning and Inference*, *82*(1–2), 215–227. https://doi.org/10.1016/S0378-3758(99)00044-0

Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in psychology are heterogeneous. *Journal of Experimental Psychology: General*, *148*(4), 601–618. https://doi.org/10.1037/xge0000558

Bryan, C. J., Rudd, M. D., & Wertenberger, E. (2013). Reasons for suicide attempts in a clinical sample of active duty soldiers. *Journal of Affective Disorders*, *144*(1–2), 148–152. https://doi.org/10.1016/j.jad.2012.06.030

Carter, B. L., & Tiffany, S. T. (1999). Meta-analysis of cue-reactivity in addiction research. *Addiction*, *94*(3), 327–340. https://doi.org/10.1046/j.1360-0443.1999.9433273.x

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396. https://doi.org/10.2307/2136404

Coppersmith, D. D. L., Millgram, Y., Kleiman, E. M., Fortgang, R. G., Millner, A. J., Frumkin, M. R., Bentley, K. H., & Nock, M. K. (2023). Suicidal thinking as affect regulation. *Journal of Psychopathology and Clinical Science*, *132*(4), 385–395. https://doi.org/10.1037/abn0000828

Crane, C., Shah, D., Barnhofer, T., & Holmes, E. A. (2012). Suicidal imagery in a previously depressed community sample. *Clinical Psychology & Psychotherapy*, *19*(1), 57–69. https://doi.org/10.1002/cpp.741

Dennis, M. L., Chan, Y.-F., & Funk, R. R. (2006). Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *The American Journal on Addictions*, *15*(s1), 80–91. https://doi.org/10.1080/10550490601006055

Dombrovski, A. Y., Clark, L., Siegle, G. J., Butters, M. A., Ichikawa, N., Sahakian, B. J., & Szanto, K. (2010). Reward/punishment reversal learning in older suicide attempters. *American Journal of Psychiatry*, *167*(6), 699–707. https://doi.org/10.1176/appi.ajp.2009.09030407

Dombrovski, A. Y., & Hallquist, M. N. (2022). Search for solutions, learning, simulation, and choice processes in suicidal behavior. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(1), Article e1561. https://doi.org/10.1002/wcs.1561

Dombrovski, A. Y., Hallquist, M. N., Brown, V. M., Wilson, J., & Szanto, K. (2019). Value-based choice, contingency learning, and suicidal behavior in mid- and late-life depression. *Biological Psychiatry*, *85*(6), 506–516. https://doi.org/10.1016/j.biopsych.2018.10.006

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Dombrovski, A. Y., Szanto, K., Siegle, G. J., Wallace, M. L., Forman, S. D., Sahakian, B., Reynolds, C. F., & Clark, L. (2011). Lethal forethought: Delayed reward discounting differentiates high- and low-lethality suicide attempts in old age. *Biological Psychiatry*, *70*(2), 138–144. https://doi.org/10.1016/j.biopsych.2010.12.025

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli mere-exposure research. *Psychological Science Research Article*, *13*(6), 513–519. https://doi.org/10.1111/1467-9280.00490

Fox, K. R., Harris, J. A., Wang, S. B., Millner, A. J., Deming, C. A., & Nock, M. K. (2020). Self-injurious thoughts and behaviors interview—Revised: Development, reliability, and validity. *Psychological Assessment*, *32*(7), 677–689. https://doi.org/10.1037/pas0000819

Fox, K. R., Ribeiro, J. D., Kleiman, E. M., Hooley, J. M., Nock, M. K., & Franklin, J. C. (2018). Affect toward the self and self-injury stimuli as potential risk factors for nonsuicidal self-injury. *Psychiatry Research*, *260*, 279–285. https://doi.org/10.1016/j.psychres.2017.11.083

Franklin, J. C., Fox, K. R., Franklin, C. R., Kleiman, E. M., Ribeiro, J. D., Jaroszewski, A. C., Hooley, J. M., & Nock, M. K. (2016). A brief mobile app reduces nonsuicidal and suicidal self-injury: Evidence from three randomized controlled trials. *Journal of Consulting and Clinical Psychology*, *84*(6), 544–557. https://doi.org/10.1037/ccp0000093

Franklin, J. C., Lee, K. M., Puzia, M. E., & Prinstein, M. J. (2014). Recent and frequent nonsuicidal self-injury is associated with diminished implicit and explicit aversion toward self-cutting stimuli. *Clinical Psychological Science*, *2*(3), 306–318. https://doi.org/10.1177/2167702613503140

Franklin, J. C., Puzia, M. E., Lee, K. M., & Prinstein, M. J. (2014). Low implicit and explicit aversion toward self-cutting stimuli longitudinally predict nonsuicidal self-injury. *Journal of Abnormal Psychology*, *123*(2), 463–469. https://doi.org/10.1037/a0036436

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 187–232. https://doi.org/10.1037/bul0000084

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211. https://doi.org/10.1080/19345747.2011.618213

gong, L., & Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, *25*(3), 684–700. https://doi.org/10.1080/10618600.2015.1044092

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *rstanarm: Bayesian applied regression modeling via Stan* (R package, version 2.21.1). https://mc-stan.org/rstanarm

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Guitart-Masip, M., Duzel, E., Dolan, R., & Dayan, P. (2014). Action versus valence in decision making. *Trends in Cognitive Sciences*, *18*(4), 194–202. https://doi.org/10.1016/j.tics.2014.01.003

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, *62*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Hans Phaf, R., Mohr, S. E., Rotteveel, M., & Wicherts, J. M. (2014). Approach, avoidance, and affect: A meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Frontiers in Psychology*, *5*, Article 378. https://doi.org/10.3389/fpsyg.2014.00378

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Holmes, E. A., Crane, C., Fennell, M. J. V., & Williams, J. M. G. (2007). Imagery about suicide in depression-"Flash-forwards"? *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(4), 423–434. https://doi.org/10.1016/j.jbtep.2007.10.004

Hooley, J. M., & Franklin, J. C. (2018). Why do people hurt themselves? A new conceptual model of nonsuicidal self-injury. *Clinical Psychological Science*, *6*(3), 428–451. https://doi.org/10.1177/2167702617745641

Jaroszewski, A. C. (2023). *Escaping an aversive context containing suicide information: Prospective study*. https://osf.io/eap6x

Jaroszewski, A. C., Huettig, J. L., Kleiman, E. M., Franz, P. J., Millner, A. J., Joyce, V. W., Nash, C. C., & Nock, M. K. (2022). Examining implicit positive affect toward suicide among suicidal and nonsuicidal adults and adolescents. *Suicide and Life-Threatening Behavior*, *52*(3), 525–536. https://doi.org/10.1111/sltb.12843

Jaroszewski, A. C., Kleiman, E. M., Simone, P. K., & Nock, M. K. (2020). First-person stimuli: Improving the validity of stimuli in studies of suicide and related behaviors. *Psychological Assessment*, *32*(7), 663–676. https://doi.org/10.1037/pas0000823

Joiner, T. (2005). *Why people die by suicide*. Harvard University Press.

Jollant, F., Bellivier, F., Leboyer, M., Astruc, B., Torres, S., Verdier, R., Castelnau, D., Malafosse, A., & Courtet, P. (2005). Impaired decision making in suicide attempters. *American Journal of Psychiatry*, *162*(2), 304–310. https://doi.org/10.1176/appi.ajp.162.2.304

Jollant, F., Lawrence, N. S., Olie, E., O'Daly, O., Malafosse, A., Courtet, P., & Phillips, M. L. (2010). Decreased activation of lateral orbitofrontal cortex during risky choices under uncertainty is associated with disadvantageous decision-making and suicidal behavior. *NeuroImage*, *51*(3), 1275–1281. https://doi.org/10.1016/j.neuroimage.2010.03.027

Kidd, S. A. (2004). "The walls were closing in, and we were trapped": A qualitative analysis of street youth suicide. *Youth & Society*, *36*(1), 30–55. https://doi.org/10.1177/0044118X03261435

Kleiman, E. M., Coppersmith, D. D. L., Millner, A. J., Franz, P. J., Fox, K. R., & Nock, M. K. (2018). Are suicidal thoughts reinforcing? A preliminary real-time monitoring study on the potential affect regulation function of suicidal thinking. *Journal of Affective Disorders*, *232*, 122–126. https://doi.org/10.1016/j.jad.2018.02.033

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and stan*. Academic Press.

Kruschke, J. K. (2021). Bayesian Analysis reporting guidelines. *Nature Human Behaviour*, *5*(10), 1282–1291. https://doi.org/10.1038/s41562-021-01177-7

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior Research Methods*, *49*(2), 457–470. https://doi.org/10.3758/s13428-016-0715-3

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *Package "emmeans."*

Lindström, B., golkar, A., & Olsson, A. (2015). A clash of values: Fear-relevant stimuli can enhance or corrupt adaptive behavior through competition between Pavlovian and instrumental valuation systems. *Emotion*, *15*(5), 668–676. https://doi.org/10.1037/emo0000075

Liu, R. T., & Spirito, A. (2019). Suicidal behavior and stress generation in adolescents. *Clinical Psychological Science*, *7*(3), 488–501. https://doi.org/10.1177/2167702618810227

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, Article 2767. https://doi.org/10.3389/fpsyg.2019.02767

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). Bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), Article 1541. https://doi.org/10.21105/joss.01541

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.

Millner, A. J., Den Ouden, H. E. M., Gershman, S. J., Glenn, C. R., Kearns, J. C., Bornstein, A. M., Marx, B. P., Keane, T. M., & Nock, M. K. (2019). Suicidal thoughts and behaviors are associated with an increased decision-making bias for active responses to escape aversive states. *Journal of Abnormal Psychology*, *128*(2), 106–118. https://doi.org/10.1037/abn0000395

Millner, A. J., Gershman, S., Nock, M., & den Ouden, H. (2018). Pavlovian control of escape and avoidance. *Journal of Cognitive Neuroscience*, *30*(10), 1379–1390. https://doi.org/10.1162/jocn_a_01224

Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P., & Robinson, O. J. (2017). Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biological Psychiatry*, *82*(7), 532–539. https://doi.org/10.1016/j.biopsych.2017.01.017

Mortier, P., Demyttenaere, K., Auerbach, R. P., Cuijpers, P., Green, J. G., Kiekens, G., Kessler, R. C., Nock, M. K., Zaslavsky, A. M., & Bruffaerts, R. (2017). First onset of suicidal thoughts and behaviours in college. *Journal of Affective Disorders*, *207*, 291–299. https://doi.org/10.1016/j.jad.2016.09.033

Naghavi, M. (2019). Global, regional, and national burden of suicide mortality 1990 to 2016: Systematic analysis for the Global Burden of Disease Study 2016. *BMJ*, *364*, Article l94. https://doi.org/10.1136/bmj.l94

Nazem, S., Forster, J. E., & Brenner, L. A. (2017). Initial validation of the self-directed violence picture system (SDVPS). *Psychological Assessment*, *29*(12), 1496–1504. https://doi.org/10.1037/pas0000448

Neath, A. A., Flores, J. E., & Cavanaugh, J. E. (2018). Bayesian multiple comparisons and model selection. *WIRES Computational Statistics*, *10*(2), Article e1420. https://doi.org/10.1002/wics.1420

Nock, M. K., & Banaji, M. R. (2007). Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry*, *164*(5), 820–823. https://doi.org/10.1176/ajp.2007.164.5.820

Nock, M. K., Jaroszewski, A. C., Deming, C. A., Glenn, C. R., Millner, A. J., Knepley, M., Naifeh, J. A., Stein, M. B., Kessler, R. C., & Ursano, R. J. (2025). Antecedents, reasons for, and consequences of suicide attempts: Results from a qualitative study of 89 suicide attempts among army soldiers. *Journal of Psychopathology and Clinical Science*, *134*(1), 6–17.

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science*, *21*(4), 511–517. https://doi.org/10.1177/0956797610364762

Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and using the brief implicit association test: Recommended scoring procedures. *PLoS ONE*, *9*(12), Article e110938. https://doi.org/10.1371/journal.pone.0110938

O'Brien, K. H. M. M., Nicolopoulos, A., Almeida, J., Aguinaldo, L. D., & Rosen, R. K. (2021). Why adolescents attempt suicide: A qualitative study of the transition from ideation to action. *Archives of Suicide Research*, *25*(2), 269–286. https://doi.org/10.1080/13811118.2019.1675561

Ribeiro, J. D., Harris, L. M., Linthicum, K. P., & Bryen, C. P. (2020). Do suicidal behaviors increase the capability for suicide? A longitudinal pretest–posttest study of more than 1,000 high-risk individuals. *Clinical Psychological Science*, *8*(5), 890–904. https://doi.org/10.1177/2167702620921511

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Shneidman, E. S. (1987). A psychological approach to suicide. In G. R. VandenBos & B. K. Bryant (Eds.), *Cataclysms, crises, and catastrophes: Psychology in action* (pp. 147–183). American Psychological Association. https://doi.org/10.1037/11106-004

Sohn, M. N., McMorris, C. A., Bray, S., & McGirr, A. (2021). The death-implicit association test and suicide attempts: A systematic review and meta-analysis of discriminative and prospective utility. *Psychological Medicine*, *51*(11), 1789–1798. https://doi.org/10.1017/S0033291721002117

Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology*, *59*(4), 239–245. https://doi.org/10.1037/h0047274

Tello, N., Harika-Germaneau, G., Serra, W., Jaafari, N., & Chatard, A. (2020). Forecasting a fatal decision: Direct replication of the predictive validity of the suicide-implicit association test. *Psychological Science*, *31*(1), 65–74. https://doi.org/10.1177/0956797619893062

Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., & Joiner, T. E., Jr. (2010). The interpersonal theory of suicide. *Psychological Review*, *117*(2), 575–600. https://doi.org/10.1037/a0018697

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Received January 25, 2024
Revision received January 22, 2025
Accepted January 23, 2025 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.